Supplementary Material for "A mapping-based universal Kriging model for order-of-addition experiments in drug combination studies"

Qian Xiao[a,*], Hongquan Xu[b]

[a]*Department of Statistics, University of Georgia, Athens, Georgia 30602, U.S.A.*
[b]*Department of Statistics, University of California, Los Angeles, California 90095, U.S.A*

---

1. Additional results for the Case Study: a Five-Drug Order-of-addition Experiment on Lymphoma Treatment

In this part, we target on comparing the proposed UK and MUK models to the main-PWO, triplet-PWO and CP models under random designs. We split the 40-run full data in the manuscript into various training and test sets, and calculate the performance measures R_1, R_2 and RMSE. Specifically, we consider seven different random sub-designs to form the training sets with run sizes ranging from 12 to 36. For each corresponding case, we use runs left in the full data as the test set. An n-run random sub-design is denoted as RD_n here. We fit the UK, MUK, main-PWO, triplet-PWO and CP models with the training data selected by RD_n, and then measure their performances via R_1, R_2 and RMSE. For each n, we replicate the study 50 times and calculate both the mean and median of the results. We show results of R_1, R_2 and RMSE in the form of "$mean(median)$" in Tables 1, 2 and 3, respectively. In all tables, we mark "NA" when the number of runs is less than the number of parameters in the model. We mark results with asterisks when numerical errors arise in the model estimation, and we exclude such problematic cases in calculating the mean or median. When using the full data as training set, there will be no R_2 and RMSE values, because there is no test set. As there is no replicate in D_{full}, we set $\tau^2 = 1$ according to the measurement accuracy in this experiment. Different choices of small τ^2 will not lead to

*Corresponding author
Email address: qx69137@uga.edu (Qian Xiao)

significantly different results (Xiao et al., 2019). When fitting the full data, the correlation between actual and predicted responses using the proposed models is roughly 1 for small τ^2.

Table 1: Mean (median) results of R_1 for different models in the case study

	CP	Main-PWO	Triplet-PWO	UK	MUK
RD_{12}	NA	0.34/(0.39)*	NA	0.69(0.71)	NA
RD_{16}	NA	0.50(0.52)*	NA	0.72(0.76)	NA
RD_{20}	0.49(0.52)*	0.64(0.65)*	NA	0.80(0.84)	0.80(0.83)
RD_{24}	0.63(0.62)	0.69(0.71)	NA	0.87(0.88)	0.88(0.89)
RD_{28}	0.72(0.75)	0.72(0.73)	NA	0.90(0.91)	0.91(0.93)
RD_{32}	0.80(0.81)	0.75(0.76)	0.80(0.83)	0.94(0.94)	0.95(0.96)
RD_{36}	0.83(0.83)	0.76(0.77)	0.95(0.96)	0.97(0.97)	0.98(0.98)
D_{full}	0.85	0.78	0.98	1.00	1.00

Table 2: Mean (median) results of R_2 for different models in the case study

	CP	Main-PWO	Triplet-PWO	UK	MUK
RD_{12}	NA	0.22(0.28)*	NA	0.52(0.55)	NA
RD_{16}	NA	0.36(0.38)*	NA	0.53(0.58)	NA
RD_{20}	0.33(0.38)*	0.48(0.51)*	NA	0.61(0.67)	0.61(0.67)
RD_{24}	0.34(0.34)	0.52(0.53)	NA	0.65(0.69)	0.67(0.69)
RD_{28}	0.43(0.47)	0.52(0.54)	NA	0.67(0.72)	0.72(0.79)
RD_{32}	0.53(0.63)	0.59(0.64)	0.52(0.56)	0.67(0.72)	0.76(0.79)
RD_{36}	0.47(0.62)	0.49(0.57)	0.69(0.81)	0.66(0.80)	0.78(0.84)

From Tables 1, 2 and 3, it is seen that both the UK and MUK models give larger R_1 and R_2, and smaller RMSE compared to the existing models in the cases of RD_{12} to RD_{32}. For the RD_{36} case, the triplet-PWO model performs well, but it is still worse than the MUK model in terms of all performance measures. It is also worse than the UK model in terms of the RMSE. Note that when the test set only includes a few data (e.g. 4 data for the 36-run case), the RMSE is a more reliable performance measure compared to the correlation (R_2). In addition, we can see that the proposed UK model has satisfactory prediction accuracy using small experimental designs. As a comparison, the triplet-PWO model can only be applied to large designs. Both the UK and MUK models require less runs than the three existing methods to achieve similar prediction accuracy. For example, they only require

Table 3: Mean (median) results of RMSE for different models in the case study

	CP	Main-PWO	Triplet-PWO	UK	MUK
RD_{12}	NA	17.71(14.76)*	NA	5.87(5.75)	NA
RD_{16}	NA	9.41(8.84)*	NA	5.99(5.79)	NA
RD_{20}	14.37(12.61)*	7.38(7.15)*	NA	5.63(5.36)	5.55(5.43)
RD_{24}	9.85(9.50)	6.58(6.38)	NA	5.20(5.14)	5.01(4.76)
RD_{28}	7.84(7.32)	6.34(6.41)	NA	5.13(5.13)	4.76(4.65)
RD_{32}	6.46(6.21)	5.91(5.87)	11.71(9.71)	4.94(4.89)	4.34(4.18)
RD_{36}	6.41(6.43)	6.05(6.25)	6.28(5.84)	5.04(4.77)	3.84(3.68)

20 runs to give better predictions compared to all existing models using 32 runs. Note that median results are often better than the mean results in the analysis, which suggests some skewness in performances using random designs. In conclusion, the proposed UK and MUK models perform better than the three existing models in this case study.

Regarding Table 3, we would like to point out that the naive model with only an intercept does better in terms of the mean RMSE than the CP model for $n = 20$, 24 and 28 cases, better than the main-PWO model for $n = 12$, 16 and 20 cases, and better than the triplet-PWO model for the $n = 32$ case. By contrast, the UK and MUK models beat the intercept-only model for all sample sizes.

When comparing the UK and MUK models, they perform similarly for RD_{20} and the latter performs better for larger training sets: RD_{24}, RD_{28}, RD_{32} and RD_{36}. Generally speaking, the MUK model may provide better predictions than the UK model when the input-output relationships are very complex and sufficient data are available, since it includes more parameters and is more flexible. When only a few observations are available, the UK model is recommended.

Note that numerical errors arise when fitting the CP and main-PWO models for the cases of RD_{16} and RD_{20} (marked with asterisks). The CP model will give errors when some column in the design matrix does not use all drugs. The PWO models may suffer from multicolinearity problems even if the design matrix does not include duplicated columns, since they take PWO indicators instead of the order-of-addition inputs. For small random designs, such problems can be common. As a comparison, the proposed UK and MUK models can be applied to various designs.

References

Xiao, Q., Wang, L., Xu, H., 2019. Application of kriging models for a drug combination experiment on lung cancer. Statistics in Medicine 38, 236–246.