Anticipating a risky future: LSTM models for spatiotemporal extrapolation of population data in areas prone to earthquakes and tsunamis in Lima, Peru

November 2023

Christian Geiß, Jana Maier, Emily So, Elisabeth Schoepfer, Sven Harig, Juan Camilo Gómez Zapata and Yue Zhu

We would like to thank the reviewer very much for the very valuable and helpful comments. We considered the remaining point and carefully revised the manuscript accordingly. Please find our detailed response to the comment below.

In the response below, the comment of the reviewer is reported in *italic*. Responses from the authors do not feature any special formatting. Quotations from the manuscript in its revised or in its original form are reported in blue.

Reviewer #1:

The manuscript represents a valuable contribution to the prediction of future hotspots of earthquake and tsunami exposure.

The title and abstract are relevant, succinctly encapsulating the paper's content. The introduction boasts a commendable array of references, which could be further enhanced in terms of quantity and relevance. The Materials and Methods section is quite comprehensive, while the Results and Discussion are well-structured, complemented by various supportive figures. The conclusions are succinct but clear.

We thank the reviewer for his/her generally positive judgement and genuinely appreciate the suggestions for improvement.

Additionally, I would like to suggest the following improvements:

The abstract lacks essential data pertinent to the conducted analysis, specifically information regarding the precision of the methods employed and the reliability of the results.

Thank you for sharing this observation. We provide more quantitative number regarding the model accuracies as follows now:

The results clearly underline the value of the LSTM-based models for forecasting gridded population data: The most accurate prediction obtained with an LSTM equipped with a bidirectional learning scheme features a root-mean-squared error of 3.63 people per 100 *100 meters grid cell, while maintaining an excellent model fit ($R^2 = 0.995$). We deploy this model for anticipation of population along a three-year interval until the year 2035.

Is Figure 2 too large? Nevertheless, the text at the top can be resized to ensure compatibility with the main text.

Thank you for this observation. We resized the text accordingly and provide a revised figure in the new manuscript.

For Figure 5a's legend, it would be advisable to ensure that the text does not overlap with the image territory.

Thank you for sharing this observation and suggestion. We changed the legend in a way that the legend does not overlap anymore with relevant image content. According to our perspective, that was the case for district boundaries that were partially affected by the placement of the legend – this is not the case anymore.

On page 8, row 217: "Is the equation in (3) the same as before?"

We modified the sentence and hope that it is less ambiguous now:

Equation (3) describes this relation:

$$t_{n+1} = t_n + i \tag{3}$$

with t_{n+1} being the forecasted year, t_n the year of the last input, and i the interval size.

Regarding Figure 6, the caption appears somewhat lengthy. Have you considered incorporating some of the description into the main text?

Thank you very much for this observation. We shortened the caption but still aimed to keep it somewhat self-contained, since it could be difficult to understand the rather complex content from the text only.

Figure 8's clarity could be improved as the details seem somewhat indistinct; the choice of subtle colors may have influenced the outcome.

Thank you for sharing this point. We agree that the LSTM-based and ConvLSTM-based outcomes feature a pattern that is not prominently distinct in this plot. However, the overall goal is to particularity highlight differences w.r.t. the MLR and RF models. We hopefully make this clear by relating to this aspect in the corresponding text:

Figure 8 provides prediction differences to the actual numbers of 2020 from a spatial perspective. Grid cells with overestimated population numbers are colored in green, whereas grid cells with underestimated population numbers are colored in red. Thereby, it can be traced that the LSTM-based and ConvLSTM-based predictions overestimate population numbers for the majority of grid cells, while both the MLR-based and RF-based predictions underestimate population numbers for the majority of grid cells (also revealed by the regression line in Fig. 7). However, both the LSTM-based models and ConvLSTM-based models follow consistently the overall trend of an area: in tendency, they exaggerate population numbers in areas of an increasing population and underestimate population numbers in areas of a decreasing population (see also Fig. 2 for a visualization of areas of increasing and decreasing population numbers in Lima and Callao). MLR-based and RF-based predictions do not reflect this overall trend, whereby over- and underestimations are more dispersedly distributed across the study area.

Concerning Figure 11, I'm unclear about the significance of the "pop increase 2020 - 2035" bar. Could you provide clarification on its meaning?

Thank you very much for identifying this unclear point. We substantially extended the caption to hopefully communicate the content of the figure better:

Maps of the predicted population affected by earthquake (upper figure) and tsunami (lower figure) for the year 2035 with corresponding hazard intensities. The solid grey bars indicate the population of the year 2020. The additional colored bar (on top) or textured bar indicate the estimated increase or decrease of the population until the year 2035, respectively. The corresponding color coding indicates the hazard intensity.

Reviewer #2:

This is a very good paper that examines a critical topic for disaster risk reduction (DRR): the estimation of the actual populations in areas exposed to hazards and, moreover, the prediction of how these populations might change within a given time threshold. Commonly, this information is either unavailable or unupdated (especially in countries belonging to the Global South), so this paper makes a significant contribution along these lines.

The manuscript is clearly arranged and written, including an introduction section that delivers a proper literature review and describes the study's rationale and objectives and a materials and methods section that explains the sources and methodologies applied in this research. The results are presented and discussed with proper clarity and depth (and the included figures give sufficient support for this). It follows my suggestions for further improving the manuscript:

We thank the reviewer for his/her generally positive judgement and genuinely appreciate the subsequent suggestions for improvement.

First, I think that the authors should deliver a more profound argument about the choice or identification of what they call driving factors ("different feature sets used for the prediction", "geospatial covariates"). They point out that "the collected driving factors represent frequently adopted variables in studies of predicting geospatial change trajectories (Gómez et al. 2020; Liu et al. 2017; Pijanowski et al. 2002)" (page 4, line 132), but given the critical importance of the specific choice of factors, I think that more explanation or rationale should be provided.

Thank you very much for discussing this point with us. We added additional content on the issue and also justified the selection of variables better, i.e., justify that we cover the main variable categories:

The compilation of a powerful and exhaustive set of geospatial covariates is frequently a challenge in the context of implementing a robust study design. For instance, Zhu (2023) lists more than 50 predictor variables which were employed in existing studies of land use and land cover prediction. Here, the collected driving factors represent frequently adopted variables in studies of predicting geospatial change trajectories (Gómez et al. 2020; Liu et al. 2017; Pijanowski et al. 2002). In detail, we internalize the main variable categories (Zhu, 2023), i.e., land use-related variables (Fig. 3a-b,f), environmental variables (Fig. 3c,d), infrastructural variables (Fig. 3e), as well as socio-economic variables (Fig. 2).

Second, I believe the implications of the findings in section 4.3 require further discussion. The paper delivers an argument in terms of identifying the exposure of populations to tsunami and earthquake hazards, but after that, the manuscript only points out that "Anticipating those patterns can help urban planners and policy makers to proactively develop effective strategies for risk mitigation" (page 13, line 309). For instance, considering a certain degree of uncertainty, how could these patterns be integrated into planning? How might they lead to a change in DRR policies, considering the possibility of building and assessing different development scenarios? I know these objectives are beyond the scope of the paper, but enhancing this section would improve the paper's multidisciplinary interest. Reviewing additional literature could help to do this.

Thank you very much for identifying those points and the suggestions. Consequently, we extended the end of section 4.3 and hopefully address the raised points:

Anticipating those patterns can help urban planners and policy makers to proactively develop effective strategies for risk mitigation. For instance, the created information about exposed population can be part of modern decentralized information systems for (multi)-risk assessment (Schöpfer et al., 2023). Here, one core element is to enable end users to explore various scenarios ("stories") of multiple hazards, cascading effects and their impacts by quantifying different *what-if* scenarios. Utilizing such a narrative-driven methodology empowers individuals to replicate diverse situations within a predetermined, multi-risk context, enabling them to assess and contrast outcomes. This multi-scenario approach proves invaluable for crafting strategies that fortify or enhance resilience, evaluating the effectiveness of proposed or already executed measures (e.g., benchmark scenarios) in the face of various hazard scenarios (acting as a 'stress test'), or in response to evolving conditions. Thereby, the importance of implementing mechanisms to visualize epistemic and aleatory uncertainties of the risk assessment procedure in graphical form is stressed to allow appropriate communication with end users.

Some minor issues:

The concept of "natural disasters" (page 1, line 29) is not typically used in current literature about disaster risk reduction. I suggest using "socio-natural disasters" or similar.

Thank you very much for this suggestion. We altered the paper accordingly and use the term "socio-natural disasters" consistently (used two times throughout the manuscript).

In page 15, line 320, I think the word "population" is missing before "distribution".

Thank you, adjusted accordingly.

Throughout the manuscript, the authors use indistinctively passive and active verb forms ("were adopted", "we carry out", "our setup", "was carried out"). Please check for consistency.

Thank you for this observation. We substituted passive wording with active sentences and aimed to consistently use past tense if applicable (we changed around 30 sentences).