REVIEWER #1

1.- “As mentioned by the authors in the introduction, EWS have been implemented in recent years at different scales and using different approaches. The authors should mention the differential aspects of the proposed Early Warning System (EWS) with respect to other systems that have been proposed, some of which are mentioned in the introduction (lines 45-75).”

As mentioned at the end of the manuscript (lines 411-415) the main advantages of the system can be summarised as follows “The Early Warning System can be easily adapted for any area of the world since the required input data can be obtained freely from public institutions and the models to compute the hydrological and the hydraulic processes (HEC-HMS and Iber+, respectively) are both freely available. Therefore, the EWS is especially interesting for developing countries where the acquisition of commercial software is not sustainable.”

This is also mentioned in the introduction “. The main contribution of the EWS presented in this work respect to the systems …” (lines 104-106). In addition, an ensemble of precipitation forecasts obtained from different institutions can be easily implemented due the capabilities of the python script that orchestrates the components of the system.

In summary, every part of the system is free, even weather forecast is free and publically available for the particular case of Galicia and North of Portugal. So, any public or private entity can implement the system which runs on a desktop PC equipped with a GPU. The cost of the whole equipment does not exceed € 1000.

2.- More details about the model implementation at the Miño basin should be given in lines 169-176. Some details that might be of interest to the reader are: infiltration model and parameters, formulation and parameters used to compute rainfall-runoff transformation, formulation and parameters used for runoff propagation in the stream network, number and size of subbasins, estimated characteristic times of the watershed and subbasins(as lag time or concentration time), is it baseflow considered? With which formulation?

More details about the HEC-HMS implementation at the Miño basin were included in the new version of the manuscript (line 199-211). For the season under study (January), baseflows lower than 25 m^3s^{-1} were obtained, which suppose around of 3% of the peak flow. For this reason, baseflow was not considered because its influence is negligible in this type of events. The methodology proposed is good enough to calibrate the parameters proposed (curve number and lag time) to obtain an accurate representation of the real events. This is mentioned in lines 211-213.

3.-HEC-HMS is a semi -distributed hydrological model, not a distributed model as stated in line 166. With the implementation of the authors (the whole watershed of 2200 km2 is divided in only 3 sub-basins) it should even be considered as an aggregated model.

According to the comments of the reviewer, the hydrologic model is re-defined as a semi-distributed model since the entire catchment is divided into three sub-basins with different characteristics. This new definition is included in line 189.

4.- How is the hydraulic inundation model initialised at the beginning of the flood hydrograph (with which discharge and/or water depths)?

The initial flow and water level for each flood event are obtained from the gauge station located at Lugo three days before the flood event. These values are provided to the hydraulic model Iber+ to run the simulations. An explanation of this procedure is included in lines 322-325.

5.- In section 4.1. the authors discuss the accuracy of the precipitation forecasts. This is very interesting since it is one of the main reasons that might explain the performance of the whole system. The authors evaluate the accuracy for a rainfall aggregation time of 24 hours. I guess they have chosen this aggregation time taking into consideration the concentration time of the watershed. However, it would be interesting to show the accuracy of the precipitation forecasts for smaller aggregation times, as for instance 12 h, 6 h, 1 h, since those times are more relevant for smaller basins.

We have only tested the accuracy of the forecast precipitation under a 24h forecast window for several reasons. First, MeteoGalicia performs its own quality tests to compare forecasts and real data, showing a good agreement at different time scales. So, the detailed analysis of accuracy at all time scales is beyond the scope of the present paper and left to the data supplier (MeteoGalicia). Second, the recorded cases linked to flooding events correspond to persistent rains (several days) with long lag/aggregation times. Shorter aggregation times can play a key role in flash-flood areas but they do not represent a critical issue for the area under study.

Finally, the main goal of the paper is the design of a EWS able to provide fast and accurate information to report alert situations and help to diminish the damage caused by floods. This goal can be achieved regardless of the source of data. In our particular case MeteoGalicia seemed the logical choice because it is the local forecast agency and provides the most detailed resolution for the Miño area (4 x 4 km).

For these reasons, we think that the accuracy analysis performed is enough taking into account the objective of the study as well as the characteristics of the area under study. A brief sentence is added in lines 259-262.

6.- How does the calibrated Lag Time in Table 3 relate to empirical formulas based on the basin size and slopes?

The values of CN and Lag time were computed using particle swarm algorithm to minimise the error of the measured streamflow and the numerical one. In this way, the characteristic values of the sub-basins are more representative of the real sub-basins. An alternative way would be the use of empirical formulas to compute the characteristic values of the curve number and lag time of the three sub-basins. However, these formulas are very dependent on several parameters which can be implemented in different ways, which means that the values obtained using different formulas can be very different, as shown Fang et al., 2008; Upegui and Gutierrez, 2011 and Grimaldi et al., 2012. A sentence is added in lines 288-293.

7.- Figure 7 shows a comparison between the water depth measured and computed at certain locations that are shown in Figure 6. Were those field values measured (and how)? Or are they values estimated by visual observation and/or photographs of the inundation? At what time during the flood were they measured/observed? Are they maximum values or values at a specific time? Those questions should be clarified in the paper.

These field values were obtained comparing photographs taken from local press and social media in January 20th, 2013 with measurements obtained in situ by the authors. These values do not correspond to the peak value of the flood event. In fact, the peak value was registered on 21th, 2013. This was reproduced by the numerical simulations. This information was included in lines 337-339.

8.- In Figure 7 the authors present a range of variation of the numerical values, but this ranges are computed rather arbitrarily (3 times the standard deviation of the water depths from 12 am to 4 pm). Why has this criterion been chosen and how does it relate to model output uncertainty? If this is not discussed, the range bars shown in Figure 7 are meaningless for model evaluation since they can be as large or small as one wants by just changing the criterion.

As we mentioned above, volunteers and local media provided in situ photographs, which creates some uncertainty about the exact moment they were taken. Timing was estimated from 12:00 a.m. and 4:00 p.m. following several proxies like light conditions, traffic and activity characteristics,… (see lines 337-341). Therefore, the in situ elevation was compared with all numerical simulations in that time interval, which allows calculating their mean value and standard deviation. Assuming that elevations correspond to a data set that follows a normal distribution three standard deviations were considered because they account for about 99.7% of the data set.

9.- How is the error in Table 5 defined and which is the probability of the real depth being inside the intervals given?

The goal of former Table 5 was to characterise the overall error between calculated and measured elevations and how that error increases with the forecast window. In addition, the bias is used to analyse whether the model overestimates or underestimates real elevations. Both indices were calculated using the difference between numerical and measured elevations at the five control points shown in Figure 6. Adding an additional error (±) to the RMSE and bias was meaningless and it was removed from the present version. In addition, the definition of RMSE and Bias is provided in lines 243-247. Finally, we have considered that the number of tables was already sufficient and, hence, the information provided in the former Table 5 was moved to the text (lines 348-360), where the accuracy of the simulation is analysed for the three forecast windows.

REVIEWER 2

1) A revision of English and namely of the verb tenses is suggested. Also choose following either British English or American English in the text (e.g., analysed (line 60) vs organized (line 90)).

Done.

Introduction

1) A literature review on available and referenced flood early warning systems (short-term forecast) to frame and compare with the one proposed in the study is missing and is suggested.

Some examples of these short-term local systems focused on river floods are: the River Forecast Centers (https://water.weather.gov/ahps/rfc/rfc.php) in the United States of America or "Sistema de Ayuda a la Decisión" (http://www.chebro.es/contenido.visualizar.do?idContenido=12789&idMenu=2902) developed by the Hydrographic Confederation of the Ebro river (Spain). In Europe the meteoalarm (http://www.meteoalarm.eu/?lang=en_UK) provides advice on exceptional weather events including floods with a temporal window of 48 h. This information was included in the new version of the manuscript in lines 73-80.

As for the difference with other systems, we mention at the end of the manuscript (lines 411-415) that the main advantages of the system can be summarised as follows "The Early Warning System can be easily adapted for any area of the world since the required input data can be obtained freely from public institutions and the models to compute the hydrological and the hydraulic processes (HEC-HMS and Iber+, respectively) are both freely available. Therefore, the EWS is especially interesting for developing countries where the acquisition of commercial software is not sustainable."

This is also mentioned in the introduction “. The main contribution of the EWS presented in this work respect to the systems …” (lines 104-106). In addition, an ensemble of precipitation forecasts obtained from different institutions can be easily implemented due the capabilities of the python script that orchestrates the components of the system.

In summary, every part of the system is free even weather forecast is free and publically available for the particular case of Galicia and North of Portugal. So, any public or private entity can implement the system which runs on a desktop PC equipped with a GPU. The cost of the whole equipment does not exceed € 1000.

A similar explanation was provided to reviewer #1.

2) Several references are made to flash flood studies (lines 49 to 53) and to flash flood EWSs (e.g., line 57). A clarification on the type and range of floods under study is required and suggested.

We agree. A sentence was included in the text between lines 80-90: “There are mainly two kind of floods derived from precipitation events: flash-floods and river-floods…”

Study Area

1) Figure 1 (right) locates it as an intermediate part of a non-identified/delimited supposedly full river catchment. This last shall be clearly identified as the Spanish part of the shared Miño river catchment, and the location (red square) shall be adjusted. Also, it is suggested that the location is done in a full map of the Iberian Peninsula.

Figure 1 was adapted according the comments of the reviewer.

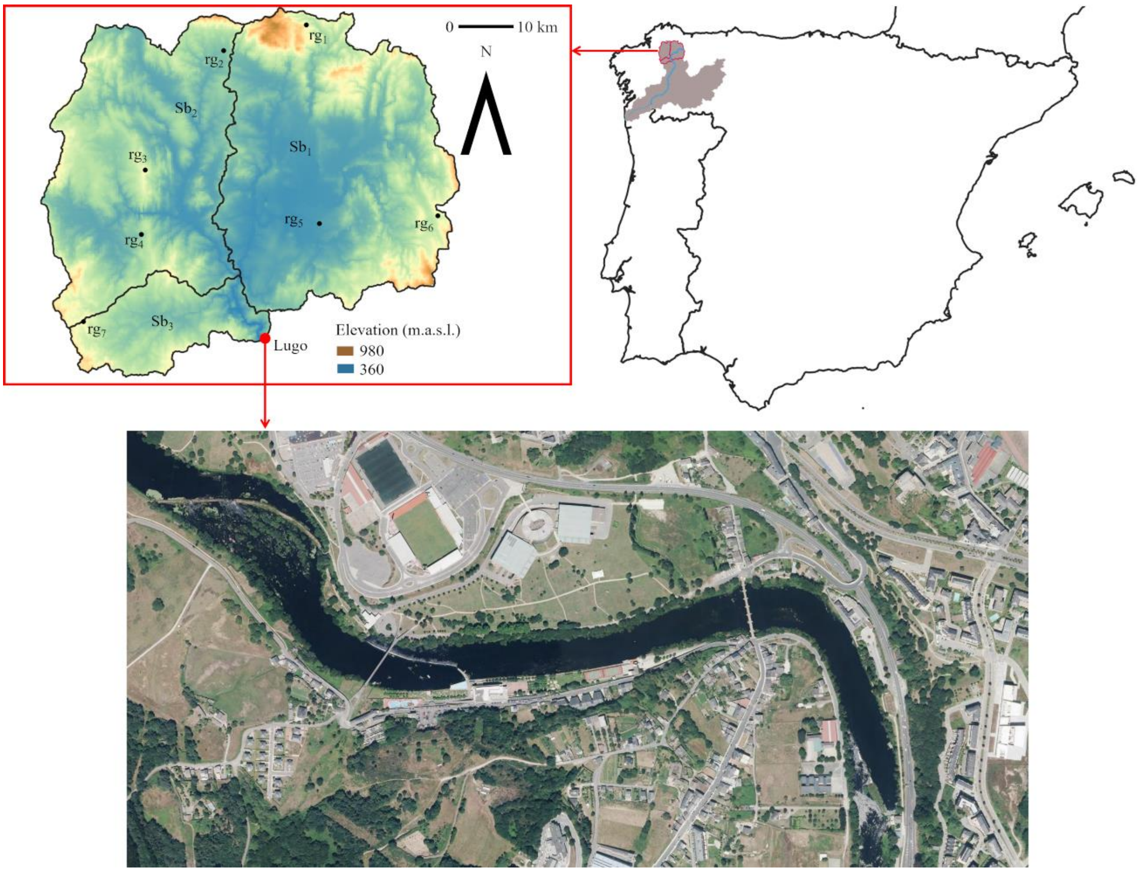

Figure 1. Area of study. In the upper right panel, the location of the entire catchment of the shared Portuguese-Spanish river (shaded area) in the Iberian Peninsula and the riverbed of the Miño river (blue line) are shown. The rain gauges (rg1, …,rg7) located in the catchment and the sub-basins (Sb_1, Sb_2 and Sb_3) of the domain (upper left panel), as well as the area of study in Lugo (lower panel) are also shown. (PNOA courtesy of © Instituto Geográfico Nacional).

2) Information on the type of flow regime is missing and is suggested.

This information was added in the new version of the manuscript (lines 120-125).

<u>Methodology</u>

1) The effect of the initial flow condition (namely for torrential regimes) in the calculations is not discussed in the text. Is that negligible? What was considered as the initial condition (e.g., the water level in the river) in the developed EWS (particularly, in the hydraulic model Iber+)?

The initial flow and water level for each flood event are obtained from the gauge station located at Lugo three days before the flood event. These values are provided to the hydraulic model Iber+ to run the simulations. An explanation of this procedure is included in lines 322-325: “The initial water depth was obtained from the data provided from the gauge station located at Lugo.”

2) Lines 127-128: “the criteria of Cox et al. (2010) are used to define safety limits …”

Why were those criteria considered suitable for that in this study? Accordance to EU regulations?

The European “Floods Directive” 2007/60/EC (FD) (The European Parliament and the Council of the European Union 2007) asks Member States to produce Flood Risk Management Plans (FRMPs) by December 2015, but it doesn’t define a specific criterion to determine the stability of the pedestrians in areas with risk of flood. We used the criterion defined by Cox et al. (2010) because it is widely used in similar works. Also, this criterion is very restrictive because it is focused on the most sensible people. However, other criteria can be easily adapted to the EWS.

Results and discussion

1) Concerning the accuracy of MeteoGalicia precipitation, data was evaluated by computing the Spearman correlation coefficient presented in Table 2. Nevertheless, the correlation is not enough to guarantee the accuracy and so some statistical indicator(s) would also be required.

2) Line 225-227: “Therefore, it can be concluded that the precipitation forecast provided by MeteoGalicia offers results very close to the real rain events for the entire time series of precipitation data (2008-2018)”.

In fact, if hourly precipitation was used, the amount of null or very small values should be a very high percentage of the sample values. In order to predict floods, the test would require higher precipitation values (e.g. values above the 90th percentile) to be considered. To clarify/correct, see Brown et al. (2010).

Some statistical indicators were added to improve the analysis of the accuracy of precipitation forecast provided by MeteoGalicia (see lines 263-279 and Table 2). In addition, following the suggestions of the reviewer, we have also analysed the accuracy of the forecast taking into account only rainy events (precipitation above 75th percentile) (see lines 263-279 and Table 3).

Table 2. Values of the correlation (Spearman's r) and normalised standard deviation (σ_n) of the precipitation forecast using the measured data as reference at each rain gauge, considering the complete time series of precipitation. The averaged values for each precipitation forecast are also shown.

	Forecast window (h)					
	1-24		25-48		49-72	
Rain gauge	r	σ_n	r	σ_n	r	σ_n
rg_1	0.84	0.80	0.82	0.81	0.77	0.80
rg_2	0.84	1.09	0.82	1.07	0.79	1.07
rg_3	0.83	1.00	0.81	0.96	0.77	0.99
rg_4	0.81	0.97	0.79	0.96	0.75	0.98
rg_5	0.81	1.13	0.80	1.10	0.76	1.12
rg_6	0.84	1.16	0.83	1.07	0.79	1.07
rg_7	0.83	1.05	0.81	1.06	0.77	1.10
Mean value	0.83	1.03	0.81	1.00	0.77	1.02

Table 3. Values of the correlation (Spearman's r) and normalised standard deviation (σ_n) of the precipitation forecast using the measured data as reference at each rain gauge, considering only rainy events (above the 75th percentile). The averaged values for each precipitation forecast are also shown.

	Forecast window (h)					
	1-24		25-48		49-72	
Rain gauge	r	σ_n	r	σ_n	r	σ_n
rg_1	0.66	0.72	0.61	0.70	0.53	0.72
rg_2	0.71	1.00	0.63	0.98	0.56	0.99
rg_3	0.70	0.98	0.61	0.93	0.59	0.98
rg_4	0.73	0.93	0.65	0.90	0.60	0.93
rg_5	0.68	1.02	0.63	1.01	0.54	1.04
rg_6	0.69	1.14	0.65	0.98	0.56	1.00
rg_7	0.74	1.03	0.68	1.02	0.63	1.10
Mean value	0.70	0.97	0.64	0.93	0.57	0.97

Concerning calibration and validation of HEC-HMS

3.1) Line 230-231: "A set of 15 extreme flood events registered during the period 2008-2018 were used to calibrate and validate the rain-runoff model HEC-HMS."

It is suggested that the time duration of all the events as also the initial flow conditions are referred and/or resumed.

A new Table 4 was included in the manuscript.

Table 4. Main characteristics of the analysed flood events

Date of the flood event	Duration (days)	Initial flow (m^3s^{-1})	Initial depth (m)
28/12/09	4	52	1.3
17/11/10	5	116	1.7
17/01/13	10	164	1.9
11/03/13	5	179	2.0
05/11/13	7	234	2.3
14/01/13	10	165	1.9
28/01/14	15	202	2.1
01/03/14	4	134	1.8
30/01/15	3	184	2.0
01/03/15	3	134	1.8
10/02/16	7	216	2.1
26/02/16	3	137	1.8
05/03/16	4	175	2.0
10/03/18	6	154	1.9
30/03/18	4	201	2.1

3.2) Line 232-233: "…with HEC-HMS using the 1-day forecast of precipitation".

Q: Why calibrate for 1 day forecast and not use the real precipitation data? Wasn´t the accuracy of forecasted data checked before? If so, shouldn't also be done a calibration for 2 and 3 days forecast windows? To clarify/justify

We decided to use forecast data to make the calibration for several reasons. First, because the EWS is based in forecasted data: starting from precipitation forecast to obtain the flood maps of the area of interest. Therefore, we need to use the forecast precipitation data to feed up the model. This means that the EWS always works with forecast precipitation, and therefore we considered more adequately to calibrate the system using forecast data, which is the information used by the system. Second, because rain gauges only record the precipitation registered at a very concrete location and this could not be representative of the entire sub-basin under interest, taking into account the complex orography of the area. The MeteoGalicia forecast precipitation offers a high resolution of 4 km, being able to take into account the orographic characteristics of the areas under interest. For these reasons, forecast precipitation was used to the calibration purposes (see lines 284-286).

Respect to the use of a 1-day forecast window for calibration purposes, we took into account the accuracy of the predicted precipitation data shown in table 2, for the complete time series, and in table 3, for the data associated to extreme events. In general, the shorter the forecast windows the higher the accuracy. Therefore, we decided to carry out the calibration only with a 1-day forecast window, taking also into account that the values obtained (curve number and lag time) are also adequate for other forecast windows because the conditions of the terrain are quite similar (rainy events).

3.3) Lines 240- 241: "The mean values of CN and Lg of each sub-basin were used to validate the model in four flood events (01/2013, 01/2014, 02/2016 and 03/2018) by means of a Taylor diagram."

So, that meaning that those values are to be used for any simulations to be performed. So, the initial flow conditions and/or flood time duration are not relevant? That should be addressed and/or justified.

The mean values of CN and Lag time are used for all simulations but each of them is initialised with the flow and water depth conditions corresponding to the date of the event. See lines 322-323.

The presentation of the results for all four flood events (and not only for one) could contribute to that and is suggested.

Following the suggestions of the reviewer we included a new figure 3.

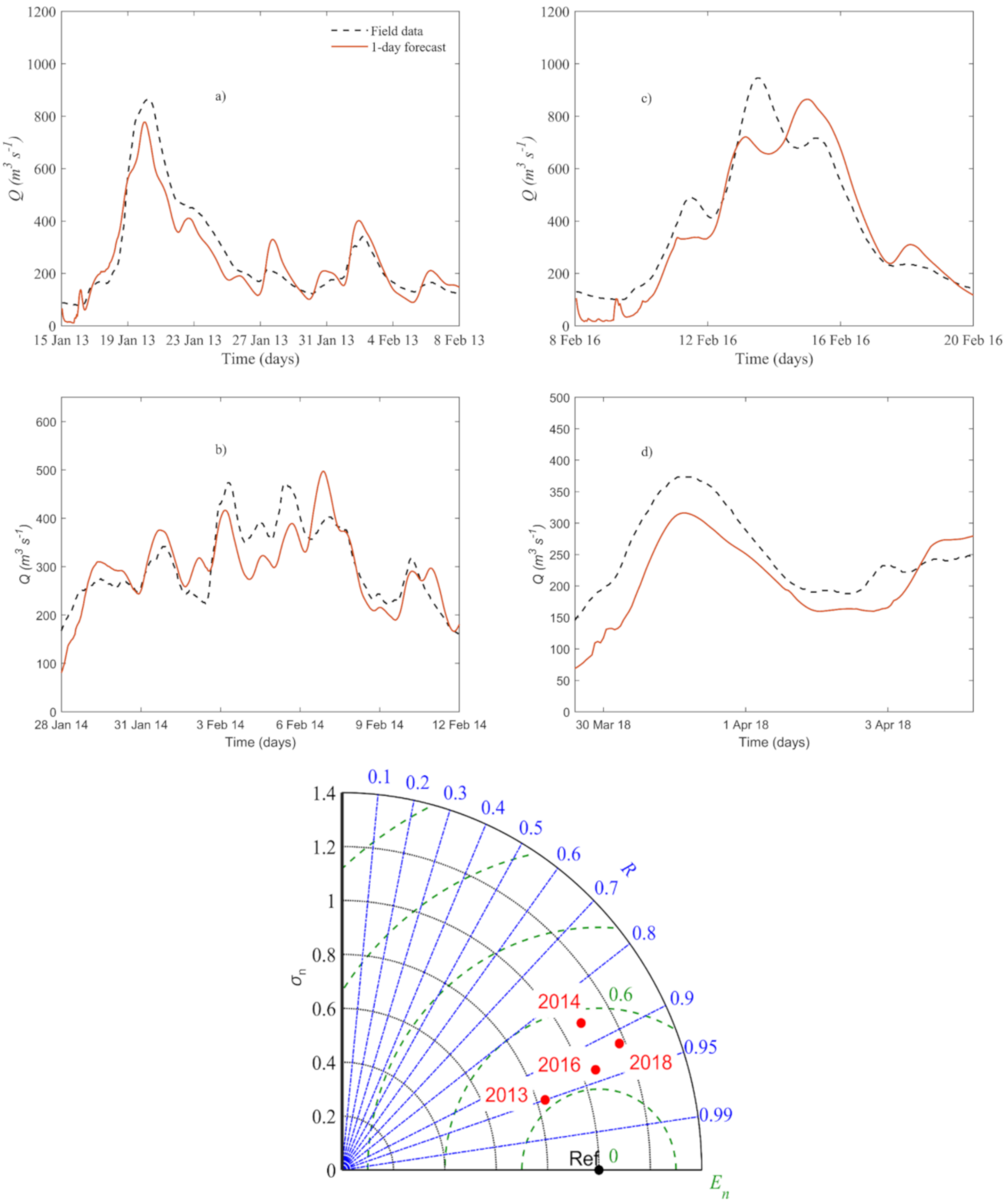

Figure 3. Time series of the registered streamflow (dashed line) and numerical streamflow (orange line) of the validation events: a) 01/2013, b) 01/2014, c) 02/2016 and d) 03/2018. Taylor diagram of the validation cases is also shown.

3.4) Lines 245-247: “These (statistical indicators) show that the mean values of CN and Lg obtained in the calibration step characterise the behaviour of the basin with a high accuracy”

That could possibly be better justified, namely by means of referenced bibliography.

Equation 1 defines the normalised standard deviation. This parameter shows the relation of the standard deviation of the numerical streamflow respect to standard deviation of the reference streamflow. A value equal to 1 means that both streamflows have the same standard deviation. Therefore, values that range from 0.8 to 1.2 means that the difference between the standard deviation of the numerical results are less than 20%.

Equation 2 defines the root mean squared difference. This parameter shows the relation of the standard deviation of the differences between the numerical and the reference streamflow respect the standard deviation of the reference streamflow. A similar parameter proposed by Moriasi et al. 2007 namely RMS (if the mean value of the numerical streamflow is similar to mean value of the reference streamflow the root mean squared difference is equivalent to RMS) is stated as very good for values that range from 0 to 0.5 and good for values that range from 0.5 to 0.6. These values proposed by Moriasi et al. 2007 are referred to monthly step but as the evaluation time step decreases, a less strict performance rating is warranted.

Equation (3) shows the definition of the correlation. It ranges from -1 to 1 and the best values of this parameter is equal to 1. It measures the degree of the collinearity of the simulated data respect to the reference one. Moriasi et al. 2007 proposed values of the correlation coefficient greater than 0.71 as acceptable values. Therefore, values that range from 0.85 to 0.9 show a good correlation between the numerical and the reference data. An explanation was introduced in lines 301-306.

The new reference was included: Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations (2007) Transactions of the ASABE, 50 (3), pp. 885-900. https://www2.scopus.com/inward/record.uri?eid=2-s2.0-34447500396&partnerID=40&md5=50b5724614f28257edef46d43db96018

Case Study

4.1) Authors are recommended to present more details about the hydraulic model setup.

- What type of boundary conditions were assumed at the inlet and outlet sections?

The inlet condition was defined by means of the input hydrograph (Critical/Subcritical) and the outlet condition was defined using a supercritical/critical outflow. (lines 323-325)

- How were the turbulence flow features modeled in this study?

Turbulence was not considered in this case since the uncertainty associated to this parameter makes very difficult to define it properly in most of the applications as pointed out in (SNCZI, 2011; Erpicum et al., 2010; Liu et al. 2013; Segura-Beltrán et al., 2016). This is mentioned in lines (325-327).

- Did authors perform any sensitivity analysis on the grid system or the Manning's coefficients used?

The Manning's coefficients mainly have influence on the timing of the peak flow. The values of the peak flow and the water depth are less dependent of this variable. Therefore, we use the default values proposed by Iber for each land use. The size of the mesh was equal to the raster resolution. Higher resolutions can be used but they present two main drawbacks. On the one hand the topography would not be real, but interpolated and, on the other hand, it is necessary a balance between accuracy and computational time. For this particular case, where only a part of the catchment area is considered, higher resolution can be used with low incidence on the computational time. However, conditions can be more restrictive when considering the whole catchment area.

- Were (all) the bridges considered in the study? If so, how? (Figure 8, bottom, shows bridge pier effects at one bridge).

The roman bridge located at Lugo was considered by means of the definition of its piers. The rest of the components of the bridge (deck, radius…) were not taken into account because of the difficulty to define the parameters associated to these parts and also because the historical records show that the bridge was never overtopped.

4.2) Line 264: What is the resolution of the topography data used in this study?

The resolution of the raster file used to define the topography of the area of study is equal to 5 m.

Technical corrections:

Line 20: model~~s~~

Done.

Line 29: estimated?

Done.

Line 49: enhanced?

Done.

Line 52: analysed?

Done.

Lines 55-58: The sentence requires revision.

The sentence was rewritten.

Line 79: to predict extreme flows?

Done.

Lines 87-85: "This model was *previously* calibrated for the area of study by means of series of historical flood events detected over the last decade."

"This model was calibrated for the area of study by means of series of historical flood events detected over the last decade." ?

Done.

Line 85: Iber+?

In this case we refer to the Iber model (not the implementation of Iber+) because the latter is integrated into project Iber (http://www.iberaula.es/).

Figure 1: Symbols Sb1, .. Sb3 were not defined in the text (sub-basins?).

This symbols are defined in the new version of the manuscript (Caption of Figure 1).

Sugg: Add the scale and north direction, particularly in the sub-figure including the city of Lugo: include Portugal (and the sea).

Figure 1 was edited.

Line 101: This sub-catchment area?

Done.

Line 106: in the entire sub-catchment?

Done.

Line 111: also affects?

Done.

Line 131: Magnitude velocity

Done.

Line 135: "The details of the components of the EWS …. are ~~shown~~ described in the following sections." presented?

Done.

Line 141: The link requires an update.

Done.

Line 154: Data from these rain gauges was used to assess the …?

Done.

Line 155: "The rain gauges selected for this study were shown in Figure 1.."

Sugg: The mentioned rain gauges are pictured … .

Done.

Line 158: "Minho River"

Miño river

Done.

Line 167: Iber+?

In this case we refer to the Iber model (not the implementation of Iber+) because the latter is integrated into project Iber (http://www.iberaula.es/).

Line 178: …normalised centred …

Done.

Line 181: meaning of subscript *n* is missing

Done.

Line 181-185: equations (1) to (3): i = 1?

Done.

Line 190: Iber+?

In this case we refer to the Iber model (not the implementation of Iber+) because the latter is integrated into project Iber (http://www.iberaula.es/).

Line 191: 2D (Two-Dimensional) (first time use of the abbreviation in the text)

Done.

Line 200: GPU: Graphics Processing Unit?

Done.

Line 205: (Definition for *bias* parameter?).

Both the bias and the RMSE are defined in lines 243-247.

Line 244: "root mean square difference" (RMSD?) (see line 252)

Done.

Q: RMSD or RMSE (line 205)

In this case we used the RMSE parameter.

Figure 8: Color-map is not clearly informative (e.g.: what are the values corresponding to the green color range?)

Figure 8 was adapted following reviewer's suggestions.