First of all, the authors want to thank the referees for the work and time devoted to review the manuscript. We know that all comments will serve to improve the quality and understanding of the work and we hope we have properly answered all the suggestions.

Lines are referred to the track-changes manuscript.

Reviewer #1:

The presented topic is quite interesting not only for local populations and policy-makers but also for the international research community. The authors presented a methodology that uses results from the CORDEX database to force a hydrological model that will feed a 2D hydraulic model. The methodology seems to be coherent and appropriated for the obtained results. The authors were careful in choosing the forcing conditions from the CORDEX project, comparing the historical results with observed precipitation data in the region. The numerical models selected are adequate, being well known tools fully developed and validated for different regions. The manuscript is well written, easy to follow and to understand, with up to date references. The authors seem to have previously experience with the topic, the region selected and the presented numerical models, which were previously configured in already published manuscripts.

However, there are some comments/issues that should be clarified.

The authors state in the introduction that "This provokes that the hydrological cycle regimes are mainly conditioned by the timing and position of winter storms, which in turn, are dependent on the NAO phase". However, is worthy to notice that there are previous studies demonstrating that the precipitation regimes in the Iberian Peninsula are not only dependent of a single atmospheric variability mode.

This information was added and this part of the text was rewritten (lines 47-55).

The authors need to clarify the methodology section. Initially, the reader understands that the hydrological model will be forced which each one of the models that provide a good characterization of the precipitation for the study area, and the authors present Figure 3 and Table 2 with the individual results for each selected model. Then, the reader realized that an ensemble was constructed with the best CORDEX models and this ensemble will force the hydrological model. It is important to use an ensemble because, since the hydrological model will be forced with results for another numerical model, the ensemble will avoid numerical inconsistencies and reduce the inaccuracies in the hydrological and *hydraulic models. The early in the calculation that the errors are minimized, the smaller inaccuracies obtained. However, it will be necessary to include how the ensemble was constructed. It is a simple average or the authors considered a weighted mean? The weighted mean could provide more accurate results by considering the previous performance of the CORDEX numerical models in the weight.*

The hydrological model was forced with the precipitation of each individual CORDEX model that surpassed the validation test. It should be noted that calculations were carried out for any particular model and the final results were averaged. Thus, the ensemble was constructed after hydrological-hydraulic simulations, for the results, by averaging the individual results provided by the hydrological (and hydraulic) model forced with each CORDEX model. In addition to the information provided by each model individually, the ensemble of the results allows offering a global view of changes minimizing the inaccuracies and errors as commented the reviewer. In climatological studies, the outputs of different models are never averaged prior to be used to force hydrological or hydraulic models. We have to take into account that several climate models do not reproduce the same meteorological situation at the same instant. Please, remember that weather models make predictions over specific areas and short timespans, while climate models analyze long timespans to predict how average conditions will change in a region over the coming decades.

As for the use of a weighted mean, it is difficult to establish the weights to carry out such a mean. The accuracy of the models was assessed in terms of two metrics and the "closely to reality" of the different methods was observed to depend on the metric, the area where the rain gage is located and the season.

This was clarified in the manuscript (lines 154-157 and 169-171).

The authors should better describe the procedures to construct the forcing with the CORDEX data. In the methodology it was not specify the version of the CORDEX project data. It is CMIP5 or CMIP6? Having a full range of numerical predictions, why the authors only selected the RCP8.5? If possible, it will be interesting to compare with a not so extreme scenario (RCP 4.5, for example).

The data from the CORDEX project used in this work were those corresponding to the CMIP5. This was clarified in the new version of the manuscript (lines 98-100).

Regarding the numerical predictions, we opted to analyze the most extreme scenario, the RCP8.5, in order to evaluate the most extreme changes and implications that the climate change can cause in the study area in terms of floods. In this sense, a more conservative perspective towards the worst scenario is preferable for this type of applications and for the development of possible mitigation-adaptation measures. Although the comparison between different scenarios can also be interesting, authors consider that it is out of the scope of this study. Moreover, some of the CORDEX models validated for the study area have not available the RCP 4.5 prediction. In addition, in the report published by Schwalm et al. (2020) in Proceedings of the National Academy of Science, they found

that, since the RCPs were developed, the historical evolution has been closest to that worst-case pathway. Thus, for the past 15 years, the greenhouse gas emissions have tracked most closely with those projected under RCP 8.5. To sum up, the worst case also seems to be the more realistic. Therefore, we decided to maintain the current analysis focused on RCP8.5 predictions. However, we acknowledge the reviewer for this recommendation because this is an interesting topic that may be addressed in future works.

This information was added in the new version of the manuscript (lines 96-104)

It is not also clear the forecasting period. The authors referred that they used historical (1990-2019) and future (2070-2099) periods, and that the data has an hourly scale. However, it is not clear if the historical simulations and the future projections were done for an specific year or if the authors calculate an average for the full period. For historical conditions, using an average period to compare with the observed data is acceptable. However, a difference of 30 years in the projections could produce significant differences in the results.

All numerical simulations were performed for the historical period, considering the entire 1990-2019 period, and also for the future, considering the entire 2070-2099 period. Therefore, both periods analyzed have the same duration (30 years), in order to maintain the coherence for comparison purposes, as comment the reviewer. Thus, in both cases (historical and future), we run 30 years and compare the results for these periods. This was clarified in the manuscript. Only a shorter period was used to validate the CORDEX models due to the availability of measured precipitation data from pluviometers. In that case, the period 2008-2020 was used to validate CORDEX models with real data from pluviometers. However, in all the simulations, periods of 30 years (1990-2019 and 2070-2099) were always taken into account. This was clarified and specified in the new version of the manuscript (Lines 139-141, 156-157, 162-167).

Why the figure 5 presents the results for the whole year? The authors explained in the methodology that the period that they use to validate the precipitation data was for the wet season (November-March).

The November-March season was used to validate the ability of CORDEX models to reproduce precipitation patterns because it is when flood events occur in the area under study, and therefore, this is the period of most interest for the scope of this study. In fact, figure 5 corroborates this point; the flood events only occur during the November-March period. Once the models were validated, the complete years were simulated in the hydrologic procedure. Taking advantage of the information obtained, for sake of clarity we show the whole year in figure 5. This also allows corroborating when the flood season occurs and its evolution.

The authors mentioned that "The developed procedure takes between 2-3 weeks to execute each model". Please, include the characteristics of the computer used to run those models.

The simulations were executed on a computer with an AMD Ryzen 7 2700X processor, 32GB of RAM and a Nvidia RTX 3080 ti GPU. This information was included in the new version of the manuscript (lines 173-174).

I recommend to the authors to include the limitations of the study. Is worthy to notice that there are several factors that could conditioning the future river flow that will reach a specific region, and not only the precipitation. The authors are representing the natural flow, but not changes in the man-made interactions with this flow. Changes in the aquifer capacity, in the river margins, in the soil characteristics, in the water use or in the hydroelectric production, among others, are non-easy predictable factors and will not be reproduced by the numerical models. However, they can have strong impacts in the floods.

We agree with the reviewer. The inclusion of the limitations of the study can help in the development of future studies. We add this information in lines 329-335.

Figure 1c should include the latitude and longitude

Done.

Figure 2: Future evolution of river flow risk instead of "risk river flow" and at Ourense city instead of in Ourense city.

Done.

Reviewer #2:

The study addresses the analysis of the future evolution of river floods in the city of Ourense (NW Spain), where flooding of the Miño river can cause significant damage. In particular, the historical and future precipitation data from the CORDEX project are used as input in a hydrological model (HEC-HMS) which, in turn, feeds a 2D hydraulic model (Iber+). For each model, hydrological simulations were carried out considering both historical (1990-2019) and future (2070-2099) periods.

Major comments follow.

In the Introduction the novelty of the study with respect to the state-of-the-art knowledge must be emphasized and the main objectives of the study must be better clarified.

We agree with the reviewer. The novelty of the study and the main objectives have been clarified in the new version of the manuscript (Lines 58-76).

Although in principle, the methodology seems appropriate, several details must be added to let the reader evaluate the correctness of the adopted approaches. In particular, the following key points should be better explained.

The capability of the EUROCORDEX RCMs models to represent precipitation over the area under investigation was tested by comparing RCMs precipitation data and field data by analyzing the entire distribution of precipitation data through the Perkins' test and also the extreme precipitation values through the P99 test. I assume that the Perkins' test is sensitive to the choice of the bin size and, in turn, the number of bins used to calculate the PDF. The authors should provide additional details on the test metrics and comment on this point, as well as on the advantage of this method with respect to statistical measures, such as bias, root mean square error, correlation, and trend analysis, commonly used to quantify model performance (see for instance doi.org/10.5194/nhess-20-3057-2020).

We agree with the reviewer that there are several analyses that can be suitable for validating the performance of climate models, depending on the scope of the study. Some works, such as the article recommended by the reviewer, use some statistical parameters such as bias, root mean square error, standard deviation, or mean values, among others, to validate climate models. However, these statistics are usually applied at monthly, seasonal or annual scales. In addition, some of them (e.g. means, standard deviations…) do not provide information of the entire data distribution, and therefore, a good fit in these statistics do not guarantee the adequate determination of some patterns of the data, which can have an important impact on the hydrological procedure. To develop our study, focused on flood

analysis, we need to use the best available temporal scale, since floods are highly dependent on daily or even more precise time scales, and also corroborate the good determination of precipitation patterns, especially those referring to extreme events. Therefore, we need to validate those CORDEX models presenting a good skill to reproduce precipitation in terms of daily scale. In addition, as discussed in Perkins et al. (2007), the monthly, seasonal or annual analysis can hide biases or systematics errors that can be detected on the daily scale. Therefore, for the reasons commented above, we opted to maintain the validation of the CORDEX models using the PDFs, since, in addition, if the model is able to simulate an entire PDF, this also demonstrates the capability to deal with rare or very extreme values that can become more common in the future, as explain in Perkins et al. (2007). Therefore, we consider that this procedure is adequate for the purposes of our study. As for the number of bins, 20 bins were considered in the study (Table 2). Results show to be independent of the number of bins. We also complemented the test based on PDFs with the statistical analysis focused on analyzing the deviation of CORDEX models when representing extreme values, those able to cause flood situations. We consider that these analysis methods allow an adequate validation of the CORDEX models for the purposes of this study. Following reviewer recommendation, we provide additional details of the validation procedure, clarifying and explaining better this selection in the new version of the manuscript (Lines 139-153).

The transformation of precipitation into the corresponding river flow was carried out using the semi-distributed model HEC-HMS. The authors should provide additional information on the hydrological model used for rainfall-runoff transformation (including the loss method for assessing the net precipitation). Also, please explain how the historical and future flows of the river were obtained on an hourly scale, given that precipitation data were at the daily scale.

Additional details on the hydraulic modeling used for flood mapping are also required.

Additional information related to hydrological and hydraulic models was provided in the new version of the manuscript (see section 2.3 Hydrological and Hydraulic Models).

The process for obtaining the river flow on an hourly scale was clarified in lines 154-160.

Minor comments

A table summarizing the physical features of the catchment (mean slope, altitude, river length, time of concentration) and a land cover map must be added.

Done (See new Table 1, Figure 2 and lines 82-88).

Please provide additional details on the catchment schematization within HEC-HMS (i.e., number of sub-catchments, connections among them and so on).

A map with the catchment schematization was added (new Figure 3).

Please clarify the meaning of "supreme water depths".

To analyze the expected changes in maximum water depths reached in specific areas of the city subjected to floods, we determine the maximum water depth reached each day under flood conditions, and then, the average of these maximum values is calculated (mean of the maxima) and also the absolute maximum (highest value) is determined (referred as supremum value in the new version of the manuscript). Thus, the supremum value is referred to the highest water depth reached by water in each specific area taking into account all the days under floods. This was clarified in the text (lines 286-290).

L 70: replace "which supposes" with "encompassing".

Done.

Addendum to my previous comments It seems that the authors applied continuous hydrologic modeling with HEC-HMS. In this case, however, further input variables are needed, such as temperature. Please clarify!

The methodology applied do not need the input of more variables since is based on the Curve Number methodology, which requires only the precipitation data. This was clarified in the text (lines 114-118).