1 Additional analysis for the simulation study in Section 5.1

We also assess models with regard to estimation of the physical parameters, $\boldsymbol{\theta}_1$. For all models the parameters in $\boldsymbol{\theta}_1$ were estimated by their posterior means. To do the comparison, we present in Table 1 the RMSE averaged over 100 independently generated datasets. As before, the row labels represent the data set, while the column labels represent the model used to fit the data sets, and the numbers in parenthesis are the RMSE standard errors.

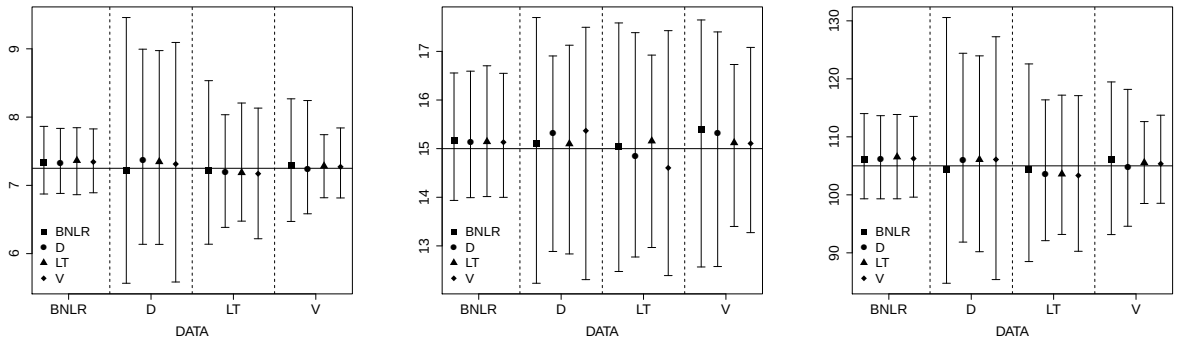
Table 1: RMSE when estimating $\boldsymbol{\theta}_1$ - Non-misaligned data

	BNLR	D	LT	V
BNLR	3.74 (2.95)	3.80 (3.05)	3.94 (3.12)	3.75 (2.87)
D	12.27 (8.00)	8.07 (6.66)	8.18 (6.31)	11.48 (7.33)
LT	8.11 (5.92)	6.45 (4.78)	5.93 (4.48)	7.19 (5.37)
V	7.49 (4.89)	6.34 (4.27)	3.58 (2.89)	3.98 (3.32)

Table 1 reaffirms what we have seen with the model comparison metrics. Models generating the data usually perform well when fitted to that dataset. When data was generated from the BNLR models, all models estimate $\boldsymbol{\theta}_1$ similarly. We also do not find any significant differences between the RSME's from the true model and from the PBBM with LT structure.

Figures 1(a)-1(c) present summaries of the estimates that each model provided according to each dataset. The x-axis indicates the dataset, while the square, circle, triangle and diamond represent the median of the 100 independent parameter estimates according to the BNLR, D, LT and V models, respectively. Moreover, the vertical solid line represents the 2.5% and 97.5% percentiles of the parameter estimates, and the horizontal solid line depicts the true value of the parameter. These figures show that when BNLR generates the data, the BNLR and the PBBM's perform similarly. On the other hand, when PBBM's generate the data, the LT has the narrowest 95% percentile intervals of the estimates around the true

values, while the BNLR has the widest 95% credible intervals of the estimates.



(a) β (b) Q (c) G

Figure 1: Median and 95% percentile interval of the estimates for β, Q and G - Non-misaligned data

Lastly, the story is unaltered when we compare BNLR and the PBBM's with respect to cross-validation using the hold out points. We computed the RMSE and the fraction of hold out points that lie within the 95% predictive interval (results not showed). The PBBM's present slightly smaller RMSE's than the BNLR when data are generated from PBBM's. When data are generated from BNLR, the models have essentially equal RMSE's. In terms of predictive interval coverage, all models perform very similarly.

2 Additional analysis of misaligned experimental data in Section 5.2

We present in Figure 2 a panel showing the bivariate posterior predictive distributions for concentrations (in the logarthmic case) in the near and far fields as estimated by PBBM with the LT structure. In principle, our framework can produce these panels as a "movie" in continuous time. Here, we show some snapshots at nine timepoints. In each graphic, the x-axis and y-axis represent the log-concentrations in the near and far fields respectively. The first

row of the panel shows density estimates at three timepoints measuring only the near field. Likewise, the second rows presents density estimates at three timepoints measuring both fields and, finally, the third row presents density estimates at three timepoints measuring only the far field. These images complete the distributional profile of the concentrations for the duration of the entire experiment and elicit the dynamic nature of the joint distribution over time.

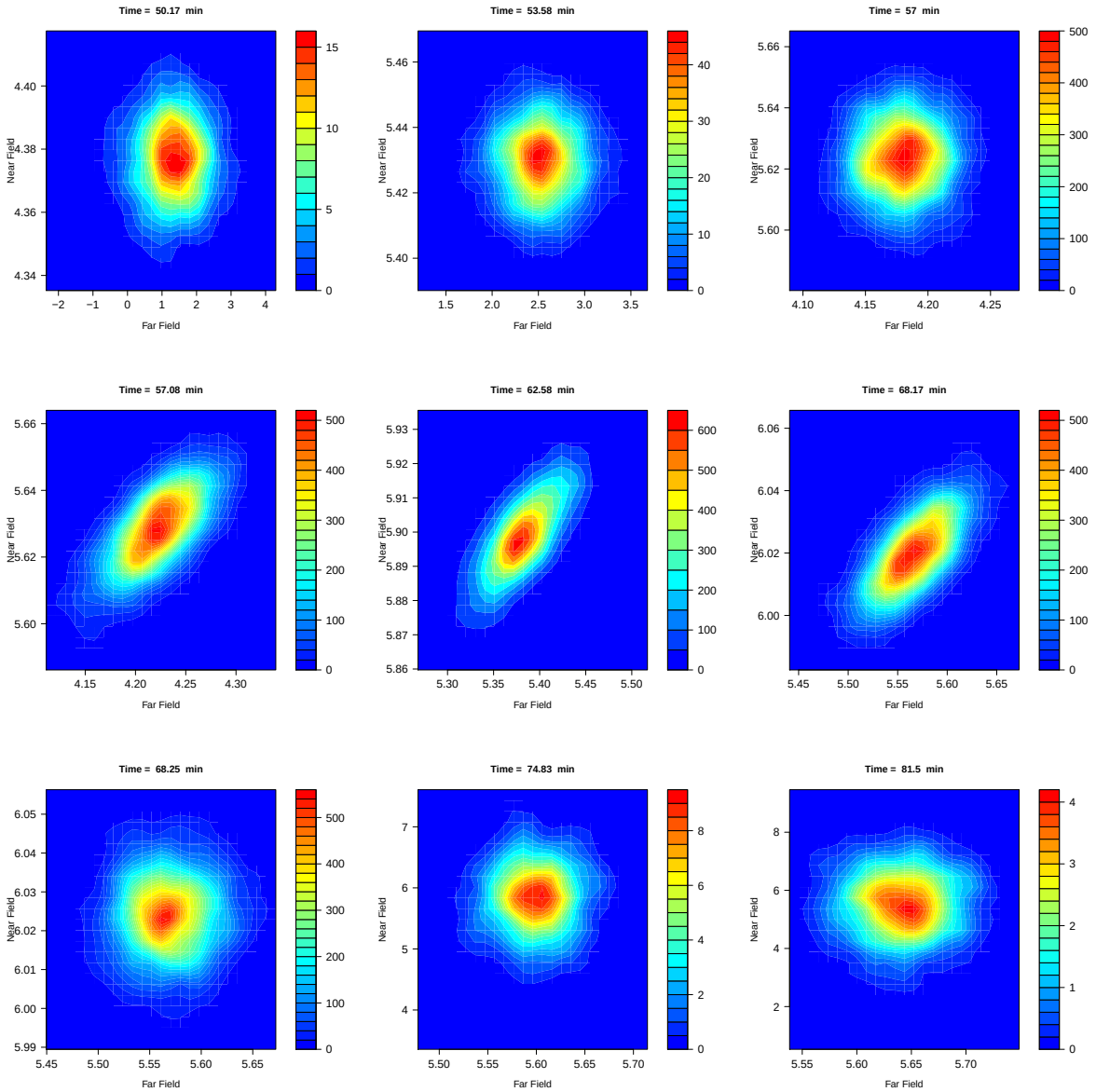


Figure 2: Posterior predictive joint distribution of the log-concentrations in the near and far fields at a selection of timepoints, as estimated from the PBBM with LT structure.