Reviewer 1: Thank you for your valuable comments. Our replies follow. **(1)** Our experiments show comparable results with other competing methods, while offering an advantage of choice of operator-valued kernels (which might not be non-negative). The results for each dataset are provided for a single test dataset. More replications might be needed to infer statistical significance of our results, which we can add. **(2)** Our stabilization problem (4) in Section 3, inspired from (Ong et al., 2004) helps in deriving the result in Representer Theorem 3.1. On the other hand, when the stabilizer $\widetilde{F}_\lambda$ from Eq. (4) belongs to the ball $\mathcal{B}_\mathcal{K}$ of fixed radius r (defined in Section 5 with $r = 1$), it enjoys the generalization bounds in Eq. (8). At least to us it is not very clear how the stabilizer behaves when it does not belong to $\mathcal{B}_\mathcal{K}$. One might suppose that the formulation similar to (Oglic and Gärtner, 2018) can be used here. However, adapting the minimization problem formulation in (Oglic and Gärtner, 2018) would lead to integral variance constraints in our case. Further, using a Gateaux derivative approach for the constrained or unconstrained minimization problem similar to that in (Oglic and Gärtner, 2018), leads to difficulties in obtaining the Representer Theorem 3.1 in our paper. As a consequence of these facts, we can only resort to an empirical cross-validation approach which we have used in our experiments to ensure that the stabilizer of problem (4) is not far away from $\mathcal{B}_\mathcal{K}$. **(3)** In contrast to scalar-valued indefinite kernels which arise naturally in many scenarios, we are not aware of natural occurrences of operator valued kernels (both positive and indefinite) in existing literature. Hence we motivate the use of generalized operator valued kernels from a function estimation and learning methodology viewpoint, which allows us to relax the requirement of positive definite kernels in learning function-valued functions. However, we will strive to improve the motivation.

Reviewer 2: Thank you for your encouraging comments. We will refine the references and include other related works.

Reviewer 3: Thank you for your inquisitive comments. Our replies follow. **(1)** Lemma 2.2 and 2.3 presented in our paper are for function-valued RKHS and $\mathcal{L}(\mathcal{Y})$-valued kernels, whereas the similar lemmas in (Alpay, 1991) are for $\mathbb{C}^{n\times n}$-valued kernels. Though our results are extensions of similar results in (Alpay, 1991), we point to the important differences here. In the proof of Lemma 2.2, we require the results in (Carmeli et al., 2006) and (Carmeli et al., 2010) to prove that $\mathcal{H}$ is a function-valued RKHS, which are not required in Alpay's proof. In deriving Corollary 2.3.1 using Lemma 2.2 and Lemma 2.3 we needed to establish arguments for operator valued kernels which were not obvious based on the arguments in (Alpay, 1991). **(2)** The derivation of representer theorem in our case requires using the definition of generalized operator-valued kernel (in Section 2) to obtain Equations (24), (25) and (26) in Appendix E which yield the required representer theorem in our setting. The derivation in our case uses Gateaux derivative with variational function approach to obtain necessary condition for stationary points for the stabilization problem (4), whereas the result in (Ong et al., 2004) uses subdifferential with respect to a vector $[f(x_1), \ldots, f(x_m)]^\top$ to obtain the representer theorem. **(3)** However, the bound on Rademacher average in Section 5 is a natural extension of the result in (Maurer, 2016). **(4)** We have cited published version of ref. [A1] in Section 5; on careful reading of ref. [A1], we found that trace class condition (Assumption 5.1) is used in [A1] as well. **(5)** The results for each dataset are provided for a single test dataset. More replications might be needed to infer statistical significance of our results, which we can add. **(6)** Thank you for the additional references. In contrast to scalar-valued indefinite kernels which arise naturally in many scenarios, we are not aware of natural occurrences of operator valued kernels (positive and indefinite) in existing literature. Hence we motivate the use of generalized operator valued kernels from a function estimation and learning methodology viewpoint, which allows us to relax the requirement of positive definite kernels in learning function-valued functions. However, we will strive to improve the motivation.

Reviewer 4: Thank you for your enlightening comments. Our replies follow. **(1)** Our stabilization problem (4) in Section 3, inspired from (Ong et al., 2004) helps in deriving the result in Representer Theorem 3.1. On the other hand, when the stabilizer $\widetilde{F}_\lambda$ from Eq. (4) belongs to the ball $\mathcal{B}_\mathcal{K}$ of fixed radius r (defined in Section 5 with $r = 1$), it enjoys the generalization bounds in Eq. (8). At least to us it is not very clear how the stabilizer behaves when it does not belong to $\mathcal{B}_\mathcal{K}$. One might suppose that the formulation similar to (Oglic and Gärtner, 2018) can be used here. However, adapting the minimization problem formulation in (Oglic and Gärtner, 2018) would lead to integral variance constraints in our case. Further, using a Gateaux derivative approach for the constrained or unconstrained minimization problem similar to that in (Oglic and Gärtner, 2018), leads to difficulties in obtaining the Representer Theorem 3.1 in our paper. As a consequence of these facts, we can only resort to an empirical cross-validation approach which we have used in our experiments to ensure that the stabilizer of problem (4) is not far away from $\mathcal{B}_\mathcal{K}$. **(2)** We will correct the typos.

References

Alpay, D. (1991). Some remarks on reproducing kernel krein spaces. *Journal of Mathematics 21*(4).

Carmeli, C., E. De Vito, and A. Toigo (2006). Vector valued reproducing kernel hilbert spaces of integrable functions and mercer theorem. *Analysis and Applications 4*(4), 377–408.

Carmeli, C., E. De Vito, A. Toigo, and V. Umanitá (2010). Vector valued reproducing kernel hilbert spaces and universality. *Analysis and Applications 8*(1), 19–61.

Maurer, A. (2016). A vector-contraction inequality for rademacher complexities. In *ALT*.

Oglic, D. and T. Gärtner (2018). Learning in reproducing kernel kreın spaces. In *ICML*.

Ong, C. S., X. Mary, S. Canu, and A. J. Smola (2004). Learning with non-positive kernels. In *ICML*.