We thank all reviewers for their valuable comments. We will address all the minor points (typos, notation issues, and remarks) as underlined and requested by the referees. Hereafter, we address comments shared by several reviewers.

Generalization beyond kNN method. First of all, although we apply our methodology to the kNN algorithm, we provide general consistency results in Theorem 3.2 for the proposed plug-in procedure that apply to *any* off-the-shelf estimators on the regression and the variance functions provided that these estimates are consistent. We illustrate this capacity in particular in the numerical experiments on popular machine learning algorithms such as random forests or svm for instance.

Additional technical aspect. As just mentioned, our methodology can be used to any estimators of the regression and the variance functions. However, our general Theorem 3.2 asks for the continuity of the cumulative distribution function of $\hat{\sigma}(X)$ (see Assumption 3.1). This assumption is violated when kNN algorithm is used and this is why, in addition to its popularity, we specify our methodology in the context of kNN. In particular Section 4.2 presents a randomized predictor based on kNN algorithm. This randomization technique used to circumvent Assumption 3.1 is rather simple but we believe that it is instructive from a methodological aspect. Even if the considered estimators of the variance function satisfies Assumption 3.1 the randomization technique can also be applied. In some sense, the construction given in Section 4.2 makes our results more general. Besides, we mention that Assumption 3.1 is one of the limitation that appears in Denis and Hebiri (2019) in the classification with reject option framework, and then Section 4.2 provides a way to handle this issue.

Extension to the regression setting with application to kNN. We totally agree with **Reviewer 1**: most of our results have direct analogues for classification (we will add more precise pointers to the final version upon acceptance). Considering the regression setting is an interesting application of the reject option problem which is new in the literature up to our knowledge and actually helps to understand a bit more the different characteristics of the model that are relevant for the rejection rule such as the variance function (which is intuitive in the end but good to be noticed). Moreover Section 4 is an original application to kNN that has no analogue and poses some additional technical tools since *i)* Assumption 3.1 does not hold with kNN and then randomization is required; *ii)* it asks for a finite sample bound on the L_∞-norm estimation error of kNN which we derive based on previous asymptotic works on kNN.

L_∞-norm bound. The obtained rates of convergence relies on Proposition C.1 (given in the supplementary material), it requires rate of convergence for the estimation of the variance function with respect to the L_∞-norm. In particular, we establish that

$$\mathbb{E}\left[\left(\sup_{x\in\mathcal{C}}\left|\hat{\sigma}^2(x)-\sigma^2(x)\right|\right)\right] \leq C\log(n)n^{-1/(d+2)} \ .$$

This result relies on the rate of convergence of $\hat{f}$ (the estimator of the regression function) *w.r.t.* L_∞ norm and can be easily extended for instance to the class of partitioning estimators or kernel estimators, under similar assumptions as in Section 4. That is to say the rate of convergence given in Theorem 4.4 applies to estimators which can be written as $\hat{f}(x)=\sum_{i=1}^{n} W_{n,i}(x)Y_i$ where $W_{n,i}$ are weight functions. We also want to point out that we do not find a generic reference for the rate of convergence for kNN estimates in L_∞-norm. But, the proof of this result shares ideas similar to the proof of Theorem 12.1 in Biau and Devroye (2015) which establishes only the consistency of the kNN estimates *w.r.t.* L_∞-norm.

About the sample size of the semi-supervised procedure. First, we want to notice that the result provided in Theorem 4.4 shows, from a theoretical perspective, the dependency with respect to the sample size of labeled and unlabeled sample. From the practical point of view, moderate values of N ($N \propto 100$) already describe a regime of convergence for the estimation of $F_{\hat{\sigma}}$. We are convinced that numerical performance of the proposed semi-supervised method is mainly dictated by the quality of the estimation of the regression function and the variance function which is one of the main issues of the regression with reject option problem.

Bibliographic remark. Finally, we will include the references to the classification with reject option shared by the referees. In particular, we will add in the text the reference to the paper of Herbei and Wegkamp (2006) to underline the relation with the classification with reject option framework. We also want to highlight that, since we focus on the regression setting, the variance function σ^2 is not necessarily bounded and thus we do not specify any upper bound for λ in the paper.