

Issue #13 (open): [REVIEW] Recommendations for visual predictive checks in Bayesian workflow

[@Guoziyang27](#) on
Jul 07, 2025 05:03: [opened]

[@Guoziyang27](#) on
Jul 07, 2025 05:03:

Conflicts of interest

- ☒ I declare that I have no known conflicts of interest with the authors.

Reviewed version

844162b

Review

This paper contributes to the evaluation of visual model checks. It does a good job of motivating and analyzing the concerns that arise around canonical visual model checks. It is enjoyable to read, and the rigorous evaluation based on PIT ECDF is a reasonable way to evaluate a potentially wide range of visual model checks.

The paper is generally well-written, and the visualizations are easy for me to understand, despite some missing details. The mathematical notations throughout the paper could also be improved.

Lack of discussion in section 4

I like how you compare two different models in Section 4, but I think the paper would benefit from more discussion about how these visualizations are used differently from each other. For example, the authors showed two different types of calibration plots, but they didn't actually discuss where they could be used differently. A table in a similar format to Table 1 could largely improve clarity.

Taxonomy of visual model checks

It would be useful to discuss the design space of visual model checks in a more systematic way. For example, visual representations and visual layouts should be different dimensions for visual predictive checks [2]. The authors tried to discuss overlaid plots as an additional dimension, but there are more to consider, such as juxtaposition and explicit-encoding [1]. The authors introduced residual plots in 4.4 for binary data, but residual plots should theoretically be applicable to all data types.

Usage of PIT ECDF in visual predictive checks

I agree that goodness-of-fit is important for visual predictive checks, but the discussion from 2.1 to 2.6 is generally applicable to any visualizations showing distributions. What is the specific point regarding visual predictive checks? The PIT ECDF quantifies the fitness of the visualized distribution to the actual distribution, but this is not the exact goal of visual predictive checks, where the latter is to check the fitness of the predictive distribution to the actual distribution. I'm not suggesting that goodness-of-fit is the wrong way to evaluate visual predictive checks, but it needs more clarification for readers to follow closely.

Notations

I am confused by the notations used in the paper. These points especially concern me: - What is the difference between $\hat{f}$ and f ? Is the former the visualized distribution or the predictive distribution? - $u(\cdot)$ and $PIT(\cdot)$ seem to be used interchangeably. It would be better to have a consistent definition. - What is F_{PIT} ? How was it calculated?

I would suggest adding a separate section for notations before getting into the main content. For example, it should be clear what notation is used for model predictions, what notation is used for visualized distributions, and what notation is used for ground-truth observations (or distributions).

Minor issues

- The headers of the second and third columns of Table 1 are not aligned with the content.
- Axis labels should not be omitted throughout the paper. The authors should also add legends for model predictions and observed data.

[1] Gleicher, Michael, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. "Visual comparison for information visualization." *Information Visualization* 10, no. 4 (2011): 289-309. [2] Guo, Ziyang, Alex Kale, Matthew Kay, and Jessica Hullman. "VMC: A Grammar for Visualizing Statistical Model Checks." *IEEE Transactions on Visualization and Computer Graphics* (2024).

Openness/Transparency

The paper includes all necessary code to replicate the plots, though I did not go through all the codes.

Submission categories

- ☐ Registered Report
- ☐ Replication Study
- ☒ Empirical Research - Quantitative
- ☐ Empirical Research - Qualitative
- ☐ Systems or design research
- ☐ Commentary
- ☐ Systematic Literature Review

Suggested outcome

Minor revisions: this paper requires some smaller changes, after which I am confident I would be able to endorse it.

Requested changes

I would like to see the analysis in section 4 to more comprehensive and the exploration of the visual checks to be more systematic.

ORCID

<https://orcid.org/0009-0004-4200-6774>

@Fumeng-Yang on
Jul 08, 2025 00:53:

Thank you for providing your review!

@TeemuSailynoja on
Aug 07, 2025 10:17:

@Guoziyang27, thank you for the time and effort you have put into reading the paper and giving us feedback. We are confident the paper will be improved by us addressing them.

