The paper presents BN-FLEMOΔ, a probabilistic flood loss model tailored for residential buildings in delta cities like Ho Chi Minh City, using empirical survey data and machine learning. The model employs automatic feature selection and Discrete Bayesian Networks to capture probabilistic dependencies and provide a probabilistic distribution of losses.

The questions addressed by the paper are relevant and the paper presents a novel interesting dataset based on a survey study in Ho Chi Minh City.

Regarding the pipeline, the authors should clarify some major points, in particular, related to the feature selection step:

1) The authors use a 10-fold cross-validation for evaluating feature selection, the selection of the Bayesian Network (BN), and performance evaluation. However, it seems that they perform an initial 10-fold cross-validation (10FCV) for feature selection, and a separate one for model evaluation. This is evident from the fact that they select seven features based on the mean results of the first 10FCV, and also from the fact that they use Python and R—one for the first phase and one for the second (which would be acceptable if one could exactly map the splits from one language to the other). If this is the case, such an approach can lead to information leakage from the whole dataset (when selecting the features) into the model evaluation phase, since the feature selection step has effectively "seen" data from all folds. This can often yield overoptimistic estimates of the model's true performance on unseen data. The pipeline should perform feature selection within the same fold in which the model is trained, or find other ways to prevent information leakage.
2) Does the dataset have any missing values other than rloss? If so, the authors should specify how they handle these missing data points when computing the multivariate linear regression (MLR). MLR, by itself, cannot handle missing values. If the excluded features have missing values, their inclusion could negatively impact performance if they are set to a specific value (for example, 0). In general, it is important to clarify how these particular data points are handled.
3) The strategy chosen for feature selection is unusual when combined with Bayesian networks. The authors evaluate the importance of each individual feature using F-statistics to measure the relationship between the variable and the target variable. This univariate approach can overlook the signal from variables that, when combined with another feature, a Bayesian network (BN) might capture. Moreover, the hyperparameter k is chosen using a linear model. These choices are not strictly speaking incorrect, but they may reduce the predictive capacity of the entire pipeline. Additionally, in BNs, it is possible to reduce the complexity of the model during the structural learning phase in ways that preserve the multivariate nature of the problem.
   a) The dataset has only 16 variables; is feature selection necessary? BN structural learning can usually handle this number of variables with an appropriate choice of algorithms and parameters, without the need for prior selection.
   b) Feature selection is part of the pipeline proposed in the paper. The authors should compare its performance not only with a model trained on the subset of selected variables but also with models trained on the entire dataset. For example, they should compare it with a random forest trained on all the features, since RF can easily handle 16 variables.
   c) To demonstrate the significance of feature selection, the authors should compare the pipeline's performance with BN models trained on all the features, using appropriate algorithms that can reduce complexity or identify Markov blankets from the data (see, e.g., Vogel 2018 and the BN literature).
   d) To better illustrate the impact of feature selection, it would be insightful to include a plot similar to Figure 2a, but showing the performance of BNs

trained with varying numbers of features. This could also be done for only the best model identified by the authors. If computational limitations exist, it would be reasonable to limit the study to a few additional features beyond seven.

e) It would be insightful to see the error bars with the standard deviation over the 10-fold in Figure 2a.

Regarding the training:

4) In line 226 they should specify what exact Bayesian algorithm they are using.

Regarding the dataset:

5) To better present the dataset, it would be useful to include the number of missing values for each feature, if any, in Table 1 or another section of the paper.
6) It would be useful to see the histogram of rloss before the discretization.

Besides these issues that need to be addressed, the paper is well-written, and the overall presentation is sufficiently well-structured. The lookup table provided in the supplemental material is useful and adequately described.