We thank the reviewers for the in-depth reviews. In this paper, we propose a new paradigm for model selection for production systems (MSPS), in which model selection is achieved by sequentially deploying a list of candidate models in order to discover the best model with the minimum number of online experiments. We show that a Gaussian process (GP)-based surrogate model can efficiently guide exploration-exploitation trade-off and outperforms a set of baseline methods. We will first answer the comments shared by multiple reviewers and then answer the individual comments.

[R2 & R4] Non-GP surrogate. As suggested by the reviewers, the surrogate models that are beyond Gaussian process (GP) have been studied in many related areas such as uncertainty quantification, Bayesian optimization, which often have non-stationality and better scalability. However, as we propose a new MSPS paradigm, the main focus of our work is to demonstrate: (i) sequential online deployment can lead to more efficient model selection than the state-of-the-art approaches (A/B tests and OPE). (ii) model uncertainty via acquisition functions can efficiently guide exploration-exploitation in our setting. Using a GP as the surrogate model allows us (i) to have a closed-form distribution of accumulative metric; (ii) to easily get high-quality uncertainty via being Bayesian non-parametric. Such a clean solution is important for developing and understanding a new method. Exploring other surrogate models that overcome the limitations of GP is beyond the scope of this paper and is an excellent direction for future research.

[R2 & R3] Experiments on production system. As reviewers 2 & 3 suggest, experiments on production systems can indeed reveal the true performance of our methods. However, we argue that such experiments are often noisy, conducted in an uncontrolled environment making a fair comparison against baselines practically impossible. On the contrary, the simulation studied in our paper provides valuable insights that are reproducible and helps gain a deep understanding of all methods presented. Such simulation studies are common in the recommender system and information retrieval research (Rohde et al. 2018; Chaney et al. 2018; Joachims et al. 2018; Jagerman et al. 2019). Furthermore, the results on production system will not be reproducible as public release of such datasets would not be possible due to IP / privacy concerns.

[R1] Missing technical details. We will add additional details about sparse GP, VI and binary observation in suppl.

[R1] How is the probability matrix $\mathbf{P}$ computed? A matrix $\mathbf{P}$ is computed for each model. Each column is associated with an input $\mathbf{x}_t$ and its values represents the probabilities of taking the corresponding actions (there are K rows, one for each action, summed to one). We collect the inputs in all the online experiments so far and use them to compute $\mathbf{P}$.

[R2] The proposed model is not time-aware... As our method focus on allowing model selection from a large candidate pool with a limited A/B test bandwidth, we target at the scenarios where candidate models are deployed to A/B tests for a few weeks and then select the best model, which is the common scenario in industry. The state-of-the-art approaches such as A/B tests and OPE share the same challenges as our method regarding time sensitive systems. Model selection for a time sensitive system is an interesting and open research question for future work.

[R2] Open sourcing. We intend to release the source code if our submission is accepted.

[R3] What are the noises in the noise distribution? How do they affect the feedback? ... what kind of uncertainty could be extracted from the noises. For example, for a recommender system, the noise can be due to human's random behavior (randomly deciding to click a link) or the information that is not available to the system such as the mood of the moment or influence of another person. As these data variance cannot be explained by the input to the system, these data variance will be learned into the noise distribution. (For a Gaussian distribution, this will lead to a higher variance). This is often referred to as aleatoric uncertainty. This noise variance will be excluded when estimating the distribution of the accumulative metric (see Sec. 3.2), because these variance cannot be reduced by collecting more data.

[R3] In line 113, ... what motivates a low-dimensional representation and/or why ... is good enough? Is the low dimensional representation still deterministic? Does it involve uncertainty? Learning latent representations is a common technique for latent variable models such as variational auto-encoder. Being low dimension (5D in our case) allows the model to better determine the representations with a limited amount of user-item interactions. Thanks to the high modeling capability of GP (a universal approximator), even complex data like natural images can be encoded into low dimensional representations. In our model, the latent representations are in the form of variational posterior (not deterministic). The uncertainty in these variational posteriors contributes to the uncertainty of the accumulative metric, which is used to guide exploration-exploitation trade-off.

[R4] Experiments with real-valued immediate feedback. We choose to conduct experiments on binary immediate feedback, because it is widely used. Binary feedback is often easy to interpret and less biased (e.g., on a streaming platform using the time that a user spends can lead to a bias towards the content with longer duration.) We do not expect our method perform differently on real valued immediate feedback. We quickly run a small experiment (20 repeats) comparing our method with IS-g on a synthetic task with 10,000 candidate models (see the figure).