Response to Referee #1

We wish to thank the Reviewer for his/her thorough review of our manuscript and for the useful comments that helped us improve the quality of the paper. The specific issues raised by this Referee are addressed in detail below:

General Comments

The authors develop a numerical tool for predicting flood hazards in real-time in an Andean watershed. The tool is based on a data-driven surrogate model of a physically-based hydrological-hydraulic modeling cascade. The topic is interesting and well-suited for NHESS.

The performance of the surrogate model in predicting water depth is, however, lower than I would have expected, and calls into question the application of this method in an operational version.

We appreciate the positive comments from the Reviewer regarding the topic of our research. We would like to clarify this paper intends to show the first approach to use surrogate models as an alternative for water depth predictions in much shorter times than running high-resolution hydrologic and hydrodynamic models. While this idea can be used as the base of future operational models, we are aware that significant improvements and additional testing should be implemented in future research, as we have now discussed in the new version of the paper.

We apologize for not describing clearly the context of the study region in the original version of the manuscript, in which we should have explained the limitations of the surrogate model implementation in the Andes region. As we discussed in the companion paper recently published in NHESS (Contreras and Escauriaza, 2020), the main difficulty for predicting the flood propagation in the Andes is the very limited historical data and remarkably complex terrain. The Quebrada de Ramon watershed has only one water depth gauge with discontinuous 40 years of measurements and no rain gauges. Additionally, the magnitude of the floods makes measuring discharge even more difficult during storms.

In this case, we only have a record of 48 events with only one validation point in the entire region. With this limitation, we implemented a surrogate model that requires 4 inputs and predicts approximately 88% of the storms with a mean error of 2.22% at the validation point. Previous applications of surrogate models using similar methodologies for storm surge predictions have been built with 350 storms and 150 validation points, for 4 inputs. Their results have been validated with 20 storms with average errors of 4% and 3% for the significant wave height and water level respectively (see Taflanidis et al., 2013, for details).

While the order of magnitude of the average errors is similar, it is expected that a larger database to build our surrogate model would improve its accuracy, mainly because a wider variety of storms would be represented in the historical series.

As acknowledged by the authors, some storms are predicted with very significant errors. In my view, the paper falls short in explaining the reasons behind this and the possible ways to improve the model performance.

In section 4.2, the authors do not observe that the surrogate model does especially good or bad depending on the characteristics of the storms (understood here as the values of the 4 input parameters), but many other relevant aspects are not explored, such as (1) the choice of input parameters itself, (2) the type of surrogate model, or (3) the number of events in the database. The latter point is mentioned in the discussion (page 18, lines 9-11), but is not tested and shown in the results. Please find below some suggested readings that can provide insights into these questions:

Berkhahn, S., Fuchs, L., Neuweiler, I. (2019). An ensemble neural network model for real-time prediction of urban floods. Journal of Hydrology, 575, 743-754.

Bermudez, M., Cea, L., Puertas, J. (2019). A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. Journal of Flood Risk Management, 12(S1), e12522

Jhong, B.C., Wang, J.H., & Lin, G.F. (2017). An integrated two-stage support vector machine approach to forecast inundation maps during typhoons. Journal of Hydrology, 547, 236– 252.

Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. Water Resources Research, 48(7), W07401.

We sincerely thank the Referee for this comment, as we now realize that we did not explain thoroughly the relations and influence of the controlling factors of the extreme flood events in this region of the Andes mountains, and especially the process to select the input variables for the surrogate model, which is now explained in the new version of the paper.

Our previous research in these regions has shown that the variables that best explain daily discharges, particularly for low exceedance probabilities, are the cumulative precipitation over the previous 3 days and the minimum temperature on the day of the maximum discharge measured at a low elevation in the valley (Castro et al., 2019). This directly justifies the selection of the minimum temperature during the storm as part of the inputs, and underscores the importance of the liquid precipitation that occurs in the entire watershed during warm events (Contreras and Escauriaza, 2020).
The cumulative precipitation over the previous 3 days represents a combination of how much has rained and the duration of the event, which is also an indication of soil saturation in the watershed. Thus, we selected the mean of the precipitation during each event and the second moment of the distribution of precipitation as variables that represent the magnitude of the rainfall event and its distribution in time.
Finally, we consider the sediment concentration that might play a significant role on the flood propagation. Contreras and Escauriaza (2020) showed differences on the order of 25% for water depths calculated with clear water or 20% of sediment concentration. Additionally, differences of up to 0.5 m were observed in the urban area for hyperconcentrated flows. The main difficulty regarding the definition of this variable is the uncertainty of sediment concentration for each event, as localized landslides, previous recent storms, or interannual changes on the vegetation covering have produced different concentrations, which are also

difficult to measure in extreme flooding conditions. Therefore we selected this variable to evaluate potential scenarios, and analyze a flood with the same rainfall event, but under different sediment concentrations.
Magnitudes of sediment concentration have been reported for the largest flood registered in the watershed, which was generated during an abnormally warm storm, with periods of intense precipitation over partially saturated soils. The sediment concentrations during the events could not be directly measured, but it was estimated to be around ~40% (Sepulveda et al., 2006; Sepulveda and Rebolledo, 2008).

Regarding the types of surrogate models, we chose the methodology based on kriging due to the simplicity of its implementation in sites with limited information. To the best of our knowledge there are no previous studies that have developed surrogate models in ungauged mountain regions, therefore we selected a simple model, with the fewest number of parameters to calibrate, and keeping the physical meaning of all the inputs and parameters. We carried out a systematic study on the number of parameters, reducing and changing the combination of inputs, as shown in the following Table, and the model presented was the best in terms of reducing the error of the predictions.

	Mean Relative Error [%]		
Inputs	Regression Order 0	Regression Order 1	Regression Order 2
P2TC	1.93 - 3.55	2.26 - 2.70	12.61 - 20.81
P2T	2.53 - 3.74	2.36 - 3.05	5.75 - 9.73
P2C	2.39 - 3.52	2.51 - 3.26	6.64 - 7.96
PTC	2.61 - 3.53	1.99 - 2.49	3.68 - 4.19
2TC	2.49 - 3.60	1.81 - 2.07	5.50 - 6.49
P2	3.18 - 17.86	3.42 - 27.27	3.83 - 20.52
PT	2.39 - 4.41	2.38 - 3.77	2.24 - 3.87
PC	2.12 - 3.64	2.27 - 3.28	2.76 - 3.86
2T	2.57 - 8.43	2.96 - 12.21	2.59 - 187.76
2C	1.32 - 3.24	1.61 - 2.06	1.93 - 2.48
TC	1.78 - 3.20	2.21 - 2.90	2.55 - 3.40
P	4.14 - 1545.12	4.18 - 1561.64	4.02 - 1450.13
2	1.26 - 15.37	1.50 - 15.45	1.53 - 15.49
T	2.29 - 46445.19	2.55 - 46328.71	2.71 - 46322.26
C	1.83 - 33.37	2.20 - 37.27	2.19 - 34.89

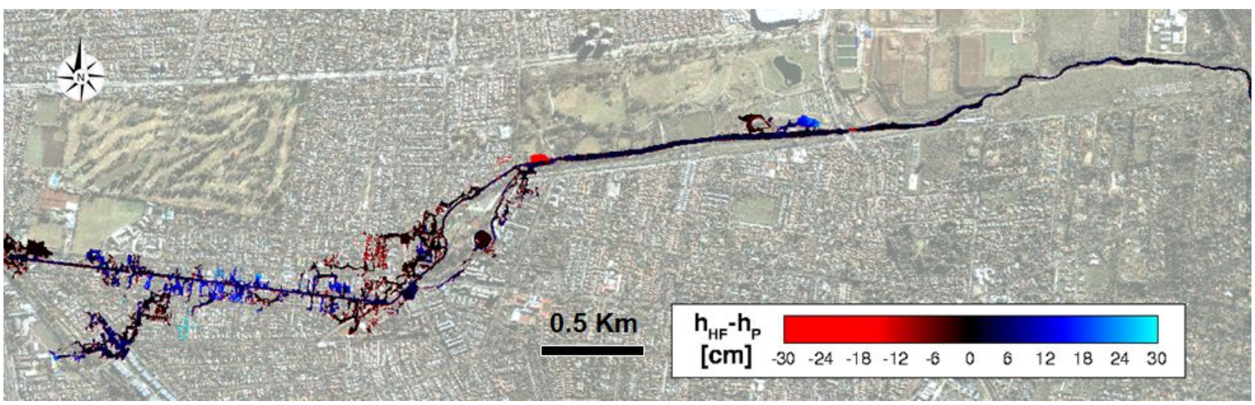
	Cases Predicted within one standard deviation [%]		
	Regression Order 0	Regression Order 1	Regression Order 2
P2TC	77.55 - 89.80	75.51 - 89.80	85.71 - 91.84
P2T	53.06 - 67.35	53.06 - 71.43	53.06 - 75.51
P2C	69.39 - 85.71	67.35 - 83.67	67.35 - 83.67
PTC	63.27 - 75.51	69.39 - 71.43	75.51 - 81.63
2TC	63.27 - 79.59	59.18 - 69.39	67.35 - 81.63
P2	40.82 - 79.59	38.78 - 77.55	46.94 - 81.63
PT	46.94 - 65.31	46.94 - 71.43	46.94 - 67.35
PC	63.27 - 75.51	69.39 - 79.59	69.39 - 83.67
2T	55.10 - 71.43	55.10 - 73.47	55.10 - 71.43
2C	63.27 - 79.59	61.22 - 71.43	63.27 - 73.47
TC	69.39 - 85.71	69.39 - 81.63	69.39 - 81.63
P	48.98 - 67.35	48.98 - 67.35	53.06 - 65.31
2	32.65 - 71.43	30.61 - 75.51	34.69 - 77.55
T	44.90 - 67.35	44.90 - 65.31	44.90 - 63.27
C	61.22 - 83.67	57.14 - 81.63	61.22 - 83.67

Even if depths are shallow, it is relevant to accurately predict flood extent for operational purposes and for extending the methodology to other sites (Page 17, lines 11-13 /Page 20, lines 5-7 of the manuscript). An additional step might be needed in the tool: a first binary classification model to predict when flooding occurs, and a second one to calculate its magnitude. I suggest evaluating the agreement between the surrogate and the physically-based depth maps obtained (ideally for all storms) by means of metrics such as the flood area index (a revision of commonly used metrics for this purpose can be found in Stephens et al. 2014. Problems with binary pattern measures for flood model evaluation. Journal of Hydrology 28 (18), 4928-4937). (Page 17, lines 14-18 / Page 21, line 19 of the manuscript).

We agree with the Reviewer, as shallow depths are also relevant for the analysis of the flooded area, and the metrics proposed by Stephens et al. (2014) have provided information on this binary classification for a smaller number of nodes. However, the analysis for all the storms in the unsteady flows of our cases, including high spatial resolution that varies in size as we move further away from the channel, makes this specific analysis a formidable task, which is outside the scope of the present investigation. We initially focused on the development and validation of the surrogate model in a mountain region, with limited available data and based on kriging interpolation.

It is important to emphasize that the nodes that are far from the main channel do not get inundated very often, and they participate in a considerably smaller number of events within the database, which adds uncertainty and calls for a careful analysis on this specific point in future research.

To provide an analysis of the errors on shallow flooded areas in the present investigation, we have now performed an averaged analysis, comparing the outcomes of the surrogate model and the deterministic simulations, as shown in the following figure. In this case, we can identify in blue and red, specific regions that might show problems with the prediction of shallow inundations.



Minor comments:

Page 4: Information on lines 9-10 seems to be repeated in lines 30-31. Page 8: How are buildings represented in the mesh of the flood inundation model of the urban area? Is it a building block method?

Response:

We have improved the description on how we represent the vegetation and buildings in the hydrodynamic model. We modified the lines 9-10 and 30-31 in the new version of the paper, so the information is not repeated. We specify on line 31 that we use a bare Earth or digital terrain model for the hydrological model. On page 8, line 17, we describe how we use a digital surface model (DSM) for the hydrodynamic model, and represent the geometry of buildings in the urban area with their elevations in the computational grid.

Page 9, Lines 16-17. "... .the zones where the sediment concentration produces significant changes on the velocity and flow depth of the flood" It's not clear to me how this is shown in Figure 5.

Response:

We apologize for not explaining in detail the effects of sediment concentrations, as we described them above. We have modified the manuscript to incorporate the adequate details on how the sediment concentration modifies the hydrodynamics of the floods in the urban region, which is based on the companion paper (Contreras and Escauriaza, (2020), https://doi.org/10.5194/nhess-20-221-2020, 2020.

Page 14, line 22: Please write "event" in full for clarity.

Response:

Corrected.

Page 16, line 9: It would be useful to indicate the CPU time to simulate in the physically-based model, for comparison purposes.

Response:

In the new version of the manuscript we have incorporated the information about the elapsed times for the simulations on page 16 lines 12 and 13. This information highlights the importance of the fast response of the surrogate model, which is almost instantaneous, contrasted with 2 to 3 days of calculation for a high fidelity simulation.

Page 19, lines 21-22: Is this filter applied in this work?

Response:

Thanks for clarifying this point. We did not apply this filter to these results. All the errors and standard deviations are influenced by the incorrectly predicted events. We have modified the manuscript to recommend filtering the predictions based on large standard deviations. We observed that this procedure might be one way to identify predictions with large errors.

Page 21, line 21: "we recommend using values of water depth in surrounding points as parts of the inputs for a specific point". As this possibility has not been tested in the paper, I suggest removing this recommendation.

Response:
We deleted the recommendation, following this comment from the Reviewer.

References:

Taflanidis, A., Kennedy, A., Westerink, J., Smith, J., Cheung, K., Hope, M., and Tanaka, S.: Rapid assessment of wave and surge risk during
landfalling hurricanes: Probabilistic approach, Journal of Waterway, Port, Coastal, and Ocean Engineering, 139, 171–182, 2013

Castro, L., Gironás, G., Escauriaza, C., Barría, P., and Oberli, C.: Meteorological Characterization of Large Daily Flows
in a High-Relief Ungauged Basin Using Principal Component Analysis, Journal of Hydrologic Engineering, 24, 05019 027,
https://doi.org/10.1061/(ASCE)HE.1943-5584.0001852, 2019

Contreras, M. T., and Escauriaza, C.: Modeling the effects of sediment concentration on the propagation of flash floods in an Andean watershed, Natural Hazards and Earth System Sciences, 20, 221–241, https://doi.org/10.5194/nhess-20-221-2020, 2020.

Sepúlveda, S., Rebolledo, S., and Vargas, G.: Recent catastrophic debris flows in Chile: Geological hazard, climatic relationships, and human response, Quaternary International, 158, 83–95, https://doi.org/10.1016/j.quaint.2006.05.031, 2006.

Sepúlveda, S. and Padilla, C.: Rain-induced debris and mudflow triggering factors assessment in the Santiago cordilleran foothills, Central Chile, Nat Hazards, 47, 201–215, https://doi.org/https://doi.org/10.1007/s11069-007-9210-6, 2008