<u>Response to Reviewer 1's comments</u>

We thank Reviewer 1 for the time to go through our manuscript in details. This manuscript describes a new and efficient method to produce a physical TC event set in the western North Pacific basin. In general, reviewers think after careful revision, the results of this study is of great interest and relevance, and it will be a useful contribution to the field of TC occurrence risk assessment. Here is our point-to-point response to Reviewer 1's comments. Reviewer 1's comments are in red.

<u>Response to Major Issues</u>

You have explained that you only used 6-hourly surface wind speed (i.e., lines 67-69 and discussed in Section 4.2 as well), is your risk assessment model able to comprehend a full TC risk? What about TC-rainfall (you only mentioned about it in the last 2-3 sentences in Conclusions) and TC-induced storm surge, which are more significant than winds in many regions of the world. For example, the recent Hurricane Florence in USA and Mangkut in South China caused significant damaged due to surge and rainfall, respectively. Therefore, under such circumstances, what is the applicability of your proposed approach? You need to think about the generalization, replicability, and adoptability of your approach from a wider perspective and not only from the study area.

We are thankful for the reviewer's comment to the overall impact of Typhoons and their differentiated reasoning in form of the different meteorological variables concerned. In this study, we present a method, which addresses the basic, critical issue in typhoon risk assessments – a robust methodology to determine the real frequency of a tropical cyclone (TC) occurrence with high socioeconomic impact *potential*. We are doing so by automatically identify and tracking severe, damage relevant tropical cyclones in a data set being representative of several thousand years of "observations", the TIGGE archive. Based on the large amount of data from the different multi-model multi-member ensembles to be analysed, in this study, we use thus a computationally inexpensive approach: by applying an impact-oriented tracking algorithm WiTRACK (Leckebusch et al., 2008; Kruschke, 2015; Befort et al., 2020) on multi-model ensemble forecasts to generate a large, physical consistent TC event set. This *identification* of major events is thus based on one meteorological variable only (wind speed), but is capable of identifying events of *general* loss relevance. We recently demonstrated in Befort et al. (2020) that the tracking algorithm used in this study (WiTRACK), can automatically identify the relevant (over 90%) of TCs with high overall socioeconomic impact (e.g., above 3,000 million RMB or 440 million US$ to mainland China). This implies the event set generated by our approach is in principle suitable for general TC risk assessments, as well as for an assessment of the hazards frequency-intensity distribution specifically. As we fully agree with Reviewer 1 that TC-rainfall and TC-induced storm surge are also important factors in assessing the full impact (risk) of TC to society, we would like to point out that the events identified and used in this study are including those where potential loss is caused not only by the direct wind force but also by secondary natural hazards, e.g. flooding by precipitation, or potential storm surge by the combined wind and pressure impact.

Therefore, the applicability of our proposed approach is directly visible and fundamental. While studies like e.g. Sajjad and Chan (2019) are focussing on an overall assessment of integrated risks (hazard x vulnerability/resilience; further comments to this topic, please find below) based on observed tracks from only some 60-70 years (of inhomogeneous data of different quality), our approach utilizes data (in this case multi-model synoptic forecasts) representative of ~10k years. To quantify robustly the probability of occurrence of e.g., a 100-year event from 60-70 years of observations is widely impossible (because it may potentially not have realised during this short observational period). Consequently, such an approach is not used e.g., in industry applications and is only of limited use for risk assessments as the **underlying** hazard component probability distribution is not sufficiently known.

The reviewer mentioned the topic of "*generalization, replicability, and adoptability*". We fully agree and this is exactly the reason why we developed this approach.

- By using an impact-based identification of the natural hazard relative to the datasets typical characteristics (i.e., the 98th percentile of surface-near wind speeds), the assessment of the hazard is accounting of potential biases in the data set under investigation (e.g., AOGCM, NWP, or RCM simulations) and is linked to overall losses out of those events relative to the climatological distribution of these data sets.
- By utilizing ensemble forecasts (i.e., TIGGE's unrealised TC events), we secure the statistical robustness of our intensity-frequency distribution assessment to create an event-set being representative of the risk of occurrence of events of potentially damaging characteristic (so called tail events).

We have demonstrated the applicability and transferability of this approach in several studies for different hazards and different regions (e.g. Leckebusch et al., 2008; Nissen et al., 2014; Osinski et al., 2016; Befort et al., 2015; Befort et al., 2016; Befort et al., 2019; Walz and Leckebusch, 2019)

A more in-depth investigation of the contribution of different drivers (e.g., extreme precipitation and storm surges) of loss and damage during a (non-realised) severe TC event would be beyond the scope of this study. Nevertheless, once a representative TC event set is derived, which provides robust information of the frequency of high impact TCs, the impact of extreme precipitation and storm surges could be integrated e.g., following the approach developed and published by us in Befort et al. (2015). **We have included the above information in Sect. 5 summary for clarity.**

Recently, Sajjad and Chan 2019 and Sajjad et al. 2020 proposed typhoon risk frameworks based on TC hazard (wind-based similar to yours), vulnerability, and disaster resilience, which provides a comprehensive information on TC risk. They found that the Pearl River delta region in Guangdong (area primarily mentioned in your case, Line) is a statistically significant hotspot of TC risk. How do you see the usefulness of your method for such frameworks? A thorough discussion regarding this is necessary

Similarly, most of the discussion in the manuscript revolves around TC-hazard and neglects the vulnerability and resilience within the regions where TCs are making landfalls. For instance, you say on Lines 236-238 that overall impacts of a storm is related to many factors such as size, duration, and intensity. However, the impacts are not only related to TC-associated factors but vulnerability and resilience are also integral parts of overall impacts and risks associated with TCs, as discussed in Sajjad and Chan 2019 and Sajjad et al. 2020. How do you incorporate these characteristics within the TC risk discussion of yours?

We think Reviewer 1 may have misread the purpose of our study because we do not share the same definition of the term "risk". The term "risk" in this study refers to the possibility of occurrence of an event (e.g. Vickery et al., 2000; Emanuel et al., 2006), as outlined in the title. In the context of the papers cited by Reviewer 1, this manuscript focuses on the hazard component in the framework of a classical impact modelling and assessment approach. Please note that in terms of an impact modelling approach, the components would be the hazard (and its risk of occurrence), the vulnerability (which would include measures of resilience), and the exposure of values at risk. **We have replaced the term "risk" by "hazard" in relevant places in the manuscript to make this point clearer**. To avoid confusion, hereinafter, we use the term "hazard" to represent the possibility of occurrence of a natural event; and the term "risk" to reflect a systemic integrated perspective of resultant impacts by the combination of information about the hazard, the vulnerability, and the exposure. This terminology would be more similar to the approach used in Sajjad and Chan (2019) and Sajjad et al. (2020), and will also account for the common practice in industry applications (e.g., CAT models).

Additionally, you need to detail the current limitations of your method. For example, how well this method could perform at higher resolution assessments, which are more important for policy and decision-making in the context of DRR efforts? What are the future prospects of your study?

Many thanks for this comment. In comparison to other methods to generate large TC event sets, our specific approach is **limited** mainly by the source of data used. The current TC event set constructed on synoptic scale forecasts archived in TIGGE, is strictly spoken representative only for the current climate state. Any longer-term climate variability (e.g. multi-decadal fluctuations like the PDO) and their impacts on any TC frequency-intensity distribution are not accounted for in this setting. Nevertheless, the presented approach would be equally applicable to data sets representing that kind of variability on longer time scales (e.g. decadal predictions or transient climate model simulations). Another limitation is obviously that we do not account for a direct assessment of the damage (loss) contribution of individual meteorological variables (e.g., precipitation leading to flooding, as mentioned above). **We added a specific section on limitations in Sect. 5 [Lines 441-447].**

It is not fully clear what Reviewer 1 refers to as "**higher resolution assessments**". The TIGGE archive provides forecast data on a spatial scale (~0.56° - 1.25°), which is not matched by any other data source of comparable length (equivalent to 10k years of observations of TCs).

Further, we intentionally linked our (forecast) model based assessment to in-situ point observations from stations: the ultimate downscaling test. As we have demonstrated in section 4.3, one of the potential applications of our event set is to improve the return-period/return level calculation of the wind hazard at the local scale. Wind speed values are used in practice to decide on payments out of e.g. parametric insurance products (Swiss Re, 2016). Consequently, reliable wind-based trigger points of typhoon parametric insurance can be determined. This will further improve the suitability and flexibility of parametric insurance for DRR applications. Ultimately, this will improve the speed of post-disaster recovery. **We added a respective paragraph to the discussion [Lines 360-364, Lines 469-477]**.

With regard to **future prospects** of this study, we discuss this in Section 5. In particular, unlike event sets generated from a stochastic approach, the TC-associated precipitation field is simulated directly by the model. This means a more complex compound TC hazard assessment, as mentioned above (limitations), can be done as well in principle. The event set that we have constructed contains all necessary information for applications in the DRR context. Once robust trigger points for the local hazard are available (including their uncertainty), the targeted application of parametric products in disaster relief application is possible. Especially, when it comes to the evaluation of the basis risk (the risk not covered by payments out of a parametric product). This study is merely the first step toward a statistically robust, full physical model based TC hazard assessment. **We added a respective paragraph to Sect. 5 [Lines 469-477]**.

Specific Comments

Lines 10-13: *The sentence is long and it is difficult to follow. Would be better to break it into two sentences, if possible.*

We thank Reviewer 1 for pointing this out, **we have modified this sentence accordingly**.

Line 22: *There is no such thing as "natural disaster" but only natural hazards. Disasters always involve human agency. Therefore, please avoid using this term and check the manuscript thoroughly for this issue. For further details, you are encouraged to see https://www.undp.org/content/undp/en/home/blog/2017/5/18/Natural-disasters-don-t-exist-but-natural-hazards-do.html*

Many thanks for pointing this out. Although we agree in principle that the disaster aspect includes a men-made perspective in its impact on human influenced structures the personal position expressed in this blog is document of a narrow understanding and perspective of nature. A natural hazard can be a disaster for the environment, even without human influences. Following good scientific practice, we prefer to use peer-reviewed literature for scientific studies and not personal comments. We used the phrase "natural disaster" as a generic term to indicate a natural event with sudden, large negative economic or environmental losses. The definition of "disaster" that we used here is similar to the definition stated in the IFRC webpage (see https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster/). We are not trying to argue whether "natural disaster" exists or not as this is beyond the scope of this study. However, the use of this phrase is in line with literature, e.g. Cavallo and Noy (2011), Smith and Matthews (2015), Ye et al. (2016), Bakkensen et al. (2018).

Nevertheless, to avoid any confusion and as we are in general agreement with the reviewer **we rephrased line 22 to: " … lead to an increase of risk to humans and for economic loss potentials from natural hazards e.g., tropical cyclones, with potentially disastrous consequences." We also checked the whole manuscript and corrected for a more precise use of the terminology in the revised manuscript.**

Line 34: *you mean "livestock"?*

We thank Reviewer 1 for pointing this out and the reviewer is correct. **We have corrected this**.

Lines 145-149: *It is mentioned that VIF is used to resolve the issue of collinearity and 17 variables are selected to construct the final LRC model. How many total variables were included initially? Are the VIF values for all of these 17 variables less than the normal threshold (i.e., VIF ≤ 7.5)? It would be useful to add the VIF values of the final variables in Table 3.*

The list of variable initially used is presented in Table 2. We have modified the caption of Table 2 for clarification.
Yes, those 17 variables stated in Table 3 have VIF value of less than 5. We are not sure whether including the VIF values of the final variables would be useful for this manuscript. However, Reviewer 1 pointed out that we did not state the criteria which we used for the variable selection. **We have included the criteria (i.e. VIF < 5) for the variable selection in the manuscript. Line 176-177:** "*Variables with VIF value larger than 5 are excluded*."

Lastly*, a thorough intermediate level editing is recommended to remove "several" grammatical and language errors throughout the manuscript.*

We thank Reviewer 1's advice, and **we have edited the manuscript accordingly**.

References

Bakkensen, L. A., Shi, X., and Zurita, B. D.: The Impact of Disaster Data on Estimating Damage Determinants and Climate Costs, Economics of Disasters and Climate Change, 2, 49-71, 10.1007/s41885-017-0018-x, 2018.

Befort, D. J., Fischer, M., Leckebusch, G. C., Ulbrich, U., Ganske, A., Rosenhagen, G., and Heinrich, H.: Identification of storm surge events over the German Bight from atmospheric reanalysis and climate model data, Natural Hazards and Earth System Sciences, 15, 1437, 2015.

Befort, D. J., Wild, S., Kruschke, T., Ulbrich, U., and Leckebusch, G. C.: Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses, Atmos Sci Lett, 17, 586-595, 10.1002/asl.694, 2016.
Befort, D. J., Wild, S., Knight, J. R., Lockwood, J. F., Thornton, H. E., Hermanson, L., Bett, P. E., Weisheimer, A., and Leckebusch, G. C.: Seasonal forecast skill for extratropical cyclones and windstorms, Q J Roy Meteor Soc, 145, 92-104, 10.1002/qj.3406, 2019.
Befort, D. J., Kruschke, T., and Leckebusch, G. C.: Objective identification of potentially damaging tropical cyclones over the Western North Pacific, Environmental Research Communications, 2, 031005, 10.1088/2515-7620/ab7b35, 2020.
Cavallo, E., and Noy, I.: Natural Disasters and the Economy — A Survey, International Review of Environmental and Resource Economics, 5, 63-102, 10.1561/101.00000039, 2011.
Emanuel, K., Ravela, S., Vivant, E., and Risi, C.: A statistical deterministic approach to hurricane risk assessment, B Am Meteorol Soc, 87, 299-314, 10.1175/Bams-87-3-299, 2006.
Kruschke, T.: Winter wind storms: Identification, verification of decadal predictions, and regionalization, Doktors der Naturwissenschaften, Institut f• ur Meteorologie, Freie Universit• at Berlin, 181 pp., 2015.
Leckebusch, G. C., Renggli, D., and Ulbrich, U.: Development and Application of an Objective Storm Severity Measure for the Northeast Atlantic Region, Meteorologische Zeitschrift, 17, 575-587, 10.1127/0941-2948/2008/0323, 2008.
Nissen, K. M., Ulbrich, U., Leckebusch, G. C., and Kuhnel, I.: Decadal windstorm activity in the North Atlantic-European sector and its relationship to the meridional overturning circulation in an ensemble of simulations with a coupled climate model, Climate Dynamics, 43, 1545-1555, 10.1007/s00382-013-1975-6, 2014.
Osinski, R., Lorenz, P., Kruschke, T., Voigt, M., Ulbrich, U., Leckebusch, G. C., Faust, E., Hofherr, T., and Majewski, D.: An approach to build an event set of European windstorms based on ECMWF EPS, Nat. Hazards Earth Syst. Sci., 16, 255-268, 10.5194/nhess-16-255-2016, 2016.
Sajjad, M., and Chan, J. C. L.: Risk assessment for the sustainability of coastal communities: A preliminary study, Science of The Total Environment, 671, 339-350, https://doi.org/10.1016/j.scitotenv.2019.03.326, 2019.
Sajjad, M., Chan, J. C. L., and Kanwal, S.: Integrating spatial statistics tools for coastal risk management: A case-study of typhoon risk in mainland China, Ocean & Coastal Management, 184, 105018, https://doi.org/10.1016/j.ocecoaman.2019.105018, 2020.
Smith, A. B., and Matthews, J. L.: Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates, Nat Hazards, 77, 1829-1851, 10.1007/s11069-015-1678-x, 2015.
Swiss Re: Natural catastrophes and man-made disasters in 2015: Asia suffers substantial losses. https://reliefweb.int/sites/reliefweb.int/files/resources/sigma1_2016_en.pdf, 2016.
Vickery, P. J., Skerlj, P. F., and Twisdale, L. A.: Simulation of Hurricane Risk in the U.S. Using Empirical Track Model, Journal of Structural Engineering, 126, 1222-1237, 10.1061/(ASCE)0733-9445(2000)126:10(1222), 2000.
Walz, M. A., and Leckebusch, G. C.: Loss potentials based on an ensemble forecast: How likely are winter windstorm losses similar to 1990?, Atmos Sci Lett, 20, e891, 10.1002/asl.891, 2019.
Ye, T., Wang, Y., Wu, B., Shi, P., Wang, M., and Hu, X.: Government Investment in Disaster Risk Reduction Based on a Probabilistic Risk Model: A Case Study of Typhoon Disasters in Shenzhen, China, International Journal of Disaster Risk Science, 7, 123-137, 10.1007/s13753-016-0092-7, 2016.

Response to Reviewer 2's comments

We thank Reviewer 2 for the time to go through our manuscript in details. This manuscript describes a new and efficient method to produce a physical TC event set in the western North Pacific basin. In general, reviewers think after careful revision, the results of this study is of great interest and relevance, and it will be a useful contribution to the field of TC risk assessment. Here is our point-to-point response to Reviewer 2's comments. Reviewer 2's comments are in red.

General comments:

This manuscript describes a new and efficient method to produce TC event set in the Pacific basin. The TC events are detected from an ensemble data archive TIGGE, using an objective impact-oriented windstorm identification algorithm WiTRACK. This dataset contributes to existing synthetic datasets (mostly statistical basin-wide methods) as in this dataset the TCs are detected in GCMs, so that the complex physical processes of TCs are captured and hence TCs are physically realistic. More extreme TCs are found in the dataset and these data will help overcome the shortage in observational record. Overall, I think the results will be a nice contribution to the field of TC risk assessment. However, I have some basic questions about the methodology, the utility of certain results and I recommend major revisions prior to publication. Also, I would recommend careful editing of the manuscript. There are many terminologies in the manuscript and please make sure it is easy for readers to follow.

We thank the reviewer advice and we have carefully reviewed and edited the manuscript.

Major comments:

1) L191: The detection rates of historical TCs are reported here. However, will the detection algorithm produce more TCs? What fraction of TCs that the detection algorithm produce is real historical TCs?

We thank the reviewer for pointing this out and highlighting a section for which we can improve clarity. Strictly spoken, the detection algorithm we apply in this study (developed for TC detection in the West-Pacific in Befort et al., 2020) does not produce TCs, but enables us to detect them automatically in the large data set. As the data set we use is the output of operational NWP's forecast models, in a very narrow sense , none of the TCs that we detected is a one-to-one equivalent to a real historical TCs. However, events which satisfy the criteria in the MPES TC identifier (MTI, Section 3.2.3) (i.e. MEPS TC events) can be considered as events which are similar to the historical event. The percentage of TCs which are in this sense similar to real events that occurred in the TIGGE is ~60%. Thus, about 40% are pure ensemble predicted events that did not realise in the observed nature or do not have a very similar twin at the same time at the same location.

We have modified the text in L191 and the caption of Table 4 for more accurate description. Lines 225-228, "*A historical TC is said to be detected in a forecast model if there exists a TC counterpart in the forecast model, which is similar to the historical TC as identified* *by the MTI (Section 3.2.3). The detection rates of historical TCs which are detected in different forecast outputs, i.e. CMA, ECMWF, JMA, and NCEP, are 91.2%, 94.7%, 89.4%, and 90.7%, respectively, ...*"

2) The authors mention that one benefit of this dataset is "The TPEPS event set includes events which are unlikely but physically possible. This provides an important and unique advantage for typhoon risk assessment." Combined with Fig. 2, TC tracks in the detected dataset is very different with observations, and TPEPS tracks appear in locations with no historical tracks. If there is no historical track in some regions, are they supposed to be no storms or there can be storms but no storm has appeared in historical records due to the low probability? This needs to be explained.

We thank Reviewer 2 for pointing out this important issue. If there is no historical track in some regions, this does not mean storms cannot occur in those regions. The fact that we have not seen a TC during the time period of known observational records in those regions could be due to the observation period is too short and the sample size is not large enough to fully represent the distribution of the underlying basic population (i.e. all possible TCs in the given climate state). For example, if we follow the necessary but insufficient conditions of TC formation which are identified by Gray (1977) from historical observations, TC formation occurs away from the equator (> 5 deg). However, Tropical Storm Vamei (2001) formed close to the equator (~1.4 deg N). This shows storm can appear in the historically "storm-free" region.

Furthermore from the statistical perspective, we can view the JRA-55 event set as a subset which is randomly selected from the TPEPS event set. This means if we randomly sample the TPEPS event set, we can obtain a subset highly similar to the JRA-55 event set. For demonstration, we have conducted bootstrap resampling on the TPEPS event set to obtain 10,000 sets of subsample. Each set of subsamples has 668 events to mimic the number of events in the JRA-55 event set. For each set of subsamples, the track density is calculated, and used to calculate uncentred pattern correlation between the resampling set of subsamples and the JRA-55 event set. In order to focus on relevant entries, for a particular grid box, if the values of track density for a resampling set and the JRA-55 event set are both less than one, such grid box is not used in the pattern correlation calculation. The mean, standard deviation, minimum, and maximum of the uncentred pattern correlation of the 10,000 set of subsamples are 0.9380, 0.0107, 0.8961, and 0.9697, respectively. This suggests the spatial pattern of the JRA-55 event set is highly similar to a small random subset of the TPEPS event set. Consequently, the JRA-55 event set can be seen as a subset randomly selected from the TPEPS event set. On the other hand, it is not be possible to deduce the basic population (e.g. the TPEPS event set) from a small sample set (e.g. the JRA-55 event set). Although the spatial distribution of the small set sample is similar to the subsamples of the basic population and thus usable as one possible realisation of the basic population, the small sample set does not contain all of the information of the underlying population. Furthermore, the statistical estimate of extremes would also be different for the small sample set (e.g. JRA-55 event set) and the basic population (e.g. TPEPS event set).

We have included the above explanation in the revised manuscript (Lines 259-280).

3) The sensitivity and performance of four ensemble data archive are not well described. For example, in some dataset, the storms are much weaker than historical storms. And some models have biases in simulating extratropical cyclone transition. More explanations and descriptions of the data archive needs to be added. Also, how these biases would have an impact on the detection algorithm?

The four data sets selected from the TIGGE archive are the state-of-the-art NWP models as used by four leading synoptic weather forecast centres worldwide. Although a full assessment of their respective models' skill and potential biases is not in the scope of this study, **we added a section with information on the general performance of these four selected NWP models (Lines 105-115)**. For the dedicated purpose of this study, the reviewer is fully correct and we need to check for biases in the underlying climatological features as provided by a time- and ensemble-aggregated view of the data set (a task normally not necessarily done in forecast model evaluation departments for all levels of severe and rare extremes). This evaluation for extreme TC occurrence is what we did in section 4.1, showing respective results in Fig.1-7. **We included a paragraph to clarify which part of the study is model validation and which part is event set building (Lines 223-225)**.

TIGGE data's main difference to the operationally used NWP output is that TIGGE did archive a lower resolution. Nevertheless, all underlying processes and feedbacks are captured in the originally resolution of the NWP products and are thus fully included. Thus, we would expect the best possible representation of dynamical processes in those forecast simulations than compared to lower resolution AOGCM simulations, e.g. for transient climate experiments. Beyond this, model resolution is known to be a limiting fact of simulating TC intensity (Bengtsson et al., 2007). One of the advantages of using WiTRACK is that it does not use raw wind speeds, instead, it uses the 98^{th} percentile relative exceedance for tracking. This means that even if the simulation wind speed of TC is systematically weaker than in historical observations, the 98^{th} percentile climatological wind should also be lower than the actual 98^{th} percentile climatological wind speed, a TC will still be tracked as long as there exists a 98^{th} percentile exceedance wind cluster. It can be shown that, within the study area, the 98^{th} percentile relative exceedance of the 4 models, which we used to construct the TIGGE event set, have similar behaviour (i.e. similar to Figure 2 of Osinski et al. (2016)). Befort et al. (2020) showed the applicability of such an approach to relate information from observations (i.e. IBTrACS data) to automatically detected TCs from a much coarser resolution reanalysis product (JRA-55). Consequently, a bias due to resolution does not have significant impact on WiTRACK as the tracking algorithm serves as a bias correction in this sense (detailed discussion on the impact of weaker wind speed in model outputs on WiTRACK can be found in Osinski et al. (2016)). **We included a paragraph to discuss this in more detail (Lines 125-127; 140-152)**.

4) The authors have compared the TIGGE PEPS TCs with JRA-55 in terms of track density, landfall frequency, etc. How about other characteristics? For example, landfall intensity along coastline?

The distribution of landfall intensity (wind speed in m/s) for TC, which made landfall with at least typhoon strength, are very similar for the JRA-55 event set and TPEPS event set. The table below shows some of the statistics of these two distributions. The two-sample

Kolmogorov-Smirnov test show these two distributions belong to the same distribution significant at the 0.05 significance level.

	JRA-55	TPEPS
Mean	23.5899	23.4044
Standard deviation	3.44527	3.84537
Median	22.58	22.2
Number of events	184	23343

We have included the above discussion into revised manuscript [Lines 345-348].

5) Fig. 7 shows the difference between TIGGE PEPS event set and observation. In the text, you have mentioned possible reasons for these differences. Is there possible way to reduce these differences, for example in the detection algorithm, to also remove low-impact storms? Also, you mentioned the ESSI, is there a way to quantify this index?

Figure 7 shows some of the differences between the TPEPS event set and the JRA-55 event set. These differences are mainly due to the finite simulation time in forecast models. Some of these differences could be reduced based on additional assumptions that would depend on the specific application of the users. A more detailed analysis of the performance with respect to a data set not affected by a finite simulation time is a reanalysis product (e.g. JRA-55). We showed in Befort et al. (2020) in JRA-55 that our tracking already focusses on the most severe part of the TC severity distribution and thus does show some expected differences to e.g. IBTrACS data.

We apologise we did not include the text associated with the SSI (and ESSI). Leckebusch et al. (2008) introduced this objective severity measure for gridded datasets of extreme storms in the North-Atlantic and the method was applied for TCs in the North-West Pacific in Befort et al. (2020). **The relevant text is included in revised manuscript [Lines 215-219]**.

Minor comments:

L41-42: more recent papers should be added. Such as the following two recent models:
- Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2018: An environmentally forced tropical cyclone hazard model. Journal of Advances in Modeling Earth Systems, 10 (1), 223–241.
- Jing, R., and N. Lin, 2020: An environment-dependent probabilistic tropical cyclone model. Journal of Advances in Modeling Earth Systems, 12 (3), e2019MS001 975.

We thank the reviewer's suggestions. **We have included these references in the revised manuscript.**

L45: I didn't understand the sentence 'the typhoon event set might not be physically consistent'. What is 'physically' consistent?

It means event sets created by stochastic perturbations will create TC events that (with respect to their inner dynamical structure) are not necessarily physically consistent anymore. As just surface footprints are stochastically modelled from existing tracks, there is no check whether those events (in the stochastically modelled from) are physically possible and how they could be realised in a fully dynamical consistent view, thus fulfilling all known physical relations and derived constraints by the means of physical laws. Consequently, the amount of unrealistic physical properties due to the oversimplified stochastic simulation is unknown and laws of physical interactions are potentially ignored. **We have modified the sentence in the revised manuscript to clarify this point [see lines 45-51]**.

L79: "The domain of this study covers the Western North Pacific (WNP), east and south-east Asia spanning from 85 E to 195E and 15 S to 75 N." Why data around equator is also used? There is no TCs forming around equator.

We thank Reviewer 2 for pointing this out and we apologise for the confusion. The domain stated in the manuscript is part of the parameter set up for WiTRACK. However, the true domain which is used for tracking is 90-180° E, and 0-70° N and **we have made this correction in the revised manuscript.**

We included regions close to the equator although TCs rarely form around the equator, it is still possible for TCs to form close to the equator, for example Tropical storm Vamei (2001). Furthermore, while the core pressure centre of the TCs might be away from the equator, the damaging wind field, as identified by the 98th percentile relative exceedance, could be quite large, impacting potentially regions close to the equator.

L102-104: Is there a reason why an old version of IBTrACS is used?

IBTrACS v03r10 was the most up-to-date official version of IBTrACS when this study was first started. Furthermore, for our study period (with 6-hourly observations), the data in v04 and v03r10 are the same.

L152: "the accuracy of the LRC is about 90%" What is the fraction of TC (or positive samples?) Does there exist issue of imbalanced data?

No, the validation set is not imbalanced. In the validation set, 49 out of 96 tracks are TCs (~51% of the validation set). **We have included a more detailed description in the revised manuscript. Lines 180-181** "*Validation using JRA-55 event set (2015-2017), which has 49 TC events and 47 non-TC events…*"

L197: "Percentage of total TC windstorms as PEPS TCs can be treated as a proxy to quantify the forecast skill of the model." In Table 5, NCEP is almost twice of that in JMA, what does this percentage mean?

This indicates the NCEP model generates more “wrong” forecasts than JMA yet these wrong forecasts are physically possible. **We included a clarifying sentence to a respective possible interpretation at lines 235-238**: *“For example, NCEP has 47.1% of TC windstorms as PEPS TCs whereas JMA has 26.5%. This indicates the NCEP model generates more “wrong” forecast than JMA however these wrong forecasts are physically possible. Yet, examining the forecast skill of models is not the focus of this study and the rest of the discussion focuses on the TPEPS TC event set.*”

L203: do you mean Fig 3? Also, more explanations should be added in the text. I can’t understand this figure.

We thank Reviewer 2 for identifying this error. The reviewer is correct that we are referring to Fig 3. Fig. 3 shows the feature scaled times series of number of TCs which are first identified in each day from May to December. The core message of Fig 3 is that the temporal variability of the TPEPS event set and the JRA-55 event set are largely similar (except for the earlier years). **We have modified the text in the revised manuscript. Lines 239-240** “*Figures 2 and 3 show the spatial pattern and temporal variability of the number of TC which are first detected for each day, …*”

L203: In Fig. 2, all TPEPS are much more similar with each other, comparing with JRA-55. How to explain this?

The major difference between the track density of TPEPS and JRA-55 is that there is an eastward bias in the TPEPS. There are several reasons that could contribute to this. The eastward bias in the track density appears to be a common feature in many GCMs (e.g. Camargo et al., 2005; Bell et al., 2013; Roberts et al., 2020), this has also been observed in seasonal forecast output (Camp et al., 2015). Finite simulation time has also contributed to this bias as TC that forms in the region east of 150 °E would not have the time to move into the western part of WNP before the end of simulation time. Differences in the amount of tracks could also contribute to the differences as more diverse tracks would be captured. **We have added a respective explanatory comment at lines 252-258**.

Fig4: The tracks in black are very easily messed up with the map. Probably change the color of coastline.

We thank Reviewer 2 for pointing this out. **We have changed the colour of the plot**.

Fig5: The y-axis is not clear to me, please add more explanation.

Fig 5 shows the climatological seasonal cycle of TC activity for the TPEPS TC event set and the JRA-55 event set. The daily number distribution, p_i, is calculated as follows:

$$p_i = \frac{n_i}{\sum_i n_i} \times 100\%$$

where n_i is the number of TC first detected on day i for the individual event set. As such, it is the probability of TC being first detected at a given day. **We have added more explanation in the caption of Figure 5.**

Fig8: The colored dots for single center are too light to see. If this figure is to show distribution, I would recommend not using same color bar for single model and for TIGGE total.

We have changed the colour scale of this figure in the revised manuscript.

Fig9: It's hard to see the distributions are in good agreement, probably can change to annual frequency instead of total number of landfall events. Also, the correlation coefficients could be used to show the landfall frequency in all TIGGE dataset is positively correlated with JRA-55.

We thank Reviewer 2 for these suggestions. **We have included pattern correlation between the spatial distribution for JRA-55 and TPEPS event sets in the revised manuscript. Line 339** "…*with uncentred pattern correlation of 0.8345*."

Fig12: I can see your points in showing the grey dashed lines. But the lower bound curves cannot show the trend properly. I would recommend add 75% or 80% confidence interval to show that the trends are same, but TIGGE PEPS event set has much narrower bounds.

We are not certain what Reviewer 2 refers to as the trend of the lower bound curves. There are two separate factors that determine the "shape" of the curve of the lower and upper bound of uncertainty. First, the return level-return period estimate has asymptotic behaviour. This means the return level estimate approaches to a certain value as the return period increases. Second, the uncertainty of the estimation increases with increasing return period. Combining these two factors we can see that the so-called "trend" in the lower bound grey curve does not exist. To show the 95% confidence interval reflects a typical setting for assessing statistical estimates uncertainty for GPD fitted return-level plots and the authors would prefer to stay with this representation.

<u>References</u>

Bell, R., Strachan, J., Vidale, P. L., Hodges, K., and Roberts, M.: Response of Tropical Cyclones to Idealized Climate Change Experiments in a Global High-Resolution Coupled General Circulation Model, J Climate, 26, 7966-7980, 10.1175/JCLI-D-12-00749.1, 2013.

Bengtsson, L., Hodges, K. I., and Esch, M.: Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses, Tellus A, 59, 396-416, 2007.

Camargo, S. J., Barnston, A. G., and Zebiak, S. E.: A statistical assessment of tropical cyclone activity in atmospheric general circulation models, Tellus A, 57, 589-604, 10.1111/j.1600-0870.2005.00117.x, 2005.

Camp, J., Roberts, M., MacLachlan, C., Wallace, E., Hermanson, L., Brookshaw, A., Arribas, A., and Scaife, A. A.: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system, Q J Roy Meteor Soc, 141, 2206-2219, 10.1002/qj.2516, 2015.

Gray, W. M.: Tropical Cyclone Genesis in the Western North Pacific, J Meteorol Soc Jpn, 55, 465-482, 1977.

Osinski, R., Lorenz, P., Kruschke, T., Voigt, M., Ulbrich, U., Leckebusch, G. C., Faust, E., Hofherr, T., and Majewski, D.: An approach to build an event set of European windstorms based on ECMWF EPS, Nat. Hazards Earth Syst. Sci., 16, 255-268, 10.5194/nhess-16-255-2016, 2016.

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L.-P., Chauvin, F., Terray, L., Valcke, S., Moine, M.-P., Putrasahan, D., Roberts, C., Senan, R., Zarzycki, C., and Ullrich, P.: Impact of Model Resolution on Tropical Cyclone Simulation Using the HighResMIP–PRIMAVERA Multimodel Ensemble, J Climate, 33, 2557-2583, 10.1175/JCLI-D-19-0639.1, 2020.

Response to Reviewer 3's comments

We thank Reviewer 3 for the time to go through our manuscript in details. This manuscript describes a new and efficient method to produce a physical TC event set in the western North Pacific basin. In general, reviewers think after careful revision, the results of this study is of great interest and relevance, and it will be a nice contribution to the field of TC risk assessment. Here is our point-to-point response to Reviewer 3's comments. Reviewer 3's comments are in red.

General Comment

I think that the topic of this study is of great interest and relevance, and that it is suitable to NHESS. Besides, the paper is generally well written, the methodology clearly illustrated and the results well presented and discussed. However, there are also a few (minor) corrections and some improvements of the text that could be made to further improve the manuscript before to proceed with its publication.

*Therefore, my recommendation is to accept the manuscript for publication after **minor revisions**.*

Specific Remarks

1. *Page 1, line 14: "... characteristics of the new event set is consistent to the..." should read "... characteristics of the new event set are consistent to the..."*

We thank Reviewer 3 for pointing this out, **we have corrected this in the revised manuscript**.

2. *Page 1, line 24: "... 67.1 billion RMB ..." Many readers could be helped to understand the economic significance of this figure by accompanying it with the corresponding value in US Dollars or Euros.*

We thank Reviewer 3's suggestion, **we have added the corresponding value in Euros in the revised manuscript**.

3. Page 2, line 45: "... (ii) the storms in the typhoon event set might not be physically consistent." Please, clarify what do you exactly mean here with "physically consistent"?

It means event sets created by stochastic perturbations will create TC events that (with respect to their inner dynamical structure) are not necessarily physically consistent anymore. As just surface footprints are stochastically modelled from existing tracks, there is no check whether those events (in the stochastically modelled from) are physically possible and how they could be realised in a fully dynamical consistent view, thus fulfilling all known physical relations and derived constraints by the means of physical laws. Consequently, the amount of unrealistic physical properties due to the oversimplified stochastic simulation is unknown and laws of physical interactions are potentially ignored. **We have modified the sentence in the revised manuscript to clarify this point [see lines 45-51]**.

4. *Page 2, line 63–64: "In this study, we show the TPEPS event set has much higher information content: more TC events and more extremely high impact TC events." Higher and more than what?*

We thank Reviewer 3 for pointing out this point. **This sentence should read "*In this study, we show the TPEPS event set has much higher information content: more TC events and more extremely high impact TC events than historical or reanalysis-based TC event set.*" (Lines 69-70)**

5. *Page 4. Line 126: "WiTRACK identifies windstorm events of all kind, including MEPS TCs, PEPS TCs, MEPS extratropical cyclones." I suppose it identifies also PEPS extratropical cyclones.*

Yes, it does. **We have added PEPS extratropical cyclones to the list for clarification**.

6. *Page 4, line 175: "The removal of these events ensures the TPEPS event set is independent of any pre-existing weather patterns." The goal here is to build a large set of typhoon events in order to provide a solid statistical evaluation of their characteristics, so why is it so important that the considered TPEPS events are independent of any pre-existing weather patterns?*

To use this as an extension of event numbers and thus as an alternative reality, the inclusion of real existing events will incorporate some bias towards observed events as all of them will create a multiple realisation in the ensemble members started at the time such a real event occurred. By not considering those ensemble members, which are closely related to observed events, will secure that indeed new events are used to build the pure EPS event set. It has to be noted though that the inclusion of those events should not change the overall track distribution, or in other words, the track distribution from pure EPS and real EPS events is fairly similar.

7. *Page 6, line 193, Figure 1: please add the units to the colour bar.*

We have added the unit to colour bar.

8. *Page 6, line 197, Table 5: Why there is such a large difference in the number of simulated TC wind storms between the TIGGE models? Is this due to the different number of ensemble members of the EPSs? The large majority of the considered TPEPS are from two EPSs: the ECMWF and the NCEP. What consequences could this fact have on the analysis results?*

The main reasons for differences in the number of detected TC windstorms between TIGGE models are they have (1) different numbers of ensemble members of the EPSs, (2) different number of runs per day, and (3) different maximum forecast lead time (c.f. Table 1). Given the spatial and temporal distributions of the individual PEPS event sets are similar to each other, the analysis on the overall TPEPS event set is reliable.

9. Page 6, line 202: Fig. 1d, should read Fig. 2d.

We thank Reviewer 3 for pointing this out, **we have corrected this in the revised manuscript**.

10. Page 6, line 203: Fig. 2 should read Fig. 3.

We thank Reviewer 3 for pointing this out, **we have corrected this in the revised manuscript**.

11. Page 6–7, line 212–220: I'm not sure I fully understand the explanation the authors provide for the discrepancy between the spatial distribution of the TPEPS event set and JRA–55 events as shown in Figure 2 (panels c and f). The fact that the JRA-55 event set can be considered as a subset of the TIGGE event set does not explain the difference in spatial distribution. According to this view, in fact, the JRA-55 events can be seen as randomly selected from a larger set (the TIGGE set), and thus they should also be spatially distributed as this event set. Also, why the higher level of the 98th percentile values of the JRA-55 wind should explain the lower number of typhoons in this area?

We agree with Reviewer 3 that the JRA-55 event set can be seen as a subset randomly selected from a larger set (i.e. the TIGGE event set). This means if we randomly sample the TPEPS event set, we can obtain a subset highly similar to the JRA-55 event set. For demonstration, we have conducted bootstrap resampling on the TPEPS event set to obtain 10,000 sets of subsample. Each set of subsamples has 668 events to mimic the number of events in the JRA-55 event set. For each set of subsamples, the track density is calculated, and used to calculate uncentred pattern correlation between the resampling set of subsamples and the JRA-55 event set. In order to focus on relevant entries, for a particular grid box, if the values of track density for a resampling set and the JRA-55 event set are both less than one, such grid box is not used in the pattern correlation calculation. The mean, standard deviation, minimum, and maximum of the uncentred pattern correlation of the 10,000 set of subsamples are 0.9380, 0.0107, 0.8961, and 0.9697, respectively. This suggests the spatial pattern of the JRA-55 event set is highly similar to a small random subset of the TPEPS event set. Consequently, the JRA-55 event set can be seen as a subset randomly selected from the TPEPS event set. On the other hand, it is **not** be possible to deduce the basic population (e.g. the TPEPS event set) from a small sample set (e.g. the JRA-55 event set). Although the spatial distribution of the small set sample is similar to the subsamples of the basic population and thus usable as one possible realisation of the basic population, the small sample set does not contain all of the information of the underlying population. Furthermore, the statistical estimate of extremes would also be different for the small sample set (e.g. JRA-55 event set) and the basic population (e.g. TPEPS event set). **We have included the above explanation in the revised manuscript (Lines 259-280).**

Upon further investigation, we found that the 98th percentile is not the reason that leads to the differences in spatial distribution. The major difference between the track density of TPEPS and JRA-55 is that there is an eastward bias in the TPEPS. There are several reasons that could contribute to this. The eastward bias in the track density appears to be a common feature in many GCMs (e.g. Camargo et al., 2005; Bell et al., 2013; Roberts et al., 2020), this has also been observed in seasonal forecast output (Camp et al., 2015). Finite simulation time has also contributed to this bias as TC that forms in the region east of 150 °E would not have the time to move into the western part of WNP before the end of simulation time. Differences in the amount of tracks could also contribute to the differences as more diverse tracks would be captured. **We have added a respective explanatory comment at lines 252-258**.

12. Page 7, line 248–249: As formulated here, this sentence seems to imply that TCs with weaker winds are also less spatially extended, which is not true.

The impact area of a TC in this study refers to the total area which has experienced TC-associated extreme wind (i.e. larger than local climatological 98th percentile wind speed). Given the fact that the wind speed of TC wind field decays radially outward, TC with weaker winds would have a smaller impact area because the outer wind speed would be below the 98th local climatological wind percentile value. **We have added more descriptions about impact area in the revised manuscript to clarify this point [Lines 306-307]**.

13. Page 7, 252–255: "… impact (Befort et al., 2020). Many of the low impact TCs …" should probably read "… impact (Befort et al., 2020), many of the low impact TCs …".

We thank Reviewer 3 for pointing this out, **we have corrected this in the revised manuscript**.

14. Figure 8: In the text of the manuscript, there are references to panels labelled with letters (a, b, … f), but the panels in Figure 8 are not labelled.

We thank Reviewer 3 for spotting this error. **We have corrected this in the revised manuscript**.

15. Page 9, line 317: "… based on minimisation of the root-mean-square-error (RMSE) of …". Of what?

We thank Reviewer 3 for pointing this out, **we have corrected this in the revised manuscript**. This is the root-mean-square-error of the quantile mapping output.

References

Bell, R., Strachan, J., Vidale, P. L., Hodges, K., and Roberts, M.: Response of Tropical Cyclones to Idealized Climate Change Experiments in a Global High-Resolution Coupled General Circulation Model, J Climate, 26, 7966-7980, 10.1175/JCLI-D-12-00749.1, 2013.

Camargo, S. J., Barnston, A. G., and Zebiak, S. E.: A statistical assessment of tropical cyclone activity in atmospheric general circulation models, Tellus A, 57, 589-604, 10.1111/j.1600-0870.2005.00117.x, 2005.

Camp, J., Roberts, M., MacLachlan, C., Wallace, E., Hermanson, L., Brookshaw, A., Arribas, A., and Scaife, A. A.: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system, Q J Roy Meteor Soc, 141, 2206-2219, 10.1002/qj.2516, 2015.

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L.-P., Chauvin, F., Terray, L., Valcke, S., Moine, M.-P., Putrasahan, D., Roberts, C., Senan, R., Zarzycki, C., and Ullrich, P.: Impact of Model Resolution on Tropical Cyclone Simulation Using the HighResMIP–PRIMAVERA Multimodel Ensemble, J Climate, 33, 2557-2583, 10.1175/JCLI-D-19-0639.1, 2020.

A New View on Risk of Typhoon Occurrence in the Western North Pacific

Kelvin. S. Ng[1], Gregor. C. Leckebusch[1]

[1]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

Correspondence to: Kelvin S. Ng (k.s.ng@bham.ac.uk)

Abstract. To study high impact tropical cyclone (TC) is of crucial importance due to its extraordinary destruction potential that leads to major losses in many coastal areas in the Western North Pacific (WNP). Nevertheless, because of the rarity of high-impact TCs, it is difficult to construct a robust hazard~~risk~~ assessment based on the historical best track records. This paper aims to address this issue by introducing a computationally simple and efficient approach ~~,~~ to build a physically consistent, high impact TC event set with non-realised TC events in the THORPEX Interactive Grand Global Ensemble (TIGGE) archive. This event set contains more than 10,000 years of TC events. ~~using data from the THORPEX Interactive Grand Global Ensemble (TIGGE) archive with the application of impact-oriented tracking algorithm, to build a physically consistent high impact typhoon event set with non-realised TC events – data equivalent to more than 10,000 years of TC events.~~ The temporal and spatial characteristics of the new event set are~~is~~ consistent to the historical TC climatology in the WNP. It is shown that this TC event set contains ~100 and ~77 times more Very Severe Typhoons and Violent Typhoons than the historical records, respectively. Furthermore, this approach can be used to improve the return period estimation of TC-associated extreme wind. Consequently, a robust extreme TC hazard ~~risk~~ assessment, reflective of the current long-term climate variability phase, can be achieved using this approach.

1 Introduction

Increasing frequency and intensity of extreme meteorological events in the recent decades (IPCC, 2012) and increasing number of human population and assets located in risk-prone regions (Desai et al., 2015) lead to an increase of risk to humans and economic loss potentials ~~to human and economic~~ from natural hazards ~~disasterrs, for example~~e.g., tropical cyclones, with potentially disastrous consequences. For example, ~~i~~In the period ~~of~~ between 1st January and 18th October 2018, total typhoon-related ~~total~~ direct economic losses in ~~Western North Pacific (WNP)~~China is evaluated to exceed ~~more than~~ ~~up to~~ 67.~~1~~ billion RMB (roughly 8.3 billion Euros) (Chinese Meteorological Administration, CMA, 2018). While natural ~~disaster~~ hazards ~~has~~ impact ~~to~~ on all society stakeholders~~ of the society~~, governments are crucial in disaster risk reduction (DRR) because of their ability to implement necessary DRR-related policy and ability to allocate resources to appropriate parties (Shi, 2012). Governments have various options for DRR investments, for example, post-disaster relief and risk financing. Using cost-benefit analysis for a case study of typhoon disasters in China, Ye et al. (2016) showed insurance premium subsidies has the highest benefit-cost ratio. This is because premium subsidies increases penetration rate of an insurance program, i.e. more protection is offered by the private sector and the risk is transferred to the private sector (Glauber, 2004). Thus, development and application of effective financial instruments for risk transfer is important.

Other than classical (re-)insurance solutions, parametric insurance solutions have been developed for test cases in areas of corn yield (Sun et al., 2014) and live~~fe~~ stock (Ye et al., 2017) for Southeast Asia and China in recent years. Swiss Reinsurance Company Ltd. (Swiss Re) insured several municipal governments in Guangdong Province, China, through parametric insurance solution (Lemcke, 2017). Parametric insurance requires no physical damage assessment after an event. As soon as a certain threshold (i.e. trigger point) is exceeded, the insured party receives the agreed compensation from the insurer. Thus it has low administrative cost and quick disbursement. However, it is a challenge to determine a robust trigger point. It is because it would require a reliable typhoon hazard~~risk~~ assessment for the region of interest. A current common approach is to generate a large typhoon event set (e.g. equivalent to 7,000 years of real world data) based on historical track data using stochastic approach (e.g. Vickery et al., 2000; Emanuel, 2006; Emanuel et al., 2006; Rumpf et al., 2007, 2009; Lee et al., 2018; Jing and Lin, 2020). There are two potential downsides with the stochastic approach: (i) such typhoon event set would be biased toward the past events, and the frequency-intensity distribution of the event set might not be the same as the underlying frequency-intensity distribution, and~~;~~ (ii) the storms in the typhoon event set which are created by stochastic approach are not necessarily physically consistent. As just surface footprints are stochastically modelled from existing tracks, there is no check whether those stochastically modelled events are physically possible and how they could be realised in a fully dynamical consistent view, i.e. fulfilling all known physical relations and derived constraints by the means of physical laws. Consequently, the amount of unrealistic physical properties due to the oversimplified stochastic simulation is unknown and laws of physical interactions are potentially ignored.

~~might not be physically consistent (i.e. with unrealistic physical properties) because they are generated using oversimplified models and complex physical interactions are potentially ignored.~~ Consequently, the trigger point derived from the common approach may not be optimal. This means insurees could be either over- or under-compensated by the insurer.

A method to increase number of extreme weather events is to make use of ensemble prediction system (EPS). Osinski et al. (2016) used European Centre for Medium-Range Weather Forecasts (ECMWF) EPS to build an event set of European windstorms. Osinski et al. (2016) pointed out there are two types of storm events produced by EPS: (i) modified EPS storm (MEPS), and (ii) pure EPS storm (PEPS). MEPSs are storms with modifications in the EPS which have real-world counterpart. PEPSs are storms in the EPS which have no real-world counterpart, i.e. unrealised. PEPSs are independent events and the number of PEPSs increases as the lead time increase until the model has no memory of the initial conditions. Thus one can form an event set of extreme weather event by using TC related PEPSs. Osinski et al. (2016) demonstrated that reliable statistics of storms under the observed climate conditions can be produced based on EPS forecasts.

Building upon the results of Osinski et al. (2016), a new approach to construct a large data volume, physically consistent TC event set is presented in this study. This event set is constructed by applying an impact-oriented windstorm tracking algorithm (WiTRACK; e.g. Leckebusch et al., 2008) ~~(WiTRACK; e.g. Leckebusch et al. 2008)~~ to a multi-model global operational ensemble forecast data archive, t~~T~~he THORPEX Interactive Grand Global Ensemble (TIGGE) (Bougeault et al., 2010; Swinbank et al., 2015). The data volume of TIGGE is about 40,000 to 50,000 years. The event set consists of all non-realised TC events which were forecasted by EPS of different centres, this event set is referred to as the TIGGE PEPS (TPEPS) event set. In this study, we show the

TPEPS event set has much higher information content: more TC events and more extremely high impact TC events than historical or reanalysis-based TC event set. The TPEPS event set can be used to produce a robust TC hazard~~risk~~ assessment and to determine a robust trigger point for parametric typhoon insurance.

In this paper, we first present a computationally simple, inexpensive and efficient method to construct a physically consistent, high information content TC event set using only the 6-hourly surface wind speed field of EPS forecast model outputs. Then we analyse the characteristics of the TPEPS event set. Validation of the new method is done by comparing with the event set which is constructed using reanalysis data. The added values of this new approach are also discussed and presented. The paper is organised as follows: data sets which are used in this study are described in Section 2. Section 3 outlined the method that has been used to construct the TPEPS event set. Results and discussions including validation and investigate the characteristic of the TPEPS event set are presented in Section 4. A summary and conclusions can be found in Section 5.

2 Data

6-hourly instantaneous 10-m wind speed data in different data archives mentioned below are used in this study because it is highly related to TC wind damages. The domain of this study covers the Western North Pacific (WNP), east and south-east Asia spanning from 90~~85~~° E to 180~~195~~° E and 0~~15~~° N~~S~~ to 70~~5~~° N. The Japanese 55-year Reanalysis (JRA-55) (Kobayashi et al., 2015) from 1979 until 2017 (resolution of 1.25°×1.25°) is used for validation of the TPEPS event set. JRA-55 (1979-2014) is also used in parameter selection in TC identification algorithm, construction of Logistic Regression Classifier (LRC) (Sect. 3.2.2), and the data in 2015-2017 are~~is~~ used for validation of LRC. ERA-Interim (ERA-I) (Dee et al., 2011) is also used in the construction of LRC.

The TIGGE data archive (Bougeault et al., 2010; Swinbank et al., 2015) is used in the construction of the PEPS TC event set. The TIGGE data archive has been used extensively in the study of TC activity forecast (e.g. Vitart et al., 2012; Belanger et al., 2012; Halperin et al., 2013; Majumdar and Torn, 2014; Leonardo and Colle, 2017; Luitel et al., 2018). ~~–~~TIGGE data archive consists of ~8-15-day ensemble forecast data from 10 numerical weather prediction centres with about 11-50 members each. In this study, only perturbed forecast outputs of EPS from selected centres are used and they are ~~Chinese Meteorological Administration (~~CMA~~)~~, ~~European Centre for Medium-Range Weather Forecasts (~~ECMWF~~)~~, Japanese Meteorological Agency (JMA), and National Centers for Environmental Prediction ~~(National Centers for Environmental Prediction (~~NCEP~~))~~ (cf. Table 1). These four data sets are chosen because they are the state-of-the-art NWP models, which is used by four leading synoptic weather forecast centres, and they are the most complete dataset in the archive for the study period 2008-2017. Model configurations and model updates are documented online at https://confluence.ecmwf.int/display/TIGGE/Models. ECMWF EPS is a variable resolution EPS, i.e. days 1-10 were run at a higher resolution than days 11-15. For computational efficiency, ECMWF EPS outputs are regridded into a lower resolution grid of 0.5625° × 0.5625°. The resolution of the selected data sets ranges from 0.5625°×0.5625° to 1.25°×1.25°. Forecast lead time of each forecast outputs ranges from 216 to 384 hours. Only forecast outputs, which are initialised during the main typhoon season, i.e. 15 May-30 November, are considered. The resultant TPEPS TC event set has data equivalent to more than 10,000 years of TC model data of the current climate state.

Many studies have evaluated the performance of these EPSs in forecasting TC activities in various ocean basins. In general, EPSs underestimate TC intensity especially for coarse resolution models (Hamill et al., 2010; Magnusson et al., 2014). TC track and genesis forecast error exists in EPS and these errors increase as lead time increases (Buckingham et al., 2010; Yamaguchi et al., 2015; Zhang et al., 2015; Xu et al., 2016). While ECMWF EPS forecast would occasionally have abnormal TC track forecast errors (i.e. track forecast error that is extremely large) and might struggled with developing a warm core in the short range forecast (Majumdar and Torn, 2014; Xu et al., 2016), ECMWF EPS appears to have better performance in TC track forecast than other EPSs (Yamaguchi et al., 2015; Zhang et al., 2015; Xu et al., 2016). Yet, a full assessment of the respective skill in models is not in the scope of this study. For the dedicated purpose of this study, an examination for biases in the underlying climatological features as provided by a time- and ensemble-aggregated view of the data set is presented in Sect. 4.1.

The International Best Track Archive for Climate Stewardship (IBTrACS) v03r10 (Knapp et al., 2010) is used for validation and identification of TC events in reanalysis and TIGGE data archive. It contains all of the available best track records from different centres around the globe up to year 2017. Since only part of the best track records of year 2017 are archived in this version of IBTrACS, best track data from Joint Typhoon Warning Centre (JTWC) is used for year 2017.

3 Methods

3.1 Identification and characterisation of TC~~typhoon~~-related windstorms

For identification and characterisation of ~~typhoon~~TC-related windstorms, an impact-oriented tracking algorithm is used – WiTRACK (Leckebusch et al., 2008; Kruschke, 2015). Befort et al. (2020)~~-~~ adapted the algorithm to TCs~~tropical cyclones~~ and showed WiTRACK is well capable to identify high impact TC~~typhoon~~ events in WNP, using coarse resolution reanalysis product, with~~in~~ comparable quality to more data intensive algorithms. A brief description of the general procedure to track a windstorm using WiTRACK is as follows: (i) clusters with wind speed above the local threshold are identified for each of the 6-hourly time step of the input dataset,~~;~~ (ii) clusters with size smaller than a predefined threshold (*minarea*) are excluded,~~;~~ (iii) clusters identified in each 6-hourly time step are connected to a track using a nearest-neighbour criterion with consideration of the size of the cluster, and~~;~~ (iv) events with lifetime less than 8 6-hourly time steps are removed. Majority of the settings of WiTRACK are identical to Befort et al. (2020), including the use of local 98th percentile wind speed as local wind threshold, except in this study *minarea* is chosen to be 15,000 km^2. The 98th percentile wind speed is chosen because over 90% of loss events with losses above 3,000 million RMB can be identified by WiTRACK as demonstrated by Befort et al. (2020). The value for *minarea* is chosen based on a series of sensitivity studies for parameter selection. The output of WiTRACK contains information about the characteristics of all identified windstorm events, including size of the windstorm at any given 6-hourly time step, the overall footprint of extreme wind associated with the windstorm events, and storm severity index (SSI; Leckebusch et al., 2008). These information are used in the identification of TC~~-typhoon~~-related pure EPS windstorm events (Sect. 3.2). As discussed in Sect. 2, TC intensity is generally underestimated by EPS and model resolution is known to be a limiting factor (Bengtsson et al., 2007; Hamill et al., 2010; Magnusson et al., 2014). One of the advantages of using WiTRACK is that it does not use raw wind speeds, instead, it uses 98^{th} percentile relative exceedance for tracking. This means that even if the simulation wind speed of TC is systematically weaker than historical observations, the 98^{th} percentile climatological wind in the models should also be lower than the observed 98^{th} percentile climatological wind. A TC will still be tracked by WiTRACK as long as there exists a 98^{th} percentile exceedance wind cluster. Consequently, a bias due to resolution does not have significant impact on WiTRACK as the tracking algorithm serves as a bias correction in this sense (detailed discussion on the impact of weaker wind speed in model outputs on WiTRACK can be found in Osinski et al. (2016)). Furthermore, it can be shown that, within the study area, the 98th percentile relative exceedance of the 4 models, which we used to construct the TPEPS TC event set, have similar behaviour (i.e. similar to Figure 2 of Osinski et al. (2016)). Consequently, individual PEPS TC event set can be combined to form a large PEPS TC event set, i.e. TPEPS TC event set.

3.2 Identifying TC~~typhoon~~-related pure EPS windstorm events

WiTRACK identifies windstorm events of all kind, including MEPS TCs, PEPS TCs, MEPS extratropical cyclones and PEPS extratropical cyclones. Therefore additional requirements are needed to identify typhoon-related PEPS TC events. ~~4~~Four post-processing procedures are used: (i) Geographic Filter (GF), (ii) Logistic Regression Classifier (LRC), (iii) MEPS TC Identifier (MTI), and (iv) Detection at Initialisation Filter (DIF).

3.2.1 Geographic Filter (GF)

GF was first introduced by Befort et al. (2020). It aims to remove non-TC-related windstorms, e.g. extratropical cyclones, cold surge outbreaks during the winter monsoon, and equatorial disturbances, from the event set by excluding windstorm events which solely identified north of 26° N and east of 100° E, and latitudinal position exclusively south of 10° N. Befort et al. (2020) found this filter can reduce the false alarm rate (i.e. the ratio between number of identified non-TC related windstorms and total number of detected windstorms) of TC identification in JRA-55.

3.2.2 Logistic Regression Classifier (LRC)

In order to reduce computational cost and increase computational efficiency, the classical methods to determine whether the atmospheric disturbance is a TC or non-TC via cold/warm core determination (e.g. Hart, 2003; Strachan et al., 2013) are not used because these methods require multiple variable fields which increase computational cost significantly. Instead, a statistical learning approach, logistic regression classifier (LRC), is used to determine whether the windstorm event is related to a TC or not. Details and background information of LRC can be found in Hastie et al. (2009) and the *caret* package in R is used for LRC training (Kuhn et al., 2018; available online at https://github.com/topepo/caret/). LRC is trained using the track characteristics of the event in the JRA-55 and ERA-Interim event set (1979-2014) as explanatory variables (Table 2). This combination of training set is chosen based on preliminary studies of constructing an optimal classifier using different combination of training set. In order to avoid issues that are associated with collinearity, a stepwise Variance Inflation Factor (VIF) selection method is used to identify independent variables. Variables with VIF value larger than 5 are excluded. 17 variables have been chosen to use in the construction of LRC (Table 3). Variables that relate to changes in storm position, lifetime of a storm, and mean wind field structure appear to be the most important variables in the LRC. This is expected as the typical trajectory, duration, and structure of TCs and other windstorms are very different. Validation using JRA-55 event set (2015-2017), which has 49 TC events and 47 non-TC events, have shown that the accuracy of the LRC is about 90% with low rate of false positives and false negatives.

3.2.3 MEPS TC Identifier (MTI)

Since there are many replicated events of forecasted historical TCs (i.e. MEPS) in the operational forecast archive, it is necessary to remove these events from our event set to avoid biases toward historical events. Instead of using the criteria suggested by Osinski et al. (2016), a set of strict criteria (MTI) is used in this study. This can ensure the statistics and climatology of TPEPS event set is not biased toward~~ the~~ historical events. The MTI eliminates forecast of MPES TC events where the forecasts of those MPES TCs were initialised (i) before, and (ii) after the time of MPES TC genesis (hereinafter type 1 and type 2 forecast events respectively). A similarity index (*SI*) (Eq. 1~~1~~) is used to eliminate type 1 forecast events:

$$d_i = \begin{cases} d_{\text{thres}} - d & d < d_{\text{thres}} \\ 0 & d \geq d_{\text{thres}} \end{cases}, \qquad (1\cancel{1}\text{a})$$

$$SI = \frac{\sum_i^{t_{\text{overlap}}} d_i}{d_{\text{thres}} \times t_{\text{overlap}}}, \qquad (1\cancel{1}\text{b})$$

where d is the great circle distance between position of historical TC and position of TIGGE TC at the overlap time step i, d_{thres} is the maximum tolerance of d, t_{overlap} is the number of overlap time steps in which both historical TC and TIGGE TC existed and it must be larger than 4. Events with SI larger than SI_{thres} are considered as MPES TC events. A series of sensitivity study have been done for determining the optimal choice of parameters (not shown) and the most optimal setting is d_{thres}=900 km and SI_{thres} = 0.1. Type 2 forecast events are found if the separation distance between the position of historical TC and the TIGGE TC at any point of their overlap time is less than 400 km. This threshold is determined by the minimum separation between historical TCs and TC in JRA-55 event set.

3.2.4 Detection at Initialisation Filter (DIF)

Any events that are detected at the time of model initialisation are removed following Osinski et al. (2016). It is because these events are likely to be related to pre-existing disturbances or structures that leads to their development. The removal of these events ensures the TPEPS event set is independent of any pre-existing weather patterns.

3.3 Adjustment procedure

More than one windstorm event could be found within a close proximity of each other over the WNP. Since the clustering algorithm in WiTRACK does not have a maximum size restriction on the cluster, multiple windstorm events in close proximity could be identified as one windstorm event by WiTRACK. An additional procedure is used to separate these merged windstorm events. This is an iterative procedure which would check whether all of the grid boxes at each 6-hourly time step of the windstorm are within 1,000 km radius from the centre of the windstorm cluster. If any of the event grid boxes are outside the 1,000 km radius, it will first remove these grid boxes and recalculate the centre of event cluster. This procedure is repeated until there is no change in the centre of cluster. This procedure addresses windstorm event with unrealistically large impact area and event SSI (ESSI). The event SSI (ESSI) is defined as

$$\mathrm{ESSI} = \sum_t^T \sum_k^K \left[\left(\max\left(0, \frac{v_{k,t}}{v_{98,k}} - 1\right)\right)^3 \times A_k \right] \tag{2}$$

where $v_{k,t}$ is the wind speed at grid box k and time step t, $v_{98,k}$ is the climatological 98th percentile wind speed at grid box k, A_k is the area-dependent weight. Summation is done over all time steps and all grid boxes affected by the windstorm. The threshold radius is chosen to be 1,000 km because typical size of TC wind field is smaller than a circle of 1,000 km radius (Lee et al., 2010; Chan and Chan, 2011).

4 Results and discussions

4.1 Statistics and Validations

In this section, we present validation of our TPEPS TC event set by comparing the climatological features as provided by a time- and ensemble-aggregated view of the TPEPS TC event set to the historical/reanalysis based event set. A historical TC is said to be detected in a forecast model if there exists a TC counterpart in the forecast model, which is similar to the historical TC as identified by the MTI (c.f. Sect. 3.2.3). The detection rates of historical TCs which are detected in different forecast outputs~~with at least one MEPS counterpart~~, i.e.~~ in~~ CMA, ECMWF, JMA, and NCEP, are 91.2%, 94.7%, 89.4%, and 90.7%, respectively, whereas only 54.2% of historical TCs in the period of 2008-2017 are detected in JRA-55 (Table 4). Since WiTRACK is a wind threshold exceedance based detection scheme and the 98th percentile wind speed value of JRA-55 within the tropical WNP is similar to these selected TIGGE data (Fig. 1), this implies JRA-55 underestimates the wind speed of wind field of TCs, which is in agreement with Murakami (2014). This also shows these selected TIGGE outputs provide a better representation of the atmosphere. Total 515,712 TC related windstorm events are detected in the selected TIGGE data set. ~38.5% of the all TPEPS events are PEPS TC events (Table 5). Percentage of total TC windstorms as PEPS TCs can be treated as a proxy to quantify the forecast skill of the model. For example, NCEP has 47.1% of TC windstorms as PEPS TCs whereas JMA has 26.5%. This indicates the NCEP model generates more "wrong" forecast than JMA however these "wrong" forecasts are physically possible. Yet, ~~this~~ examining the forecast skill of models is not the focus of this study and the rest of the discussion focuses on the TPEPS TC event set.

Figures 2 and 3 show the spatial pattern and temporal variability of the number of TC which are first detected for each day, respectively, of the TPEPS and JRA-55 event sets. While individual model might have bias in certain spatial and temporal domain, for example the region with the highest track density of JMA is at the eastern WNP in Fig. 2~~1~~d in comparison to other models, and NCEP failed to capture the peak activity prior 2012 in Fig. 3~~2~~, the overall patterns of the TPEPS event set match the JRA-55 event set. This is expected because (i) TC formation depends on the environmental conditions and initial disturbance (Gray, 1977; Ritchie and Holland, 1997; Nolan, 2007). During the period of active TC season, environmental conditions over WNP are usually favourable for TC formation but often there is no suitable disturbance in the region. Since EPS simulates the chaotic behaviour of the atmosphere, it would forecast disturbances which would be possible to form but not realised in the real atmosphere. Hence PEPS TCs can be formed in those period of time over WNP.~~;~~ And (ii) the trajectory of TCs depends mainly on the large scale environmental flow of the region (Chan, 2010). This implies

PEPS TCs would also follow the typical trajectory of real TCs given that the large scale flow is correctly represented in the forecast models. Thus, in general the spatial and temporal patterns of the TPEPS event set match the patterns of JRA-55 event set. There are several possible reasons which lead to the differences in spatial pattern between TPEPS event set and JRA-55 event set. The eastward bias in the track density appears to be a common feature in many GCMs (e.g. Camargo et al., 2005; Bell et al., 2013; Roberts et al., 2020), this has also been observed in seasonal forecast output (Camp et al., 2015). Finite simulation time has also contributed to this bias as TC that forms in the region east of 150 °E would not have sufficient time to move into the western part of WNP before the end of simulation time. Differences in number of tracks could also contribute to the differences in spatial pattern as more diverse tracks would be captured in larger event set.

Some TPEPS events appear in locations where no historical TC event is observed (Figs. 2c and 2f). While there is no historical TC event in some locations, this does not imply TC cannot occur in those regions. The historical data, which cover 39 years of observations, may not have enough samples to construct a distribution that can correctly represent the basic population (i.e. all possible TCs in the given climate). For example, the occurrence of Tropical Storm Vamei that formed close to the equator (~1.4° N) does not satisfy the classical "necessary but insufficient" conditions of TC formation, which are identified by Gray (1977) based on historical observations. This shows TC can appear in historically "TC-free" region. Furthermore, from the statistical perspective, the JRA-55 event set can be viewed as a subset which is randomly selected from the TPEPS event set. To provide more evidence to support this view, we have conducted bootstrap resampling on the TPEPS event set to obtain 10,000 sets of subsamples. Each set of subsamples has 668 events to mimic the number of events in the JRA-55 event set. Uncentred pattern correlation between the track density of the JRA-55 event set and the track density of each set of subsamples are calculated. In order to focus on the relevant entries, if the values of track density of a grid box for a resampling set and the JRA-55 event set are both less than 1, such grid box is neglected in the pattern correlation calculation. The mean, standard deviation, minimum and maximum of the uncentred pattern correlation of the 10,000 set of subsamples are 0.9380, 0.0107, 0.8961, and 0.9697, respectively. This suggests the spatial pattern of the JRA-55 event set is highly similar to some small random subsets of the TPEPS event set. Thus, the JRA-55 event set can been seen as a subset which is randomly selected from the TPEPS event set. On the other hand, it is not be possible to deduce the basic population (e.g. the TPEPS event set) from a small sample set (e.g. the JRA-55 event set). Although the spatial distribution of the small set sample is similar to the subsamples of the basic population and thus usable as one possible realisation of the basic population, the small sample set does not contain all of the information of the underlying population. Furthermore, the statistical estimate of extremes would also be different for the small sample set and the basic population.

~~The spatial discrepancy near the dateline between the JRA-55 event set and the TPEPS event set (Figs. 2c and 2f) can be explained by considering the amount of data used in the construction of event sets. The JRA-55 event set is constructed based on 39-year of reanalysis data whereas the TPEPS event set is constructed using more than 10,000 years of TC model data from operational forecast models. Since both event sets are constructed from physical models (i.e. GCMs), the JRA-55 event set can be considered as a subset of the TIGGE event set. Furthermore, in the region 0-20 °N and 160-180 °E, the 98th percentile values of JRA-55 is higher than all TIGGE models considered in this study (Fig. 1). Consequently, systems with the same strength would be identified in that region in the TPEPS event set but not in the JRA-55 event set.~~

Some of the examples of TPEPS TC tracks and impact footprints are shown in Fig. 4. The trajectory of these TPEPS TC tracks is indistinguishable to historical TC trajectories in WNP. This shows these TPEPS TC events are realistic and physically possible events. Figure 5 shows the climatological daily number distributions of TC first detections~~s~~ for TPEPS TC event set and JRA-55 event set. Although the peak activities period of JMA is slightly lagged behind and the over- and under-estimation of the peak of activity for CMA and NCEP are observed, respectively, the seasonal cycle of TPEPS TC event set is well captured and this matches to the seasonal cycle of the JRA-55 event set. This shows our new approach is capable to produce spatially and temporally realistic events.

In general, the temporal evolutions of the number of first storm detections of TPEPS event set during the integration time has an increasing trend in the short lead time followed by a roughly constant behaviour (Fig. 6). In short lead time (i.e. close to initialisation of forecast), the true state of the atmosphere is well simulated by forecast models, thus EPSs are likely to produce storms that actually occurred (i.e. MEPS storms) and less likely to produce PEPS storms (Osinski et al., 2016). As lead time increases, more PEPS storms are produced due to increasing uncertainty of the state and the chaotic behaviour of the atmosphere in EPSs. When EPS has no memory of the initialisation state of the atmosphere, the probability distribution of formation of PEPS TCs becomes a uniform distribution.

The overall impact of any storm is related to the many factors for example lifetime of the storm, the size of the storm, and the intensity (or strength) of the storm (e.g. Vickery et al., 2000; Mori and Takemi, 2016; Kim and Lee, 2019). Here we investigate whether there are systematic biases in the TPEPS TC event set which would affect these quantities. The lifetime distribution of TPEPS TCs matches to the JRA-55 event set but proportionally overestimates for short-lived TCs and underestimates for long-lived TCs (Fig. 7a). These differences are the consequence of the finite simulation time in forecast models. If the same restriction (i.e. finite simulation time window) is applied to the JRA-55 TC event set (grey shaded areas in Fig. 7), the lifetime distribution of TPEPS TCs would be in good agreement to the JRA-55 TCs. Similar conclusion can be reached in the comparison of the distribution of time required to reach lifetime maximum intensity (LMI) (Fig. 7b). However, finite simulation time of EPSs cannot explain the difference in the distribution of impact area, which is the total area that has experienced TC-associated extreme wind (i.e. larger than local climatological 98th percentile wind speed), between TPEPS and JRA-55 event sets despite they have the same type of distribution (Fig. 7c). The difference in the distributions of impact area maybe due to the fact that wind speed of the TC wind fields is underestimated in JRA-55 as discussed above. Consequently, many weaker TCs, which would have small impact areas, are not detected and thus they are not necessarily included in the JRA-55 TC event set.

4.2 Robust TC hazard~~risk~~ assessment

To demonstrate the benefit of our approach, TC records in IBTrACS, JRA-55 TC event set, and TPEPS TC event set are stratified into intensity classes according to their lifetime maximum intensity (c.f. Table 6). Since WiTRACK is an impact-oriented, wind speed percentile based tracking scheme which tracks TCs with potential impact (Befort et al., 2020)~~-~~, ~~-~~m~~M~~any of the low impact TCs (i.e. TCs in the Tropical Depression and Tropical Storm (TD&TS) category) are not detected and thus not included in the TPEPS TC event set. Focusing onto the categories of high impact TC, i.e. Typhoon (TY), Very Strong Typhoon (VST), and Violent Typhoon (VTY), the

TPEPS event set contains 302.14, 102.54, and 77.02 times more TY, VST, and VTY than the IBTrACS records, respectively. This means our new approach can capture much more extremely high impact events such that a more robust analysis of extreme TC events can be done.

The key advantage of this new approach is that it constructs a physically consistent and high information content TC event set with good and realistic representation of the current climate state using a computationally inexpensive algorithm. Since more physically consistent and physically possible TCs are included, more extreme events can be captured in the TPEPS event set. Consequently, a robust TC hazard~~risk~~ assessment can be obtained. Some of the examples are presented in this subsection.

Figure 8 shows the location of first detection of TCs with LMI at least typhoon strength, which made landfall within the given domain (105-180° E, 0-30° N) for TPEPS and JRA-55 TC event set. The spatial pattern of the TPEPS TC event set (Fig. 8f) matches the spatial pattern of the JRA-55 TC event set. The data in the JRA-55 TC event set are sparse and it does not provide sufficient information about whether TCs, which made landfall in this region, are typically first identified in the WNP or in the South China Sea (SCS). The TPEPS TC event set, on the other hand, provides a clearer picture and suggests events, which made landfall in this domain, are typically first identified in the SCS and western WNP. This is consistent with the known climatology. As TCs within the SCS and western WNP usually follow the western and northwestern trajectory and subsequently made landfall over the Vietnam, south and southeast mainland China, Taiwan, and the Philippines.

Figure 9 shows the number of TC landfall events, which made landfall with at least typhoon strength, with the focus of southern and southeast mainland China, and Taiwan. Much more landfall events have been captured by TPEPS TC event set (11449) than the JRA-55 TC event set (100). The spatial distribution of TPEPS TCs is in good agreement with the JRA-55 TCs with uncentred pattern correlation of 0.8345. TCs, which made landfall with at least typhoon strength, are more likely to made landfall along the coast of the southern Fujian Province and the eastern Guangdong Province than any other coastal area of South and Southeast mainland China. Furthermore, higher TC landfall frequency is observed on the side of islands (i.e. Hainan Island and Taiwan) which faces the open ocean than the other side of islands. This is consistent with observations. The TPEPS TC event set also provides information about the frequency of TC landfall at locations where no landfall events had observed in the JRA-55 TC event set, e.g. locations along the coastline of Guangdong Province. Furthermore, the distribution of landfall intensity for TCs, which made landfall with at least typhoon strength, for the TPEPS TC event set is very similar to the JRA-55 TC event set (the null hypothesis, i.e. the distributions are the same, is not rejected at the 0.05 significance level of the two-sample Kolmogorov-Smirnov test).

4.3 Application

The TPEPS TC event set is constructed based on physical models, i.e. GCMs, which provide a good representation of the atmosphere of the real world. The wind field associates to a TPEPS TC event is realistic and local effects, such as local topography, have been taken into account. This implies the wind information of the TPEPS TC event set can be used for estimates return periods of local extreme wind events associated with typhoon with high confidence. F~~F~~igure 10 shows the number of TC-related 6-hourly extreme wind (i.e. wind speed higher than the local 98th percentile climatological wind speed) data entries in each of the grid box within Guangdong Province in the Southern China. The JRA-55 TC event set can only construct a TC-related 6-hourly extreme distribution with ~25 (inland) and ~325 (coastal) data entries whereas such distribution can be constructed with at least 500 to over 28,000 data entries using the TPEPS TC event set. This implies the estimated return period using the TPEPS TC event set would be more reliable than using the JRA-55 TC event set and similarly the observation data alone. This is of importance from the DRR perspective as wind speed values are used in practice to decide on payments out of parametric insurance products (Swiss Re, 2016). Consequently, reliable wind-based trigger points of typhoon parametric insurance can be determined. This will further improve the suitability and flexibility of parametric insurance for DRR applications. Ultimately, this will improve the speed of post-disaster recovery. A demonstration for such ~~calculation~~application is given below. ~~This is demonstrated as follows.~~

Four surface observation stations are chosen for this demonstration, they are Baiyun International Airport (BAIYUN INTL; 23.392° N, 113.299° E; from 1945-2019), Baoan International Airport (BAOAN INTL; 22.639° N, 113.811° E; from 1957-2019), Shanwei (22.783° N, 115.367° E; from 1956-2019), and Shangchuan Dao (21.733° N, 112.767° E; from 1959-2019). For each selected surface station, the grid box of each EPS that corresponds to the surface station is identified (Fig.11). Resolution of models is known to be a factor to limit the wind speed of TCs (Bengtsson et al., 2007). This means for the same TC, the associated wind speed would be lower in low resolution model and higher for high resolution model. In order to utilise the extreme wind information from EPSs with different resolution, the cube of 98th percentile relative exceedance of wind speed (EXCE) is used. Since EXCE is a ratio, it is a resolution independent quantity and the tail behaviours of the EXCE distribution for these models are similar, which is in agreement with Osinski et al. (2016). Information from different models can be combined using EXCE. EXCE entries, which correspond to TC in the TPEPS TC event set, are extracted for those grid boxes. This forms a set of "observations" of the impacts of high impact TCs at those grid boxes in the model space. We assume all of the EXCE entries are independent and identically distributed (iid) random variables. This is a reasonable assumption, due to the fast moving nature of TCs, diverse possible direction of the movement of wind field, and rapid decay of wind field over land for a 6-hour interval, local observations often have only one extreme wind observation~~s~~ for~~of~~ a TC event. In order to translate this information to the physical world, quantile mapping is used for mapping EXCE to the observed surface wind speed which exceeded local climatological 98th percentile. Historical in situ surface wind data are obtained from the Integrated Surface Database (ISD) (Smith et al., 2011). Quantile mapping is done using the R package *qmap* (Gudmundsson et al., 2012; Gudmundsson, 2016). Due to different geographic configuration and climatology of each in situ observation station, different quantile mapping strategies have been employed. The optimal strategy is chosen based on minimisation of the root-mean-square-error (RMSE) of the quantile mapping output (see Gudmundsson (2016) for more details). Using above information, the return period-return level plot (using threshold exceedance approach) is constructed using the R package *extRemes* (Gilleland and Katz, 2016). For detail discussion of calculation of return period and return level, readers are referred to Elsner et al. (2006), Jagger and Elsner (2006), and Gilleland and Katz (2016). Figure 12 shows the return period-return level plot of ~~4~~four selected stations which are derived using our proposed approach with the TPEPS TC event set and using in situ observational data. The width of the 95% confidence interval which is calculated using our proposed approach is much sharper than the 95% confidence interval which is calculated using in situ observational data. In other words, the uncertainty can be reduced by using the TPEPS TC event set because more observations are used in the calculation.

The above application of the TPEPS TC event set can provide crucial information for the DRR community. As discussed in the introduction, typhoon parametric insurance can be an effective financial instrument for typhoon risk transfer. However, an effective typhoon parametric insurance requires a robust trigger point, which is determined by the meteorological information, e.g. wind speed. If the trigger point is too high, disbursements would not be made even if a catastrophic meteorological disaster has occurred, i.e. under-compensation; If the trigger point is too low, disbursements would be made even if no catastrophic event has occurred. Using the TPEPS TC event set, the estimated return period has smaller uncertainty than the estimation made by in situ observational data, such that an optimal trigger point for typhoon parametric insurance can be determined.

5 Summary and Conclusions

In this study, a new and efficient method, which addresses the critical issue in typhoon risk assessments – a robust methodology to determine the real frequency of TC occurrence with high socioeconomic impact potential by constructing a physically consistent TC event set, is presented ~~to produce a physically consistent TC event set with high information content in the WNP has been presented~~. This is achieved by applying an objective impact-oriented windstorm identification algorithm – WiTRACK, on 6-hourly 10-m horizontal wind field of selected ensemble data set from a multi-centre grand ensemble data archive – TIGGE. While WiTRACK identifies major events based on one meteorological variable only, it is capable of identifying events of general loss relevance as demonstrated by Befort et al. (2020). This implies the event set generated by our approach is in principle suitable for general TC risk assessments, as well as for an assessment of the hazards frequency-intensity distribution specifically.Befort et al. (2020) Several sensitivity tests with different parameter settings are done using JRA-55 data to obtain the optimal setup for WiTRACK. Since WiTRACK can identify all types of windstorm events, ~~4~~four post-processing procedures are used to identify PEPS TCs, these procedures include a geographic filter and logistics regression classifier. The TPEPS event set has the climatological spatial and temporal pattern of TCs which match to the historical climatological pattern of TC in WNP. More than 302, 102, and 77 times of TY, VSTY, and VTY, respectively, are found in the TPEPS TC event set in comparison to the IBTrACS record. A robust representation of extreme TC events in WNP can be obtained using the TPEPS TC event set because of the high number of physically consistent extreme events. Consequently, a robust hazard ~~risk~~ assessment of land-affecting ~~typhoons~~ TCs in the WNP can be produced using the event set constructed by this new method. Furthermore, the return-period of typhoon-related extreme wind events e.g. Typhoon Haiyan (2013) and Typhoon Mangkhut (2018), can be determined with sharper confidence intervals in a similar manner as Walz and Leckebusch (2019). As a result, policymakers and related stakeholders can improve the current typhoon related disaster reduction and mitigation strategy. ~~Furthermore~~Furthermore, a robust trigger point for parametric typhoon hazard insurance can be determined using our proposed approach by reducing the uncertainty of estimated return period of a meteorological extreme event. This will improve the suitability and flexibility of parametric insurance for DRR applications. Consequently, this will improve the speed of post-disaster recovery.

The TC event set constructed using the method described in this study~~paper~~ has several unique properties in comparison to the TC event set constructed by other methods (Vickery et al., 2000; Emanuel et al., 2006; Rumpf et al., 2009; Kim and Lee, 2019):

(i) ~~(i)~~ Many methods in the literature (e.g. Emanuel et al., 2006; Rumpf et al., 2009) use historical best track data to construct a spatial probability function that determine the genesis location of synthetic TCs and a parametric track model, that matches to the historical observations, to determine the movement of synthetic TCs. Consequently, these synthetic tracks are highly likely to be identified in the region where TCs were identified from the historical observations and highly ~~unlikely~~ rare in the region where TCs were never identified but physically possible. In contrast, TPEPS TCs are detected at any physically possible locations over the WNP. This means, besides the events, which are similar to the historical observations, the TPEPS TC event set also includes events that occur in the region where no historical event was observed. ~~There are two underlying reasons: (1) The physical reason – TC has low probability of occurrence in those regions due to typical unfavourable environment. For example, following the classical "necessary but insufficient" conditions of TC formation which are identified by ; (2) The statistical reason – the observation period is too short and it does not provide enough samples for the underlying population.~~Consequently, the TPEPS TC event set provides an important and unique advantage for typhoon hazard ~~risk~~ assessment. In comparison to other methods to generate large TC event sets, our specific approach is limited mainly by the source of data used. The current TC event set constructed using medium range forecasts archived in TIGGE, is strictly spoken representative only for the current climate state. Any longer-term climate variability (e.g. multi-decadal fluctuations like the Pacific Decadal Oscillation (PDO)) and their impacts on any TC frequency-intensity distribution are not accounted for in this setting. Nevertheless, the presented approach would be equally applicable to data sets representing that kind of variability on longer time scales (e.g. decadal predictions or transient climate model simulations).

~~These events are typically associated with low probability of occurrence, yet, they are still physically possible.~~

~~The TPEPS event set includes events which are unlikely but physically possible. This provides an important and unique advantage for typhoon risk assessment.~~

(ii) In the literature, the structure of wind field of synthetic TCs follows a predefined, analytical model, e.g. parametric vortex structure developed by Holland (1980) or modified Rankine vortex. For the TPEPS TC event set, complex physical processes in GCMs determine the structure of wind field of TCs, therefore the structure of wind field of TCs is realistic. This is an advantage for robust wind hazard~~risk~~ assessment of land-affecting TCs because the resultant wind field includes the complex atmosphere-land interaction which depends on the local topography. Consequently, the TPEPS TC event set can be used as addition observations for the estimation of return period of TC-related extreme wind as demonstrated above.

(iii) Many of the TC risk~~risk~~ assessments are done based on wind risk, and/or wind-induced coastal risk but not TC-related precipitation risk (Vickery et al., 2000; Emanuel et al., 2006; Rumpf et al., 2009; Mendelsohn et al., 2012; Mori and Takemi, 2016; Marsooli et al., 2019; Kim and Lee, 2019). A reason is that historical damages due to TC-related wind are much better documented than TC-related precipitation damages (Emanuel et al., 2006). However, damages due to TC-related precipitation, e.g. flooding, should not be ignored. Based on the pay-out of the National Flood Insurance Program of the United States for the flood event of Hurricane Ike (2008), Smith and Katz (2013) estimated the insured flood damage as 5.~~376~~4 billion USD. Furthermore, some of the high impact

TCs in WNP have typical typhoon intensity but the amount of rainfall is extremely high, e.g. Typhoon Morakot (2009) (Wu, 2012). Since precipitation is one of the output variables of these medium range ensemble forecasts, precipitation-related impact can be examine by integrating the realistic precipitation information from forecast outputs into the TPEPS TC event set. Furthermore a spatial distribution of TC related hazard, e.g. extreme wind and extreme precipitation, of the TPEPS TC event set can be constructed using the notion of TC hazard footprint (Chen et al., 2018). Consequently, a more thorough typhoon risk assessment can be achieved. This is currently under our investigation.

In conclusion, the event set that we have constructed contains all necessary information for applications in the DRR context. This event set can improve the hazard component in an overall assessment of integrated TC risks (e.g. Sajjad and Chan, 2019) by providing a robust probability of occurrence of extreme TC event. Furthermore, using this event set, a robust trigger points of parametric insurance for the local hazard can be determined. Once such trigger points for the local hazard are available (including their uncertainty), the targeted application of parametric insurance products in disaster relief application is possible. Especially, when it comes to the evaluation of the basis risk. This study is merely the first step toward a statistically robust, full physical model based TC hazard assessment. The impact of TC-related extreme precipitation and storm surges can be integrated following the approach developed by Befort et al. (2015).

Data availability. JRA-55 (Kobayashi et al., 2015) and ERA-I (Dee et al., 2011) are freely available for academic use at the UCAR Research Data Archive: https://rda.ucar.edu/datasets. The TIGGE dataset (Bougeault et al., 2010; Swinbank et al., 2015) used in this study can be accessed through ECMWF server: https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=pf/. IBTrACS (Knapp et al., 2010) and ISD (Smith et al., 2011) are available at the United States National Centers for Environmental Information, National oceanic and Atmospheric Administration: https://www.ncdc.noaa.gov/ibtracs/index.php, and https://www.ncdc.noaa.gov/isd, respectively. JTWC best track data used in this study is obtained from the United States Navy Website: https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks.

Author contribution. KSN and GCL originated the idea, developed the methodology, performed data analysis, and wrote the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgments. The authors thank three reviewers for their helpful and constructive comments. The authors thank Drs. D. Befort and M. Angus for valuable discussion. This work was supported by the Building Resilience to Natural Disasters using Financial Instruments grant INPAIS (Integrated Threshold Development for Parametric Insurance Solutions for Guangdong Province China, Grant Ref: NE/R014264/1, through Natural Environment Research Council (NERC). The computations described in this paper were performed using the BlueBEAR HPC service at the University of Birmingham.

References

Bakkensen, L. A., Shi, X., and Zurita, B. D.: The Impact of Disaster Data on Estimating Damage Determinants and Climate Costs, Economics of Disasters and Climate Change, 2, 49-71, 10.1007/s41885-017-0018-x, 2018.

Befort, D. J., Fischer, M., Leckebusch, G. C., Ulbrich, U., Ganske, A., Rosenhagen, G., and Heinrich, H.: Identification of storm surge events over the German Bight from atmospheric reanalysis and climate model data, Natural Hazards and Earth System Sciences, 15, 1437, 2015.

Befort, D. J., Wild, S., Kruschke, T., Ulbrich, U., and Leckebusch, G. C.: Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses, Atmos Sci Lett, 17, 586-595, 10.1002/asl.694, 2016.

Befort, D. J., Wild, S., Knight, J. R., Lockwood, J. F., Thornton, H. E., Hermanson, L., Bett, P. E., Weisheimer, A., and Leckebusch, G. C.: Seasonal forecast skill for extratropical cyclones and windstorms, Q J Roy Meteor Soc, 145, 92-104, 10.1002/qj.3406, 2019.

Befort, D. J., Kruschke, T., and Leckebusch, G. C.: Objective identification of potentially damaging tropical cyclones over the Western North Pacific, Environmental Research Communications, 2, 031005, 10.1088/2515-7620/ab7b35, 2020.

Belanger, J. I., Webster, P. J., Curry, J. A., and Jelinek, M. T.: Extended Prediction of North Indian Ocean Tropical Cyclones, Weather and Forecasting, 27, 757-769, 10.1175/WAF-D-11-00083.1, 2012.

Bell, R., Strachan, J., Vidale, P. L., Hodges, K., and Roberts, M.: Response of Tropical Cyclones to Idealized Climate Change Experiments in a Global High-Resolution Coupled General Circulation Model, J Climate, 26, 7966-7980, 10.1175/JCLI-D-12-00749.1, 2013.

Bengtsson, L., Hodges, K. I., and Esch, M.: Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses, Tellus A, 59, 396-416, 2007.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.-Y., Parsons, D., Raoult, B., Schuster, D., Dias, P. S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., and Worley, S.: The THORPEX Interactive Grand Global Ensemble, B Am Meteorol Soc, 91, 1059-1072, 10.1175/2010BAMS2853.1, 2010.

Buckingham, C., Marchok, T., Ginis, I., Rothstein, L., and Rowe, D.: Short- and Medium-Range Prediction of Tropical and Transitioning Cyclone Tracks within the NCEP Global Ensemble Forecasting System, Weather and Forecasting, 25, 1736-1754, 10.1175/2010WAF2222398.1, 2010.

Camargo, S. J., Barnston, A. G., and Zebiak, S. E.: A statistical assessment of tropical cyclone activity in atmospheric general circulation models, Tellus A, 57, 589-604, 10.1111/j.1600-0870.2005.00117.x, 2005.

Camp, J., Roberts, M., MacLachlan, C., Wallace, E., Hermanson, L., Brookshaw, A., Arribas, A., and Scaife, A. A.: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system, Q J Roy Meteor Soc, 141, 2206-2219, 10.1002/qj.2516, 2015.

Cavallo, E., and Noy, I.: Natural Disasters and the Economy — A Survey, International Review of Environmental and Resource Economics, 5, 63-102, 10.1561/101.00000039, 2011.

Chan, J. C. L.: Movement of Tropical Cyclones, in: Global Perspectives on Tropical Cyclones, World Scientific Series on Asia-Pacific Weather and Climate, Volume 4, World Scientific, 133-148, 2010.

Chan, K. T. F., and Chan, J. C. L.: Size and Strength of Tropical Cyclones as Inferred from QuikSCAT Data, Mon Weather Rev, 140, 811-824, 10.1175/MWR-D-10-05062.1, 2011.

Chen, W., Lu, Y., Sun, S., Duan, Y., and Leckebusch, G. C.: Hazard Footprint-Based Normalization of Economic Losses from Tropical Cyclones in China During 1983–2015, International Journal of Disaster Risk Science, 9, 195-206, 10.1007/s13753-018-0172-y, 2018.

CMA: Member report, ESCAP/WMO Typhoon Committee 13th Integrated Workshop, 36, 2018.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q J Roy Meteor Soc, 137, 553-597, 10.1002/qj.828, 2011.

Desai, B., Maskrey, A., Peduzzi, P., De Bono, A., and Herold, C.: Making Development Sustainable: The Future of Disaster Risk Management, Global Assessment Report on Disaster Risk Reduction, United Nations Office for Disaster Risk Reduction (UNISDR), Genève, Suisse, Geneva: UNISDR, 2015.

Elsner, J. B., Jagger, T. H., and Tsonis, A. A.: Estimated return periods for Hurricane Katrina, Geophys Res Lett, 33, 10.1029/2005GL025452, 2006.

Emanuel, K.: Climate and tropical cyclone activity: A new model downscaling approach, J Climate, 19, 4797-4802, Doi 10.1175/Jcli3908.1, 2006.

Emanuel, K., Ravela, S., Vivant, E., and Risi, C.: A statistical deterministic approach to hurricane risk assessment, B Am Meteorol Soc, 87, 299-314, 10.1175/Bams-87-3-299, 2006.

Gilleland, E., and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in R, Journal of Statistical Software; Vol 1, Issue 8 (2016), 10.18637/jss.v072.i08, 2016.

Glauber, J. W.: Crop Insurance Reconsidered, American Journal of Agricultural Economics, 86, 1179-1195, 10.1111/j.0002-9092.2004.00663.x, 2004.

Gray, W. M.: Tropical Cyclone Genesis in the Western North Pacific, J Meteorol Soc Jpn, 55, 465-482, 1977.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383-3390, 10.5194/hess-16-3383-2012, 2012.

Gudmundsson, L.: qmap: Statistical transformations for post-processing climate model output. R package version 1.0-4. 2016.

Halperin, D. J., Fuelberg, H. E., Hart, R. E., Cossuth, J. H., Sura, P., and Pasch, R. J.: An Evaluation of Tropical Cyclone Genesis Forecasts from Global Numerical Models, Weather and Forecasting, 28, 1423-1445, 10.1175/WAF-D-13-00008.1, 2013.

Hamill, T. M., Whitaker, J. S., Fiorino, M., and Benjamin, S. G.: Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter, Mon Weather Rev, 139, 668-688, 10.1175/2010MWR3456.1, 2010.

Hart, R. E.: A cyclone phase space derived from thermal wind and thermal asymmetry, Mon Weather Rev, 131, 585-616, Doi 10.1175/1520-0493(2003)131<0585:Acpsdf>2.0.Co;2, 2003.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, Springer Series in Statistics, Springer-Verlag New York, 745 pp., 2009.

Holland, G. J.: An Analytic Model of the Wind and Pressure Profiles in Hurricanes, Mon Weather Rev, 108, 1212-1218, 10.1175/1520-0493(1980)108<1212:AAMOTW>2.0.CO;2, 1980.

IPCC: Managing the risks of extreme events and disasters to advance climate change adaptation, Cambridge, A special report of Working Groups I and II of the Intergovernmental Panel on Climate Change, 2012.

Jagger, T. H., and Elsner, J. B.: Climatology Models for Extreme Hurricane Winds near the United States, J Climate, 19, 3220-3236, 10.1175/JCLI3913.1, 2006.

Jing, R., and Lin, N.: An Environment-Dependent Probabilistic Tropical Cyclone Model, J Adv Model Earth Sy, 12, e2019MS001975, 10.1029/2019MS001975, 2020.

Kim, G. Y., and Lee, S.: Prediction of extreme wind by stochastic typhoon model considering climate change, Journal of Wind Engineering and Industrial Aerodynamics, 192, 17-30, 10.1016/j.jweia.2019.05.003, 2019.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The International Best Track Archive for Climate Stewardship (IBTrACS) Unifying Tropical Cyclone Data, B Am Meteorol Soc, 91, 363-376, Doi 10.1175/2009bams2755.1, 2010.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93, 5-48, 10.2151/jmsj.2015-001, 2015.

Kruschke, T.: Winter wind storms: Identification, verification of decadal predictions, and regionalization, Doktors der Naturwissenschaften, Institut f• ur Meteorologie, Freie Universit• at Berlin, 181 pp., 2015.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., and Hunt, T.: Classification and Regression Training. 2018.

Leckebusch, G. C., Renggli, D., and Ulbrich, U.: Development and Application of an Objective Storm Severity Measure for the Northeast Atlantic Region, Meteorologische Zeitschrift, 17, 575-587, 10.1127/0941-2948/2008/0323, 2008.

Lee, C.-S., Cheung, K. K. W., Fang, W.-T., and Elsberry, R. L.: Initial Maintenance of Tropical Cyclone Size in the Western North Pacific, Mon Weather Rev, 138, 3207-3223, 10.1175/2010MWR3023.1, 2010.

Lee, C.-Y., Tippett, M. K., Sobel, A. H., and Camargo, S. J.: An Environmentally Forced Tropical Cyclone Hazard Model, J Adv Model Earth Sy, 10, 223-241, 10.1002/2017MS001186, 2018.
Lemcke, G.: A resilient world: NatCat parametric insurance solutions for China's Provincial Government, Sigma event 2017: Catastrophes-Protecting the unisured. Solutions for a resilient world, Zurich, Switzerland, 6-7 April, 2017.
Leonardo, N. M., and Colle, B. A.: Verification of Multimodel Ensemble Forecasts of North Atlantic Tropical Cyclones, Weather and Forecasting, 32, 2083-2101, 10.1175/WAF-D-17-0058.1, 2017.
Luitel, B., Villarini, G., and Vecchi, G. A.: Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones, Journal of Hydrology, 556, 1026-1037, 10.1016/j.jhydrol.2016.09.019, 2018.
Magnusson, L., Bidlot, J.-R., Lang, S. T. K., Thorpe, A., Wedi, N., and Yamaguchi, M.: Evaluation of Medium-Range Forecasts for Hurricane Sandy, Mon Weather Rev, 142, 1962-1981, 10.1175/MWR-D-13-00228.1, 2014.
Majumdar, S. J., and Torn, R. D.: Probabilistic Verification of Global and Mesoscale Ensemble Forecasts of Tropical Cyclogenesis, Weather and Forecasting, 29, 1181-1198, 10.1175/WAF-D-14-00028.1, 2014.
Marsooli, R., Lin, N., Emanuel, K., and Feng, K.: Climate change exacerbates hurricane flood hazards along US Atlantic and Gulf Coasts in spatially varying patterns, Nature Communications, 10, 3785, 10.1038/s41467-019-11755-z, 2019.
Mendelsohn, R., Emanuel, K., Chonabayashi, S., and Bakkensen, L.: The impact of climate change on global tropical cyclone damage, Nature Climate Change, 2, 205-209, 10.1038/nclimate1357, 2012.
Mori, N., and Takemi, T.: Impact assessment of coastal hazards due to future changes of tropical cyclones in the North Pacific Ocean, Weather and Climate Extremes, 11, 53-69, 10.1016/j.wace.2015.09.002, 2016.
Murakami, H.: Tropical cyclones in reanalysis data sets, Geophys Res Lett, 41, 2133-2141, 10.1002/2014GL059519, 2014.
Nissen, K. M., Ulbrich, U., Leckebusch, G. C., and Kuhnel, I.: Decadal windstorm activity in the North Atlantic-European sector and its relationship to the meridional overturning circulation in an ensemble of simulations with a coupled climate model, Climate Dynamics, 43, 1545-1555, 10.1007/s00382-013-1975-6, 2014.
Nolan, D. S.: What is the trigger for tropical cyclogenesis?, Aust. Met. Mag., 56, 241-266, 2007.
Osinski, R., Lorenz, P., Kruschke, T., Voigt, M., Ulbrich, U., Leckebusch, G. C., Faust, E., Hofherr, T., and Majewski, D.: An approach to build an event set of European windstorms based on ECMWF EPS, Nat. Hazards Earth Syst. Sci., 16, 255-268, 10.5194/nhess-16-255-2016, 2016.
Ritchie, E. A., and Holland, G. J.: Scale interactions during the formation of Typhoon Irving, Mon Weather Rev, 125, 1377-1396, 1997.
Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L.-P., Chauvin, F., Terray, L., Valcke, S., Moine, M.-P., Putrasahan, D., Roberts, C., Senan, R., Zarzycki, C., and Ullrich, P.: Impact of Model Resolution on Tropical Cyclone Simulation Using the HighResMIP–PRIMAVERA Multimodel Ensemble, J Climate, 33, 2557-2583, 10.1175/JCLI-D-19-0639.1, 2020.
Rumpf, J., Weindl, H., Höppe, P., Rauch, E., and Schmidt, V.: Stochastic modelling of tropical cyclone tracks, Mathematical Methods of Operations Research, 66, 475-490, 10.1007/s00186-007-0168-7, 2007.
Rumpf, J., Weindl, H., Höppe, P., Rauch, E., and Schmidt, V.: Tropical cyclone hazard assessment using model-based track simulation, Nat Hazards, 48, 383-398, 10.1007/s11069-008-9268-9, 2009.
Sajjad, M., and Chan, J. C. L.: Risk assessment for the sustainability of coastal communities: A preliminary study, Science of The Total Environment, 671, 339-350, https://doi.org/10.1016/j.scitotenv.2019.03.326, 2019.
Sajjad, M., Chan, J. C. L., and Kanwal, S.: Integrating spatial statistics tools for coastal risk management: A case-study of typhoon risk in mainland China, Ocean & Coastal Management, 184, 105018, https://doi.org/10.1016/j.ocecoaman.2019.105018, 2020.
Shi, P.: On the role of government in integrated disaster risk governance—Based on practices in China, International Journal of Disaster Risk Science, 3, 139-146, 10.1007/s13753-012-0014-2, 2012.
Smith, A., Lott, N., and Vose, R.: The Integrated Surface Database: Recent Developments and Partnerships, B Am Meteorol Soc, 92, 704-708, 10.1175/2011BAMS3015.1, 2011.
Smith, A. B., and Katz, R. W.: US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases, Nat Hazards, 67, 387-410, 10.1007/s11069-013-0566-5, 2013.
Smith, A. B., and Matthews, J. L.: Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates, Nat Hazards, 77, 1829-1851, 10.1007/s11069-015-1678-x, 2015.
Strachan, J., Vidale, P. L., Hodges, K., Roberts, M., and Demory, M.-E.: Investigating Global Tropical Cyclone Activity with a Hierarchy of AGCMs: The Role of Model Resolution, J Climate, 26, 133-152, 10.1175/JCLI-D-12-00012.1, 2013.

Sun, B., Guo, C., and Cornelis van Kooten, G.: Hedging weather risk for corn production in Northeastern China: The efficiency of weather-indexed insurance, Agricultural Finance Review, 74, 555-572, 10.1108/AFR-01-2014-0001, 2014.

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., Keller, J. H., Matsueda, M., Methven, J., Pappenberger, F., Scheuerer, M., Titley, H. A., Wilson, L., and Yamaguchi, M.: The TIGGE Project and Its Achievements, B Am Meteorol Soc, 97, 49-67, 10.1175/BAMS-D-13-00191.1, 2015.

Swiss Re: Natural catastrophes and man-made disasters in 2015: Asia suffers substantial losses. https://reliefweb.int/sites/reliefweb.int/files/resources/sigma1_2016_en.pdf, 2016.

Vickery, P. J., Skerlj, P. F., and Twisdale, L. A.: Simulation of Hurricane Risk in the U.S. Using Empirical Track Model, Journal of Structural Engineering, 126, 1222-1237, 10.1061/(ASCE)0733-9445(2000)126:10(1222), 2000.

Vitart, F., Prates, F., Bonet, A., and Sahin, C.: New tropical cyclone products on the web, ECMWF Newsletter, 130, 2012.

Walz, M. A., and Leckebusch, G. C.: Loss potentials based on an ensemble forecast: How likely are winter windstorm losses similar to 1990?, Atmos Sci Lett, 20, e891, 10.1002/asl.891, 2019.

WMO: Typhoon Committee Operational Manual, World Meteorological Organization, World Meteorological Organization, 2019.

Wu, C.-C.: Typhoon Morakot: Key Findings from the Journal TAO for Improving Prediction of Extreme Rains at Landfall, B Am Meteorol Soc, 94, 155-160, 10.1175/BAMS-D-11-00155.1, 2012.

Xu, W., Qi, L., Du, Y., and Xia, L.: Analysis on Abnormal Tropical Cyclone Track Forecast Error of Ecmwf-Ifs in the Western North Pacific, Tropical Cyclone Research and Review, 5, 12-22, https://doi.org/10.6057/2016TCRRh1.02, 2016.

Yamaguchi, M., Vitart, F., Lang, S. T. K., Magnusson, L., Elsberry, R. L., Elliott, G., Kyouda, M., and Nakazawa, T.: Global Distribution of the Skill of Tropical Cyclone Activity Forecasts on Short- to Medium-Range Time Scales, Weather and Forecasting, 30, 1695-1709, 10.1175/WAF-D-14-00136.1, 2015.

Ye, T., Wang, Y., Wu, B., Shi, P., Wang, M., and Hu, X.: Government Investment in Disaster Risk Reduction Based on a Probabilistic Risk Model: A Case Study of Typhoon Disasters in Shenzhen, China, International Journal of Disaster Risk Science, 7, 123-137, 10.1007/s13753-016-0092-7, 2016.

Ye, T., Li, Y., Gao, Y., Wang, J., and Yi, M.: Designing index-based livestock insurance for managing snow disaster risk in Eastern Inner Mongolia, China, International Journal of Disaster Risk Reduction, 23, 160-168, 10.1016/j.ijdrr.2017.04.013, 2017.

Zhang, X., Chen, G., Yu, H., and Zeng, Z.: Verification of Ensemble Track Forecasts of Tropical Cyclones During 2014, Tropical Cyclone Research and Review, 4, 79-87, https://doi.org/10.6057/2015TCRR02.04, 2015.

Tables

Centre	Number of members	Runs per day	Resolution	Implementation date	Forecast lead time (hr)
CMA	14	2 (00, 12 UTC)	0.5625°×0.5625°	20070515	240
		2 (00, 12 UTC)		20140805	360
ECMWF	50	2 (00, 12 UTC)	0.5625°×0.5625°	20061001	360
JMA	50	1 (12 UTC)	1.25° × 1.25°	20060301	216
	50	1 (12 UTC)		20130328	264
	26	2 (0, 12 UTC)		20140226	264
NCEP	20	4 (0, 6, 12, 18 UTC)	1.0° × 1.0°	20070327	384

Table 1. Information of selected data sources from TIGGE archive.

Variables
Time average of area of cluster
Time average of longitude of cluster centre
Time average of latitude of cluster centre
Time average of maximum extent of cluster
Time average of mean wind speed
Time average of standard deviation of wind speed
Time average of minimum wind speed
Time average of maximum wind speed
Time average of longitude of location of maximum wind
Time average of latitude of location of maximum wind
Time average of storm severity index (SSI)
Standard deviation of time series of area of cluster
Standard deviation of time series of longitude of cluster centre
Standard deviation of time series of latitude of cluster centre
Standard deviation of time series of maximum extent of cluster
Standard deviation of time series of mean wind speed
Standard deviation of time series of standard deviation of wind speed
Standard deviation of time series of minimum wind speed
Standard deviation of time series of maximum wind speed
Standard deviation of time series of longitude of location of maximum wind
Standard deviation of time series of latitude of location of maximum wind
Standard deviation of time series of storm severity index
Number of 6-hourly time steps
Area of windstorm event footprint
Event SSI
Difference of latitude between the initial and final locations
Difference of longitude between the initial and final locations
Total distance travelled

Table 2. List of explanatory variables which are initially considered in the LRC model. ~~can be obtained from the WiTRACK output.~~

Variable	t-value
Difference of latitude between the initial and final locations	12.5707
Difference of longitude between the initial and final locations	9.9983
Time average of standard deviation of wind speed	9.3709
Time average of minimum wind speed	8.5015
Time average of maximum extent of cluster	5.1416
Number of 6-hourly time steps	4.8719
Standard deviation of times series of latitude of location of maximum wind	3.4302
Standard deviation of times series of mean wind speed	2.3640
Standard deviation of times series of area of cluster	2.2447
Event SSI	1.9621
Standard deviation of times series of maximum extent of cluster	1.7922
Time average of latitude of cluster centre	1.4493
Standard deviation of time series of SSI	0.9980
Standard deviation of times series of longitude of location of maximum wind	0.9237
Standard deviation of times series of standard deviation of wind speed	0.7268
Time average of longitude of location of maximum wind	0.4204
Standard deviation of time series of minimum wind speed	0.2613

Table 3. List of explanatory variables and their associated t-value which are used in the construction of LRC.

Year	IBTrACS	CMA	ECMWF	JMA	NCEP	JRA-55
2008	21	19	19	19	17	10
2009	22	20	20	20	14	10
2010	13	13	13	13	13	6
2011	21	19	20	17	19	14
2012	24	23	23	23	23	16
2013	29	28	28	27	28	15
2014	19	12	17	17	17	13
2015	22	20	21	20	21	17
2016	26	25	25	24	25	13
2017	30	28	29	23	29	9
Total	227	207	215	203	206	123
Detection Rate		91.2%	94.7%	89.4%	90.7%	54.2%

Table 4. (From the left) Annual n~~N~~umber of historical TCs in IBTrACS (second column); Annual number of historical TCs detected in the respective forecast models ~~with at least one MEPS counterpart identified~~ (third to sixth columns); Annual number of historical TCs detected in JRA-55 (seventh column).

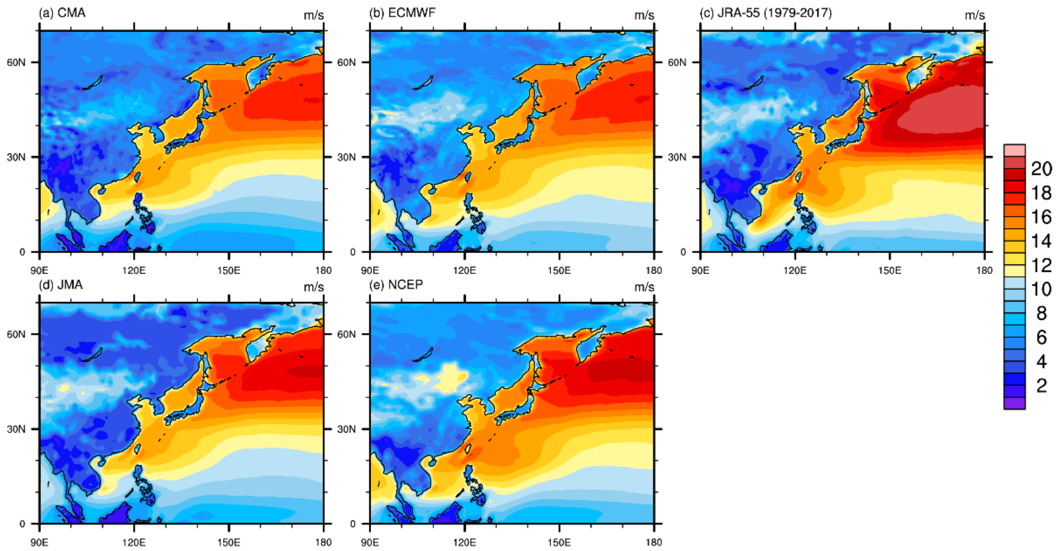
Centres	Number of TC windstorms	Number of Pure EPS TCs	% of TC windstorms as pure EPS TCs
CMA	39535	13322	33.7
ECMWF	215737	74091	34.3
JMA	56537	14964	26.5
NCEP	203903	96052	47.1

Table 5. Statistics of TCs in the selected TIGGE data.

Intensity Class	IBTrACS	JRA-55	TPEPS
TD&TS	252	32	27643
STS	208	126	70759
TY	231	254	69794
VSTY	231	193	23686
VTY	85	63	6547
Total	1007	668	198429

Table 6. Number of TC records in IBTrACS, JRA-55 TC event set, and TPEPS TC event set, for different intensity classes. The classes are Tropical Depression (TD) and Tropical Storm (TS), Severe Tropical Storm (STS), Typhoon (TY), Very Strong Typhoon (VST), and Violent Typhoon (VTY). The intensity classes for IBTrACS are defined according to WMO (2019). The intensity classes for JRA-55 TC and TPEPS TC are derived from the WMO (2019) intensity classes by using quantile mapping of intensity records of JRA-55 TC and IBTrACS records.

Figures



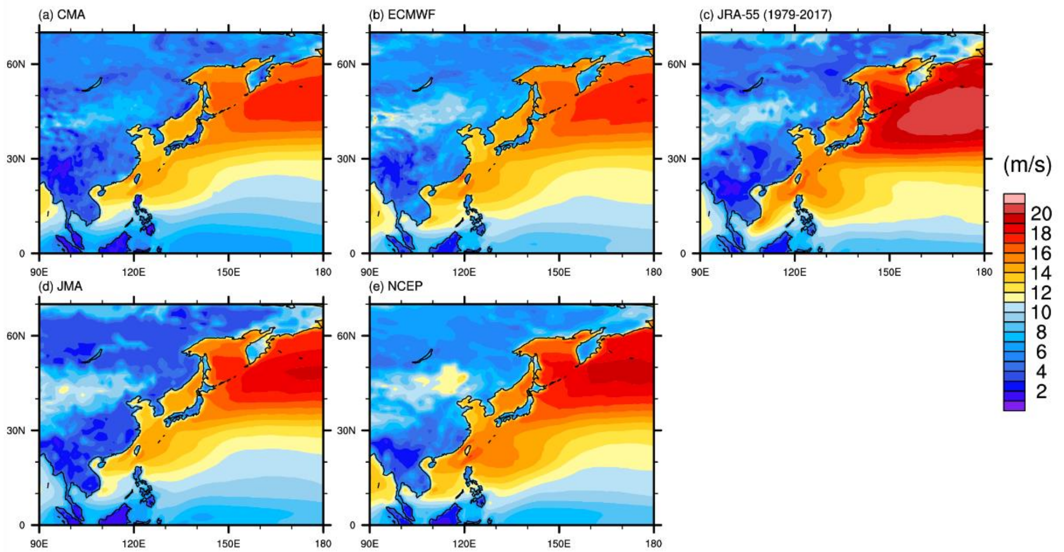


Figure 1. Local 98th percentile wind speed for each grid box in the region for TIGGE: (a) CMA, (b) ECMWF, (d) JMA, (e) NCEP, and (c) JRA-55.

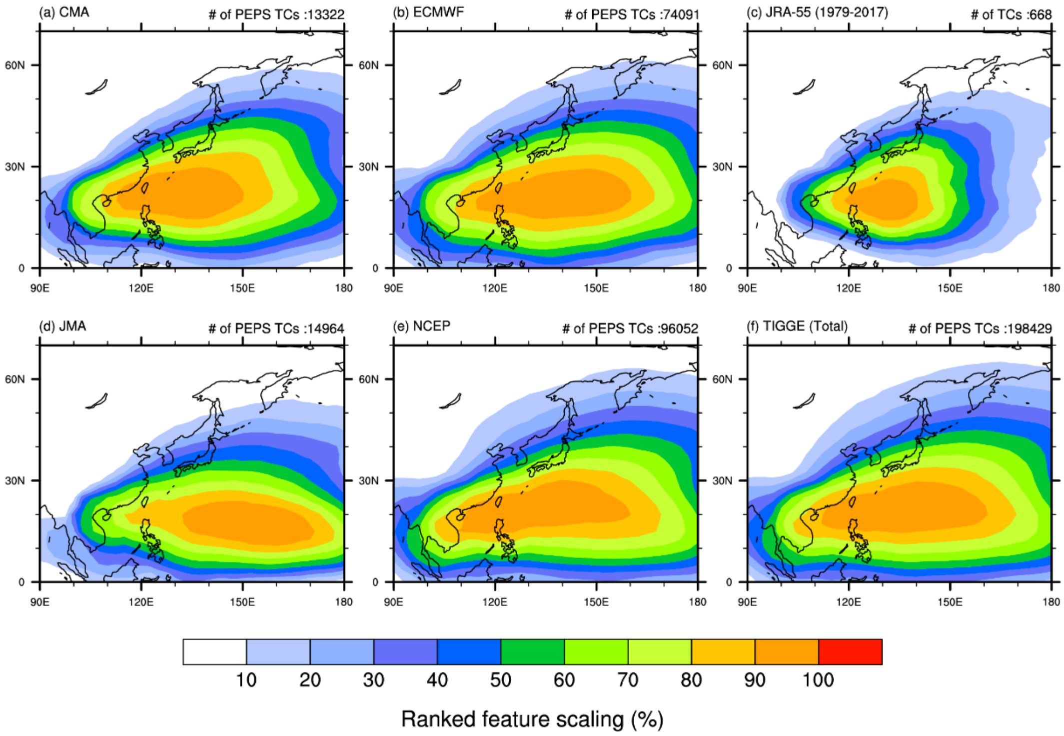


Figure 2. Ranked feature scaled track density (%) of different data sets: (a) CMA, (b) ECMWF, (c) JRA-55, (d) JMA, (e) NCEP, and (f) TIGGE total. Number of TCs in the corresponding event set is stated on the top right of each panel.

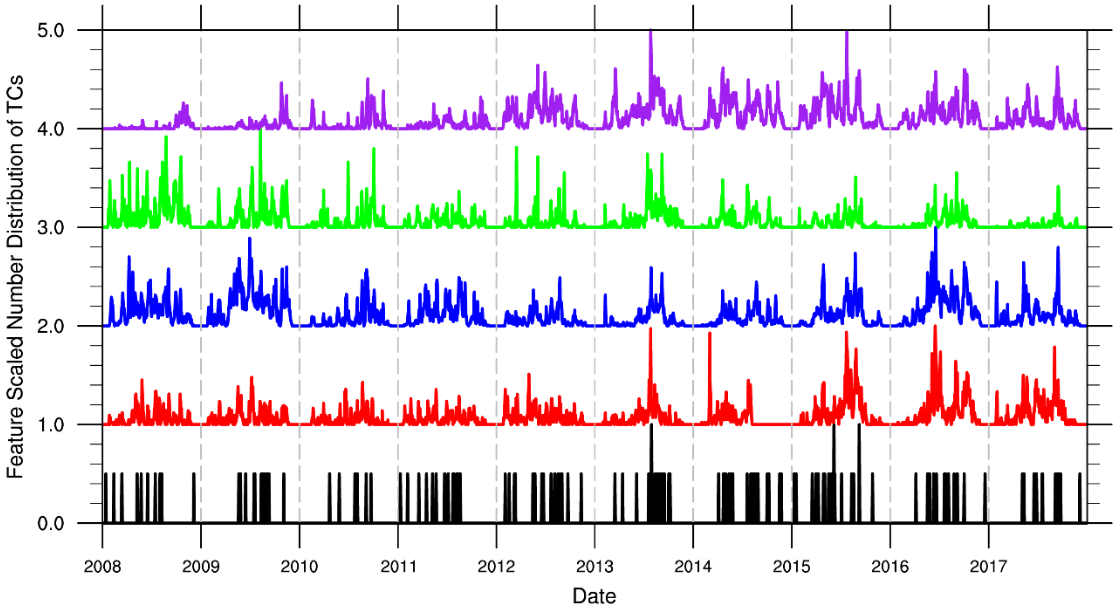
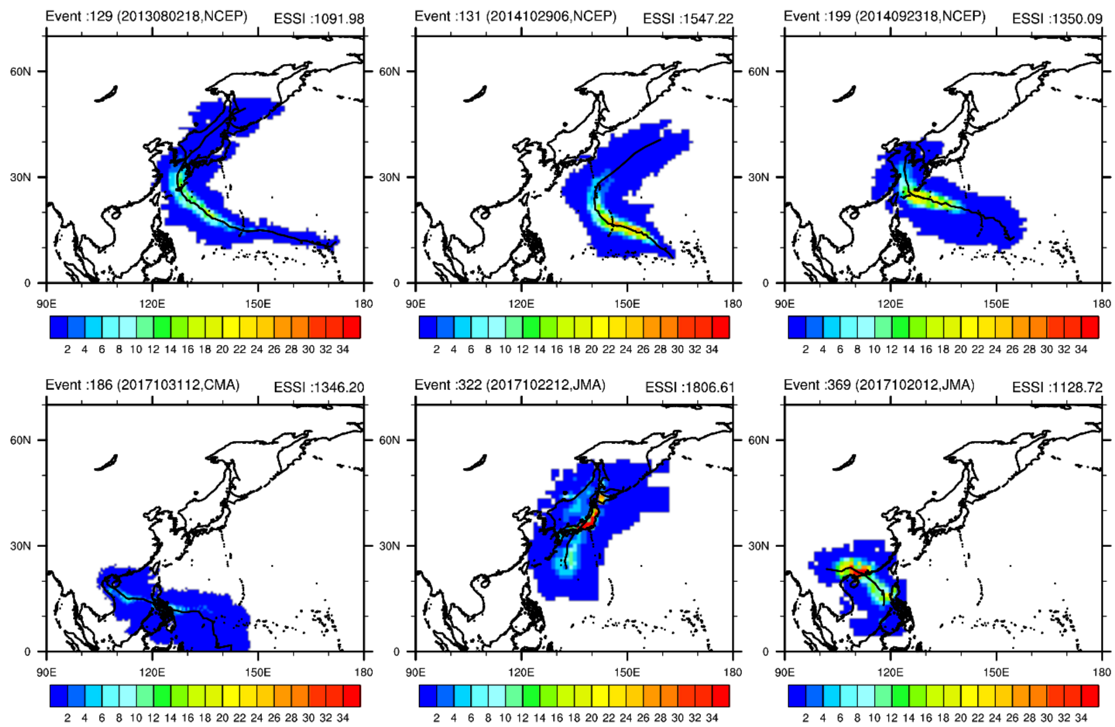


Figure 3. Feature scaled time series of number of TCs which are first identified in each day~~formation~~ in the~~of~~ TPEPS TC event set (CMA: red, ECMWF: blue, JMA: green, NCEP: purple) and JRA-55 event set (black). For visual convenience, the time series of CMA, ECMWF, JMA, and NCPE are shifted by 1, 2, 3, 4, respectively.



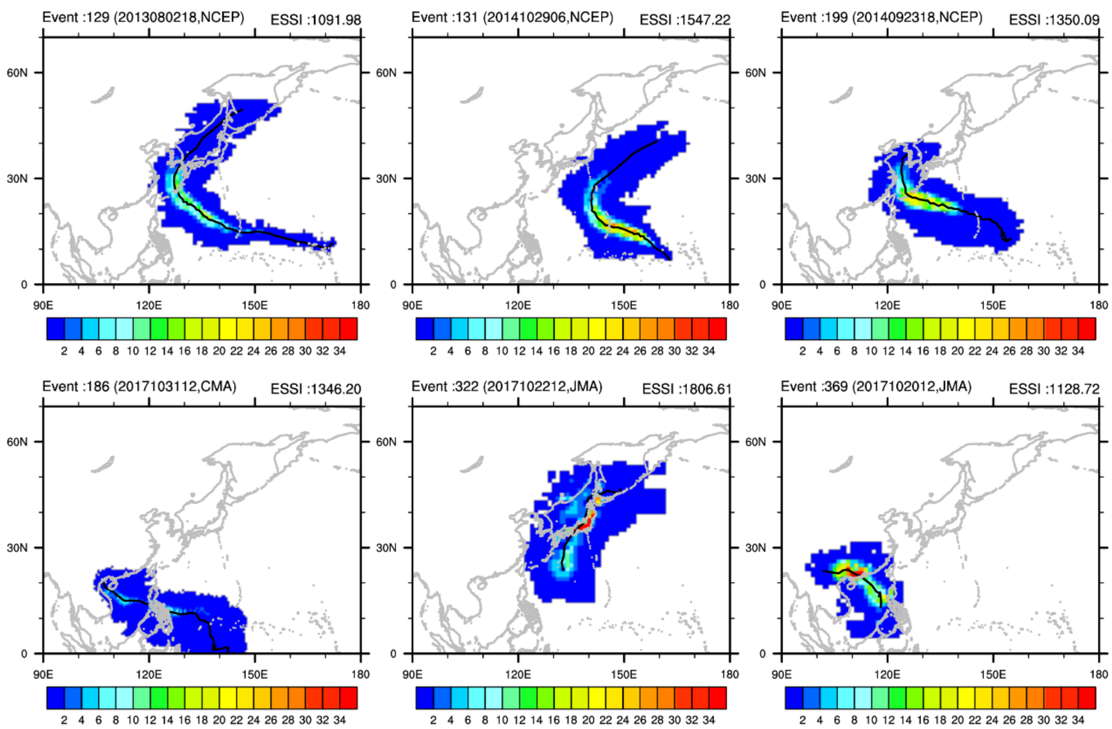


Figure 4. Some of the PEPS TC impact footprint (colour contours) and tracks (black line within the colour contours) of the TPEPS TC event sets. The colour contours show the cumulative SSI of the PEPS TCs over their respective lifetime at individual grid box. ESSI of each PEPS TC is shown on the top right of each panel.

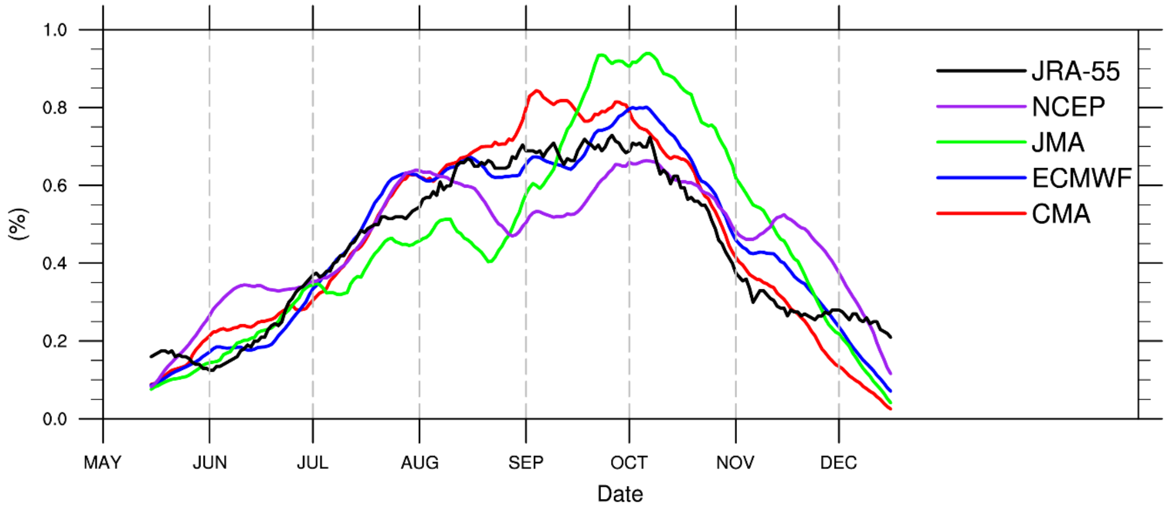


Figure 5. Climatological daily number distribution of TC first detection for TPEPS TC event set (CMA: red, ECMWF: blue, JMA: green, NCEP: purple) and JRA-55 event set (black), i.e. the probability of TC being first detected at a given day in the model. 30-day moving average is used in order to remove high frequency signal.

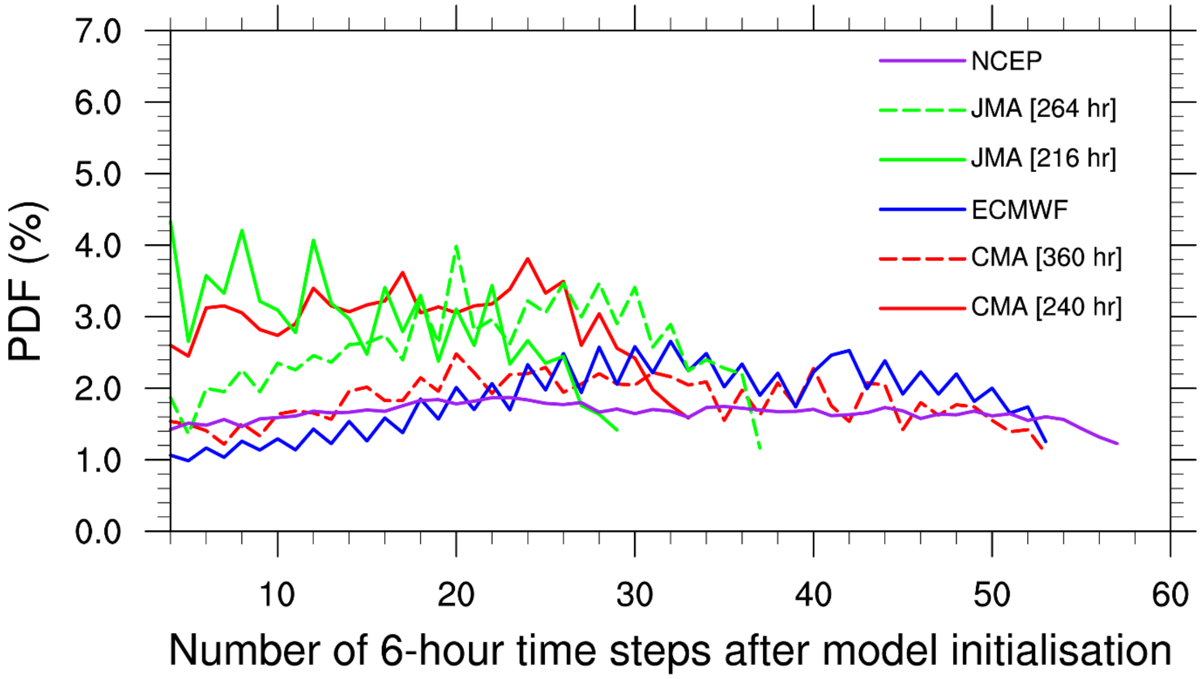


Figure 6. Temporal evolution of frequency of first storm detections of TPEPS event set (CMA: red, ECMWF: blue, JMA: green, NCEP: purple).

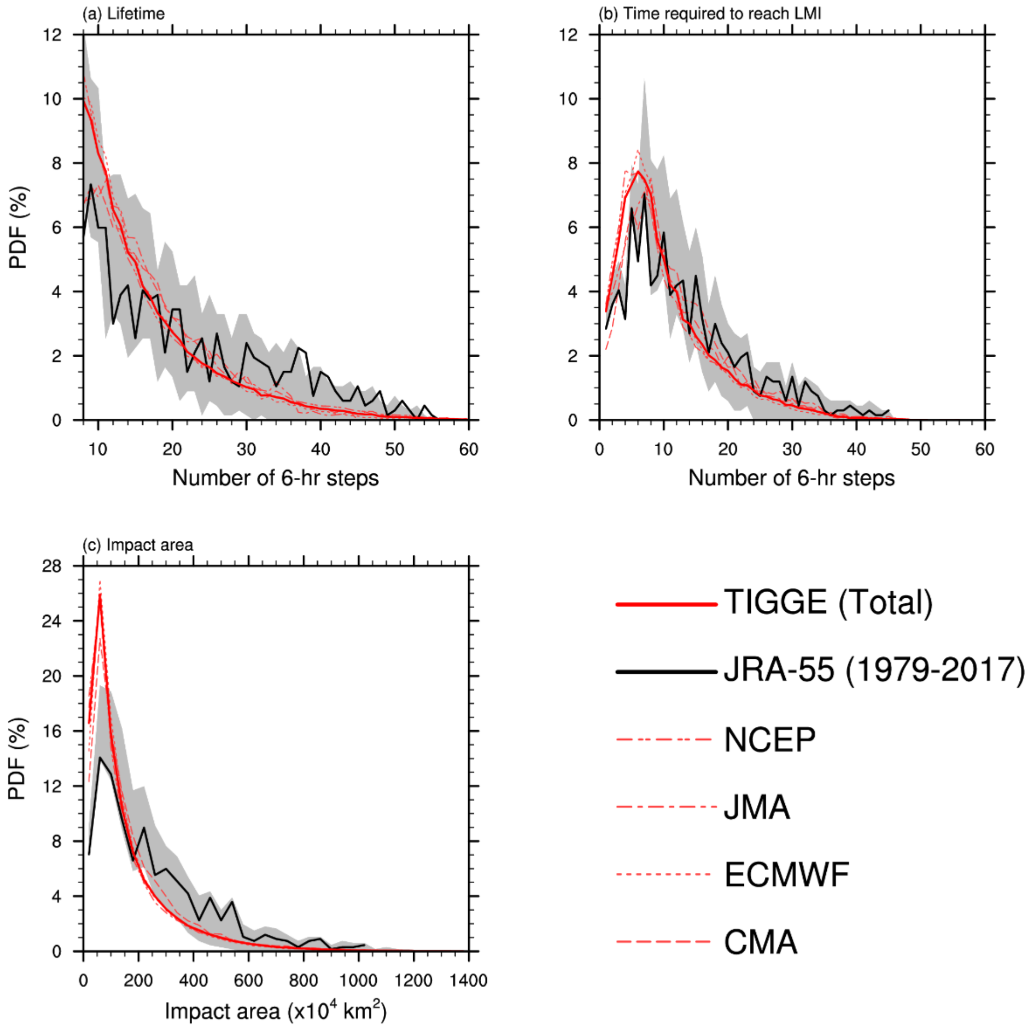
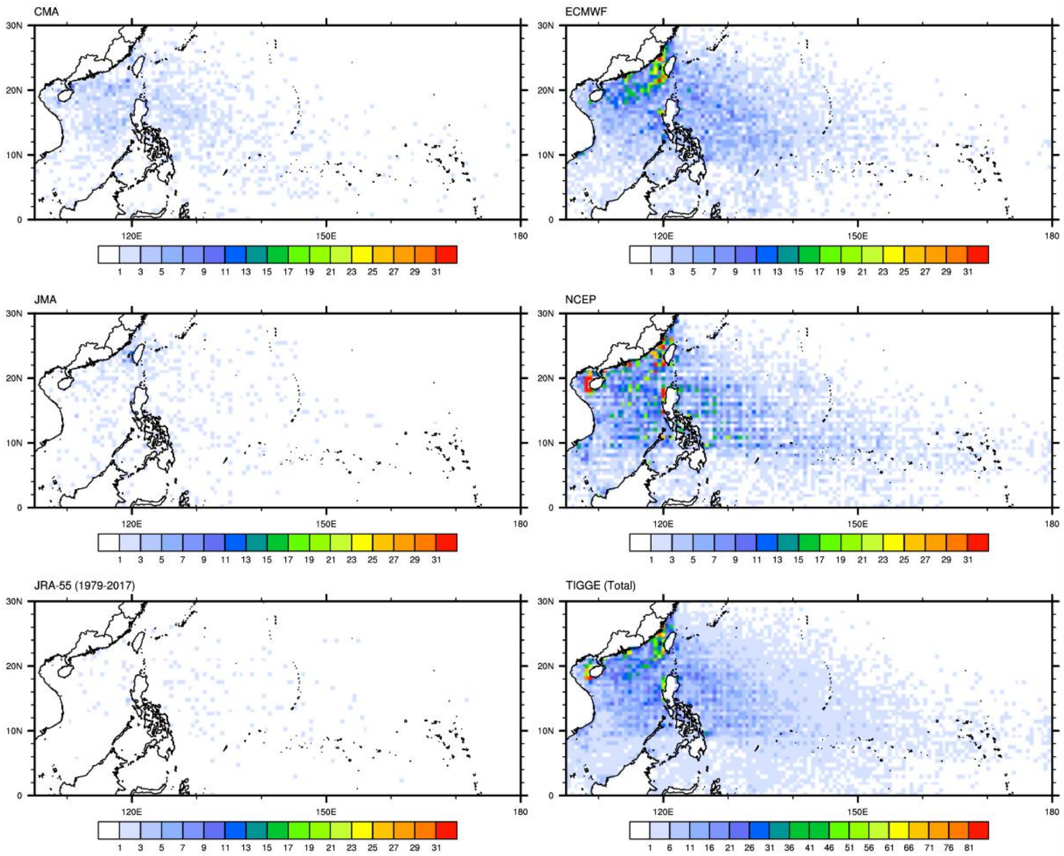


Figure 7. The distribution of (a) lifetime, (b) time required to reach LMI, and (c) impact area of TCs in TPEPS TC event set (red lines) and JRA-55 event set (black line). The grey area indicates the spread of the lifetime distribution of JRA-55 if finite simulation windows are applied to the JRA-55 event set.


CMA
ECMWF
JMA
NCEP
JRA-55 (1979-2017)
TIGGE (Total)

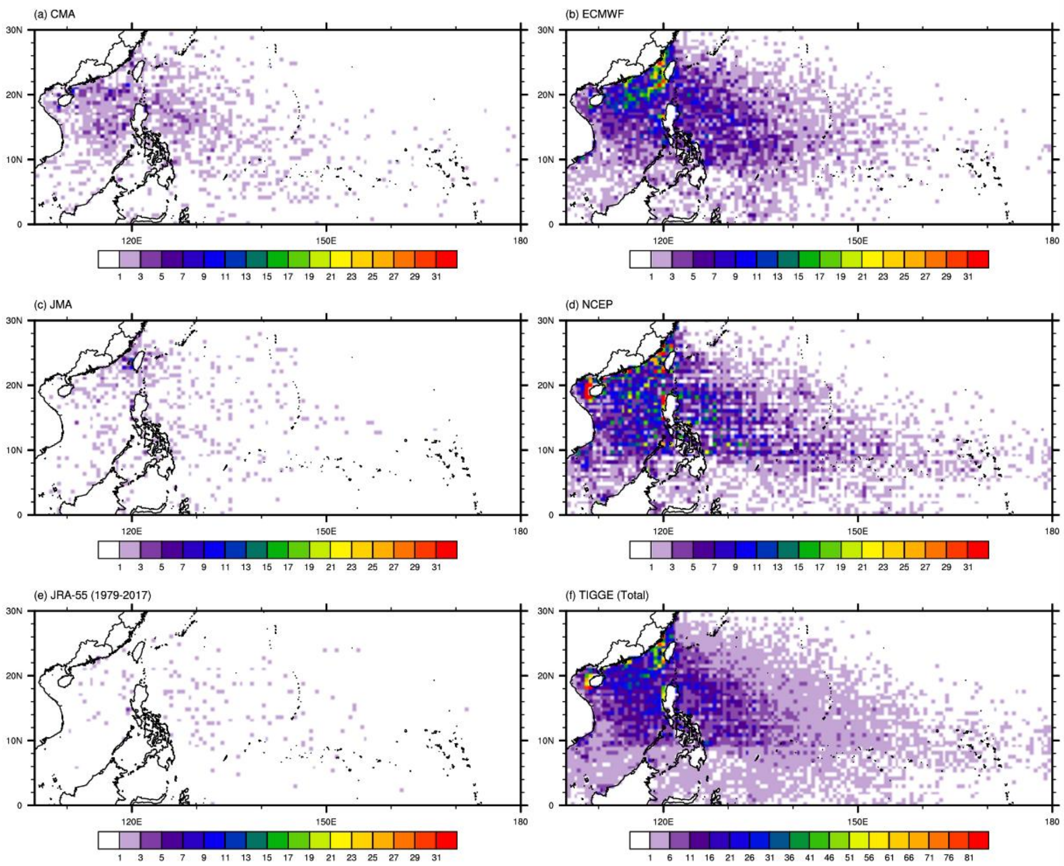

(a) CMA
(b) ECMWF
(c) JMA
(d) NCEP
(e) JRA-55 (1979-2017)
(f) TIGGE (Total)

Figure 8. The spatial distribution of location of first detection of TCs (with LMI at least typhoon strength) which made landfall within the domain 105-180 °E, 0-30 °N for TPEPS TC event set and JRA-55 event set.

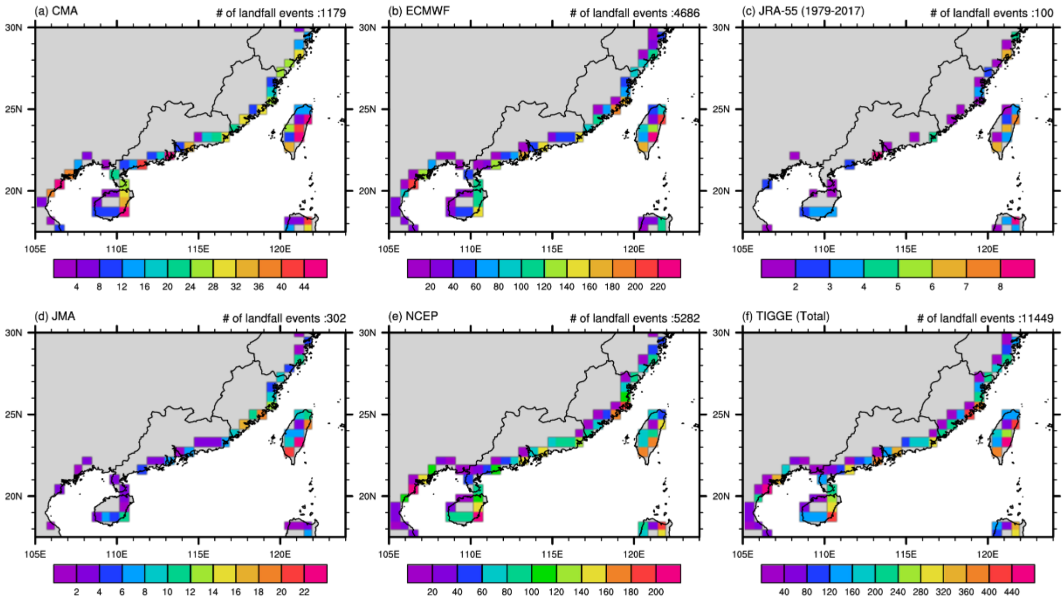


Figure 9. Spatial distribution of number of landfall events (landfall with at least typhoon strength) for TPEPS TC event sets and JRA-55 event set (colours). The total number of landfall events in each panel is shown on the top right of each panel.

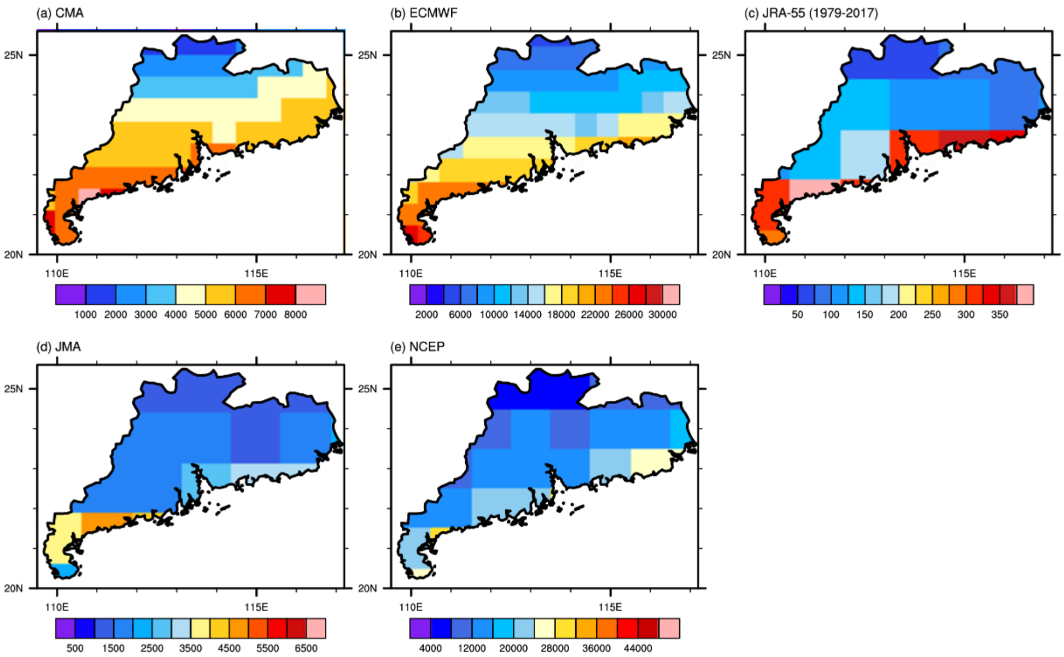
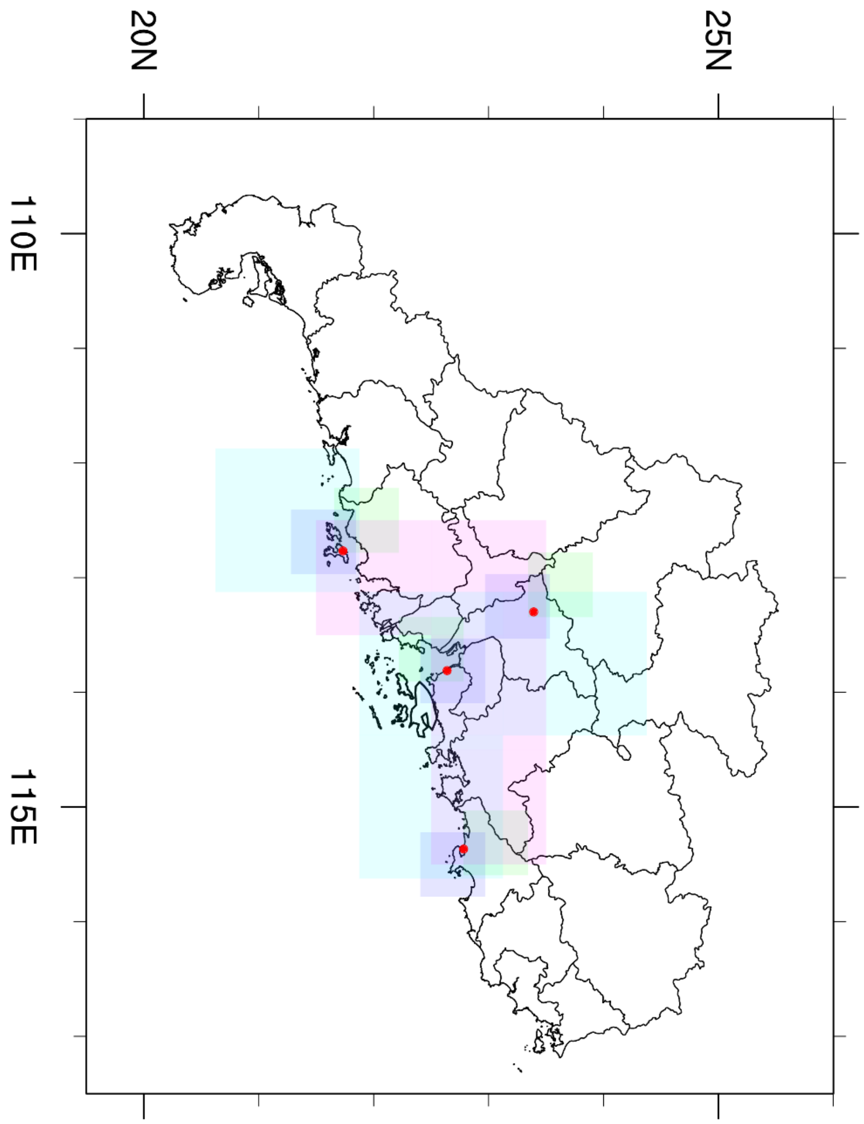


Figure 10. Number of TC-related 6-hourly data entries in each of the grid box in Guangdong Province, China, for TPEPS TC event sets and JRA-55 event set.



809

810 **Figure 11**. Locations of the selected surface observation stations (red dots) in Guangdong, China with
811 corresponding grid boxes from 4 EPS outputs: CMA (green), ECMWF (blue), JMA (cyan), and NCEP (magenta).
812 Information of prefectural boundaries is obtained from GADM version 3.6 Level 2 (available at
813 https://gadm.org/data.html)

814

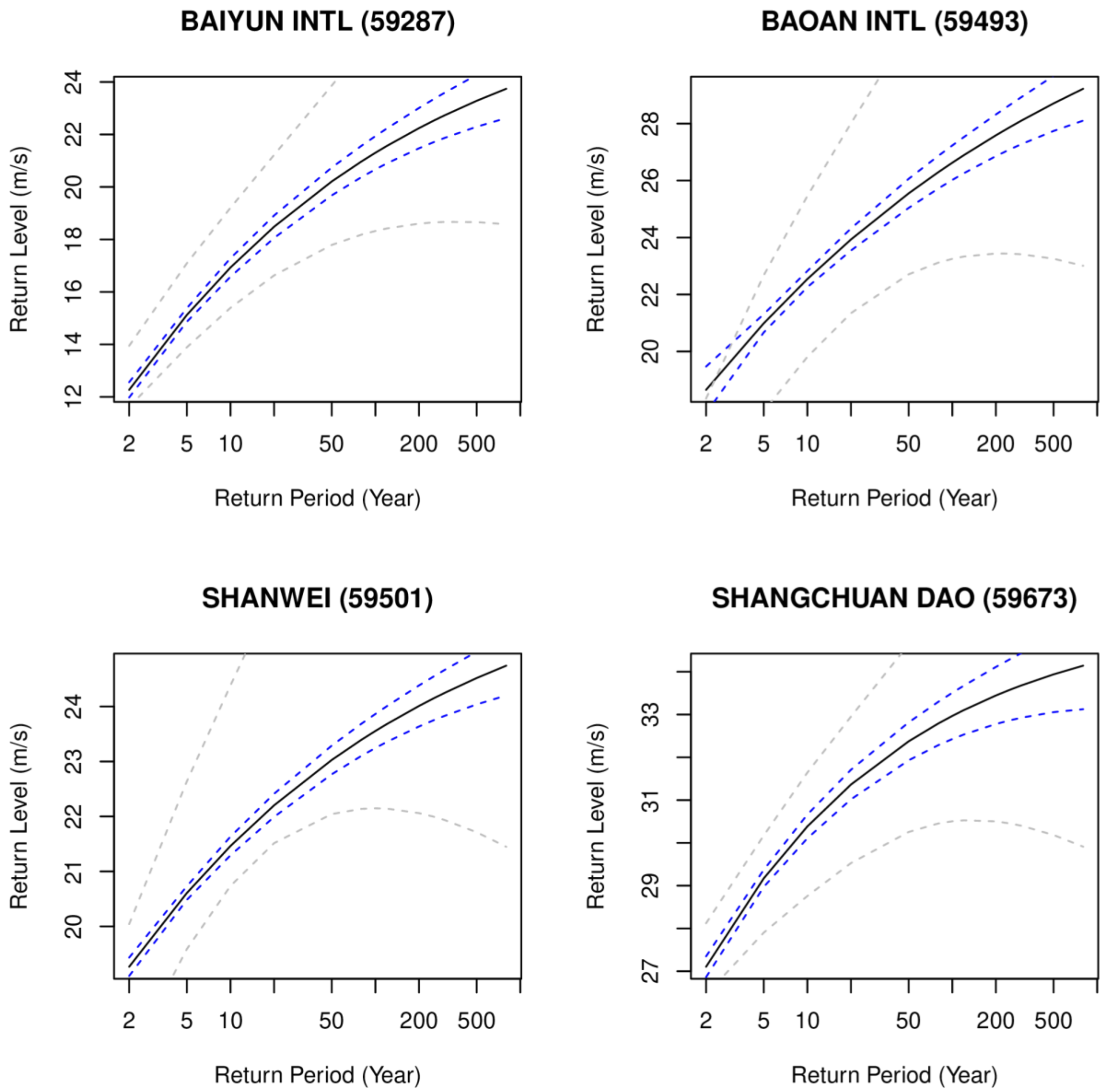


Figure 12. Return period-return level plot for 4 selected surface observation stations: Baiyun International Airport, Baoan International Airport, Shanwei, and Shangchuan Dao. Black lines indicate the best estimate of return period-return level. Blue lines indicate the 95% confidence interval calculated using TIGGE PEPS event set. Grey lines indicate the 95% confidence interval calculated using in situ observations.