Thanks for the thoughtful reviews, we address the reviewers comments and questions below:

Reviewer 1:

Regarding the Gaussianity/universality assumption, we will discuss it in terms of earlier works, in particular [Elkhalil et al., 2020] (referenced in Review 3), and we will clarify how it is supported by the numerical experiments.

Thank you for pointing the works of [Rifkin & Lippert, 2007] on the leave-one-out error! We will cite these works and discuss the similarities and differences with the resulting estimators, in particular, the role that the trace plays in the KARE and the computational implications and benefits.

We will add a discussion of [Gerfo et al., 2008] in relation to the decomposition of the risk along the principal components of the data.

Reviewer 2:

We agree that adding more details on the mathematical tools (in particular around the Stieltjes transform) would benefit the audience; we found it hard to do this with the page limit, but we should be able to do this with the extra allowed page if the paper is accepted.

About Def 1, this is our working definition; we will add a remark explaining how it arises from the kernel optimization problem (the theorem you refer to). There is random noise on the outputs and the inputs are also random, which makes the kernel Gram matrix a random matrix (though conceptually, we don't view the input randomness as a "noise".

Expliciting the eigendecomposition of a concrete kernel is indeed difficult; the SCT is hence more a theoretical tool allowing one to reveal the KARE, which is not dependent on this eigendecomposition and easily computable. The eigendecomposition of integral operator T_K of the RBF kernels for Gaussian inputs is explicited on page 4 of the Appendix; we will add a reference to this in the main (in addition to Figure 2).

We will add some comments to clarify the exposition of the Part 4, and clarify the connection with the previous parts (in particular, how we arrive to the KARE).

We will take advantage of the extra page allowed if the paper is accepted to discuss the numerical experiments in the main.

Reviewer 3:

We will improve the discussion on the Gaussianity assumption, see response to Reviewer 1 above. We will emphasize that although the input distribution influences the observation distribution, the moments of the observations are what matters to study the expected risk; in other words the input dependence can be understood through the lens of the observation distribution moments. We will discuss this in relation with existing RMT results.

Regarding non-asymptotic vs data distribution agnostic results, both are crucial in our paper: (1) For the SCT, the non-asymptotic results are indeed the main challenge (2) For the KARE, a central point of interest is the fact that it is agnostic to the data distribution.

The effect of the dimension d is only indirect in our result indeed. You are right that convergence rates of the theorems depends on the SCT ϑ. In the RBF with Gaussian data case, the dependence of the SCT on the dimension d is made explicit in the Appendix on page 4, where we can add a note in relation to your question.

Regarding Lemma 6 of the Appendix, you are perfectly right, we will add the references you mentioned. We will add a few lines to clarify the difference between the kernel function $K(x, x')$, the kernel operator K and the Gram matrix G.

The discussion on the Stieltjes should indeed be augmented, and we will do this thanks to the extra space available.

Regarding the organization, we will detail the contributions part a little more, in particular by making precise references to the key theorems.

Reviewer 4:

We agree with all your comments we will update the paper accordingly.

Regarding the question about shift-invariant kernels: we did experiments with non-shift-invariant kernels and we will add them to the appendix. You are right that our theoretical work does not assume shift-invariance.