Supplement to "A Comparison of Bayesian Multivariate Versus Univariate Normal Regression Models for Prediction"

Xun Li Joyee Ghosh Gabriele Villarini

In this Supplement to the article "A Comparison of Bayesian Multivariate Versus Univariate Normal Regression Models for Prediction", we report i) the second level models for missing covariates needed for the TC activity application, ii) prediction results for the TC activity data when forecasts are issued in June and July, iii) estimated correlation matrix of residuals, and iv) additional simulation results to check for sensitivity analysis.

Second Level Models for Missing SST Forecasts in June, July, and August

The SST forecasts of CMC2$_{Atl}$ and CMC2$_{Trop}$ had missing observations in June, July, and August, so we specified a model for them in the second level. In addition, in June, NASA$_{Atl}$ and NASA$_{Trop}$ had missing observations, so those required a model. We provide a list of the predictors that were included in the second level models below. These were chosen after considering model diagnostics for normality and independence.

June				July		August	
NASA$_{Atl}$	NASA$_{Trop}$	CMC2$_{Atl}$	CMC2$_{Trop}$	CMC2$_{Atl}$	CMC2$_{Trop}$	CMC2$_{Atl}$	CMC2$_{Trop}$
intercept	intercept	intercept	intercept	intercept	intercept	intercept	intercept
GFDLA$_{Atl}$			GFDLA$_{Atl}$	GFDLA$_{Atl}$	GFDLA$_{Atl}$	GFDLA$_{Atl}$	GFDLA$_{Atl}$
GFDLB$_{Atl}$			GFDLB$_{Atl}$	GFDLB$_{Atl}$	GFDLB$_{Atl}$	GFDLB$_{Atl}$	GFDLB$_{Atl}$
GFDL$_{Atl}$	GFDL$_{Atl}$		GFDL$_{Atl}$	GFDL$_{Atl}$	GFDL$_{Atl}$	GFDL$_{Atl}$	GFDL$_{Atl}$
GFDLA$_{Trop}$		GFDLA$_{Trop}$	GFDLA$_{Trop}$	GFDLA$_{Trop}$	GFDLA$_{Trop}$	GFDLA$_{Trop}$	GFDLA$_{Trop}$
GFDLB$_{Trop}$			GFDLB$_{Trop}$	GFDLB$_{Trop}$	GFDLB$_{Trop}$	GFDLB$_{Trop}$	GFDLB$_{Trop}$
GFDL$_{Trop}$	GFDL$_{Trop}$	GFDL$_{Trop}$	GFDL$_{Trop}$	GFDL$_{Trop}$	GFDL$_{Trop}$	GFDL$_{Trop}$	GFDL$_{Trop}$
	NASA$_{Atl}$		NASA$_{Atl}$	NASA$_{Atl}$	NASA$_{Atl}$	NASA$_{Atl}$	NASA$_{Atl}$
			NASA$_{Trop}$	NASA$_{Trop}$	NASA$_{Trop}$	NASA$_{Trop}$	NASA$_{Trop}$
			CMC2$_{Atl}$		CMC2$_{Atl}$		CMC2$_{Atl}$
						GFDLA$_{Atl_{July}}$	GFDLA$_{Atl_{July}}$

Table 1: The covariates for the second level linear regression models.

Leave-One-Out Prediction Results for TC Activity Data for SST Forecasts Issued in June and July

Response	Method	Cor.Pearson	Cor.Spearman	RMSE	MAE	Coverage	Length
TS	ss-ind-ind	0.44	0.31	2.96	2.15	0.88	11.19
	ss-dep-ind	0.33	0.18	2.69	2.36	1.00	11.98
	ss-ind-dep	0.43	0.31	2.98	2.16	0.88	11.09
	ss-dep-dep	0.30	0.13	2.72	2.39	1.00	12.07
Hurricane	ss-ind-ind	0.40	0.49	2.43	1.98	0.75	7.68
	ss-dep-ind	0.45	0.48	2.20	1.74	0.88	8.37
	ss-ind-dep	0.39	0.49	2.46	2.00	0.75	7.59
	ss-dep-dep	0.47	0.48	2.20	1.73	0.88	8.47
PDI	ss-ind-ind	0.69	0.67	0.59	0.45	0.88	2.19
	ss-dep-ind	0.68	0.31	0.67	0.54	0.88	2.33
	ss-ind-dep	0.67	0.67	0.60	0.46	0.88	2.16
	ss-dep-dep	0.66	0.31	0.68	0.55	0.88	2.35
ACE	ss-ind-ind	0.68	0.55	0.43	0.35	0.88	1.65
	ss-dep-ind	0.61	0.36	0.48	0.42	0.88	1.77
	ss-ind-dep	0.66	0.55	0.44	0.36	0.88	1.63
	ss-dep-dep	0.59	0.36	0.49	0.42	0.88	1.79

Table 2: Results for June with linear regression models.

Response	Method	Cor.Pearson	Cor.Spearman	RMSE	MAE	Coverage	Length
TS	ss-ind-ind	0.63	0.69	2.30	1.69	1.00	11.01
	ss-dep-ind	0.59	0.66	2.26	2.00	1.00	11.50
	ss-ind-dep	0.63	0.67	2.33	1.69	1.00	10.92
	ss-dep-dep	0.55	0.66	2.32	2.06	1.00	11.64
Hurricane	ss-ind-ind	0.35	0.44	2.48	1.96	0.88	7.74
	ss-dep-ind	0.49	0.49	2.14	1.70	0.88	8.30
	ss-ind-dep	0.34	0.44	2.51	2.00	0.88	7.68
	ss-dep-dep	0.51	0.49	2.14	1.72	0.88	8.47
PDI	ss-ind-ind	0.52	0.55	0.67	0.49	0.75	2.11
	ss-dep-ind	0.71	0.52	0.67	0.53	0.88	2.28
	ss-ind-dep	0.50	0.55	0.68	0.50	0.75	2.08
	ss-dep-dep	0.71	0.60	0.67	0.54	0.88	2.30
ACE	ss-ind-ind	0.55	0.60	0.48	0.39	0.88	1.55
	ss-dep-ind	0.67	0.71	0.47	0.37	0.88	1.69
	ss-ind-dep	0.54	0.60	0.49	0.39	0.88	1.53
	ss-dep-dep	0.67	0.71	0.47	0.38	0.88	1.72

Table 3: Results for July with linear regression models.

Estimates of the Residual Correlation Matrix

Response	Correlation matrix of the observed data				The estimated correlation matrix			
	TS	Hurricane	PDI	ACE	TS	Hurricane	PDI	ACE
TS	1.00	0.86	0.73	0.78	1.00	0.76	0.55	0.63
Hurricane	0.86	1.00	0.83	0.87	0.76	1.00	0.73	0.79
PDI	0.73	0.83	1.00	0.99	0.55	0.73	1.00	0.96
ACE	0.78	0.87	0.99	1.00	0.63	0.79	0.96	1.00

Table 4: **Dep-data-dep-prior Spike-and-slab**: Correlation in the observed data and the estimated correlation matrix of residuals for June.

Response	Correlation matrix of the observed data				The estimated correlation matrix			
	TS	Hurricane	PDI	ACE	TS	Hurricane	PDI	ACE
TS	1.00	0.86	0.73	0.78	1.00	0.72	0.50	0.59
Hurricane	0.86	1.00	0.83	0.87	0.72	1.00	0.70	0.76
PDI	0.73	0.83	1.00	0.99	0.50	0.70	1.00	0.96
ACE	0.78	0.87	0.99	1.00	0.59	0.76	0.96	1.00

Table 5: **Dep-data-dep-prior Spike-and-slab**: Correlation in the observed data and the estimated correlation matrix of residuals for July.

Response	Correlation matrix of the observed data				The estimated correlation matrix			
	TS	Hurricane	PDI	ACE	TS	Hurricane	PDI	ACE
TS	1.00	0.86	0.73	0.78	1.00	0.72	0.50	0.59
Hurricane	0.86	1.00	0.83	0.87	0.72	1.00	0.66	0.73
PDI	0.73	0.83	1.00	0.99	0.50	0.66	1.00	0.95
ACE	0.78	0.87	0.99	1.00	0.59	0.73	0.95	1.00

Table 6: **Dep-data-dep-prior Spike-and-slab**: Correlation in the observed data and the estimated correlation matrix of residuals for August.

Additional Simulation Results for Sensitivity Analysis

We make the following changes to the *High correlation* scenario (please refer to Section 4.1). Our goal is to show that most results in the paper are not sensitive to these changes, and our main conclusions remain the same.

1. *A different set of hyperparameters for the inverse Wishart distribution for the residual covariance matrix* $\boldsymbol{\Sigma}$: Following Kundu *et al.* (2021), we choose the prior on $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} \sim IW\left(\nu, d_{\Sigma} \boldsymbol{I}_q\right)$$

   where we set the hyperparameters at $\nu = q + 2$ and $d_{\Sigma} = 1$. This choice of hyperparameters gives the prior mean for $\boldsymbol{\Sigma}$ as the identity matrix. In the rest of the paper, we use $d_{\Sigma} = 0.5$, since the variables are already standardized to have variance 1, and $d_{\Sigma} = 1$ puts more prior mass on values that are not possible under such standardization.

2. *Different initial values*: Here we change the initial values for the Markov chains from the null model (in the rest of the paper) to the full model.

3. *Mean vs. Median*: Here we report RMSE calculated using the mean of the posterior predictive distribution instead of the median in the rest of the paper.

4. *HPD prediction intervals*: Here we evaluate the length and coverage of 90% HPD prediction intervals instead of equal-tailed intervals in the rest of the paper.

The results after making the aforementioned changes, averaged over 100 datasets, are presented in Table 7. For convenience, we report the original results for $H = 0.95$ from the paper in Table 8 below.

Methods	ME	$\text{RMSE}_{\text{median}}$	$\text{RMSE}_{\text{mean}}$	Length		Coverage	
				Equal-tailed	HPD	Equal-tailed	HPD
ss-ind-ind	0.247	1.015	1.015	3.686	3.675	0.924	0.923
ss-dep-ind	0.125	1.002	1.002	3.630	3.617	0.924	0.922
ss-ind-dep	0.227	1.013	1.013	3.675	3.664	0.924	0.923
ss-dep-dep	0.124	1.001	1.001	3.628	3.615	0.924	0.923
hs-ind-ind	0.432	1.036	1.036	3.759	3.748	0.923	0.924
hs-dep-ind	0.204	1.010	1.010	3.695	3.684	0.927	0.925
hs-ind-dep	0.394	1.030	1.030	3.689	3.678	0.919	0.917
hs-dep-dep	0.211	1.011	1.011	3.685	3.674	0.924	0.924

Table 7: **Additional results** for all methods for $H = 0.95$. Here $\text{RMSE}_{\text{median}}$ refers to the RMSE using medians of posterior predictive distributions, $\text{RMSE}_{\text{mean}}$ refers to the RMSE using means of posterior predictive distributions, "ss-ind-ind" refers to the ind-data-ind-prior spike-and-slab, "ss-dep-ind" refers to the dep-data-ind-prior spike-and-slab, "ss-ind-dep" refers to the ind-data-dep-prior spike-and-slab, "ss-dep-dep" refers to the dep-data-dep-prior spike-and-slab, "hs-ind-ind" refers to the ind-data-ind-prior horseshoe, "hs-dep-ind" refers to the dep-data-ind-prior horseshoe, "hs-ind-dep" refers to the ind-data-dep-prior horseshoe, and "hs-dep-dep" refers to the dep-data-dep-prior horseshoe.

Methods	ME	$\text{RMSE}_{\text{median}}$	$\text{RMSE}_{\text{mean}}$	Length (Equal-tailed)	Coverage (Equal-tailed)
ss-ind-ind	0.260	1.017	1.017	3.485	0.907
ss-dep-ind	0.124	1.002	1.001	3.432	0.908
ss-ind-dep	0.239	1.015	1.014	3.475	0.908
ss-dep-dep	0.124	1.001	1.001	3.430	0.907
hs-ind-ind	0.454	1.038	1.038	3.541	0.904
hs-dep-ind	0.202	1.010	1.010	3.476	0.910
hs-ind-dep	0.423	1.034	1.033	3.472	0.900
hs-dep-dep	0.214	1.011	1.011	3.477	0.907

Table 8: **Original results** for all methods for $H = 0.95$. Here $\text{RMSE}_{\text{median}}$ refers to the RMSE using medians of posterior predictive distributions, $\text{RMSE}_{\text{mean}}$ refers to the RMSE using means of posterior predictive distributions, "ss-ind-ind" refers to the ind-data-ind-prior spike-and-slab, "ss-dep-ind" refers to the dep-data-ind-prior spike-and-slab, "ss-ind-dep" refers to the ind-data-dep-prior spike-and-slab, "ss-dep-dep" refers to the dep-data-dep-prior spike-and-slab, "hs-ind-ind" refers to the ind-data-ind-prior horseshoe, "hs-dep-ind" refers to the dep-data-ind-prior horseshoe, "hs-ind-dep" refers to the ind-data-dep-prior horseshoe, and "hs-dep-dep" refers to the dep-data-dep-prior horseshoe.

References

Kundu, D., Mitra, R., and Gaskins, J. T. (2021). Bayesian variable selection for multioutcome models through shared shrinkage. *Scandinavian Journal of Statistics* **48**, 295–320.