Cholesky-based multivariate Gaussian regression

Thomas Muschinski[1,2], Georg J. Mayr[2], Thorsten Simon[1,2], Achim Zeileis[1]

[1] Department of Statistics, Universität Innsbruck, Innsbruck, Austria
[2] Department of Atmospheric and Cryospheric Science, Universität Innsbruck, Innsbruck, Austria

E-mail for correspondence: Thomas.Muschinski@uibk.ac.at

Abstract: Multivariate Gaussian regression has applications in many fields, but is made difficult by the high model complexity and positive-definite requirement on the estimated covariance. We implement multivariate Gaussian regression through a Cholesky-based reparameterization of the covariance matrix. The distributional parameters—the means and the entries of the Cholesky factor—can be made to depend on covariates through flexible additive predictors, allowing for nonlinear variations in mean and covariance. The reparameterization is compared to reference methods for estimating a fixed covariance. An application for weather prediction (surface temperature) illustrates the flexibility of the approach.

Keywords: Covariance modeling; Cholesky decomposition; Multivariate Gaussian; MCMC simulation.

1 Cholesky-based multivariate Gaussian regression

Multivariate modeling has a wide range of applications from longitudinal analyses of biomarker data to postprocessing of numerical weather predictions. Employing multivariate Gaussian distributions in the framework of distributional regression allows one to specify very flexible models. For the bivariate Gaussian case, the correlation may be modeled directly (e.g. Klein et al. 2015), but for higher dimensions two main difficulties occur: (i) high complexity resulting from the large number of distributional parameters and (ii) ensuring a positive definite covariance. To tackle the latter issue, we factorize the covariance by the Cholesky decomposition (Pourahmadi 2011). To deal with its high complexity, we regularize the Cholesky-based multivariate Gaussian regression models (Umlauf et. al 2018).

This paper was published as a part of the proceedings of the 35th International Workshop on Statistical Modelling (IWSM), Bilbao, Spain, 19–24 July 2020.

The Cholesky decomposition of a positive definite symmetric matrix Σ has the form

$$\Sigma = LL^{\mathrm{T}} \qquad \text{and} \qquad \Sigma^{-1} = L^{-1\mathrm{T}}L^{-1}, \tag{1}$$

and is unique if the main diagonal of the lower triangular L is positive. The log-likelihood of a multivariate Gaussian distribution for the k-dimensional observation vector y can then be written in terms of μ and L^{-1} by

$$\ell(\mu, L^{-1}|y) = -\frac{k}{2}\log(2\pi) + \log(|L^{-1}|) - \frac{1}{2}(y-\mu)^{\mathrm{T}}(L^{-1})^{\mathrm{T}}L^{-1}(y-\mu). \tag{2}$$

We designate the nontrivial elements of $L^{-1\mathrm{T}}$ by λ_{ij}, with $i \leq j$, and link all distributional parameters to additive models:

$$\mu_i = \eta_{\mu,i}, \quad \log(\lambda_{ii}) = \eta_{\lambda,ii}, \quad \text{and} \quad \lambda_{ij} = \eta_{\lambda,ij} \quad \text{for} \quad i < j. \tag{3}$$

The reparameterization is available as a family for the R package **bamlss** (Umlauf et. al 2018) that implements optimizers for regularized estimation.

2 Simulation study

We test the proposed regression method with data simulated from a known multivariate Gaussian distribution of dimension 10. The distribution has zero mean, heteroscedastic marginal variances $\Sigma_{ii} = i$ and a first order autoregressive correlation matrix with $\rho = 0.5$.

Two different model setups are used to estimate the true distributional parameters from 50 simulated y and the process is repeated 1000 times. In Model 1, all η_i (see Eq. 3) are modeled as intercepts only. Model 2 is the same as Model 1 except that off-diagonal entries of L^{-1} (i.e. λ_{ij}, $i \neq j$) are regularized with a ridge penalty.

The estimates' representations of the true covariance and precision is evaluated using the spectral norm of the corresponding matrix differences, and compared to three reference methods for covariance estimation: (i) the sample covariance, (ii) a shrinkage covariance estimate and (iii) the graphical lasso (glasso).

The unregularized Model 1 has similar performance to the sample covariance; the regularized Model 2 performs better than both the shrinkage estimate and glasso (Fig. 1). For estimating a stationary covariance structure, the proposed multivariate distributional regression approach performs well despite the number of distributional parameters (65) exceeding the number of simulated vectors used for estimation (50). The true strength of the method, though, lies in the flexible manner in which distributional parameters can be modeled on covariates.

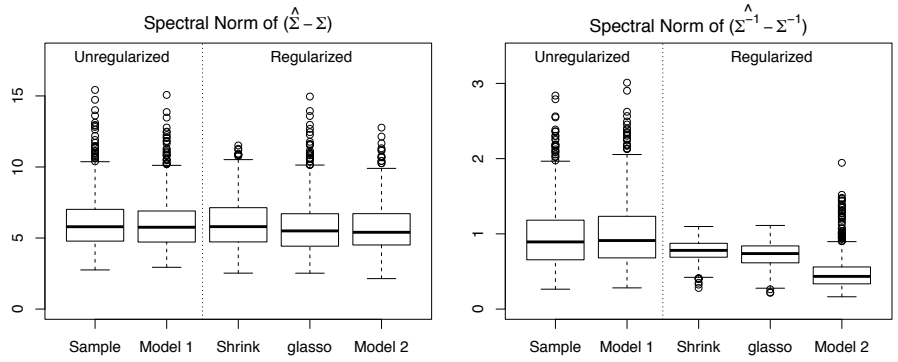

FIGURE 1. Spectral norm of differences between the estimated and true covariance (left) and precision matrices (right). Boxplots represent 1000 simulations. Smaller values indicate that the estimated covariance (precision) matrices are closer to the truth.

3 Multivariate forecasting of surface temperatures

The goal of numerical weather prediction is forecasting future atmospheric states from current observations using governing physical equations. The resulting predictions are postprocessed by statistical methods to improve their skill. For forecasting the temporal evolution of surface temperature over several future (lead) times, the error correlation between lead times must be considered. Our proposed method accomplishes this task with a multivariate approach by postprocessing the predictions (GEFS reforecasts, Hamill 2013) for several lead times simultaneously.

To illustrate, we model 00 UTC surface temperature at Innsbruck, Austria, for 8 lead times (+8 d, +9 d, ..., +15 d) with an 8-dimensional Gaussian distribution. Seasonal variations in both predictive skill and error correlations are permitted by letting Cholesky factor entries depend on the day of the year (`yday`) and mean parameters have a linear dependency on the corresponding forecasts $\texttt{ens}_i$, but with seasonally varying coefficients:

$$\begin{aligned} \mu_i &= (\beta_{0,i} + f_{0,i}(\texttt{yday})) + (\beta_{1,i} + f_{1,i}(\texttt{yday})) \cdot \texttt{ens}_i \\ \log(\lambda_{ii}) &= \beta_{0,ii} + f_{ii}(\texttt{yday}) \\ \lambda_{ij} &= \beta_{0,ij} + f_{ij}(\texttt{yday}), \end{aligned} \tag{4}$$

where f are nonlinear cyclical functions of `yday`.

Five years of data were used to estimate the model parameters and reveal pronounced seasonal cycles in the effects of the μ models (Fig. 2). Each of the modeled λ_{ij} are also allowed to have such seasonal dependencies, which are significant for $i = j$ and also for several λ_{ij} with lag 1 (i.e. $j = i + 1$). At higher lags, the seasonal effects become insignificant.

Seasonally varying Cholesky factor estimates ($\widehat{L^{-1}}$) result in distinct $\widehat{\Sigma}$ for every `yday`. Taking $\widehat{\Sigma}$ for January 1 and July 1, we see that not only are variances in winter nearly twice as large as in summer, the errors are also more

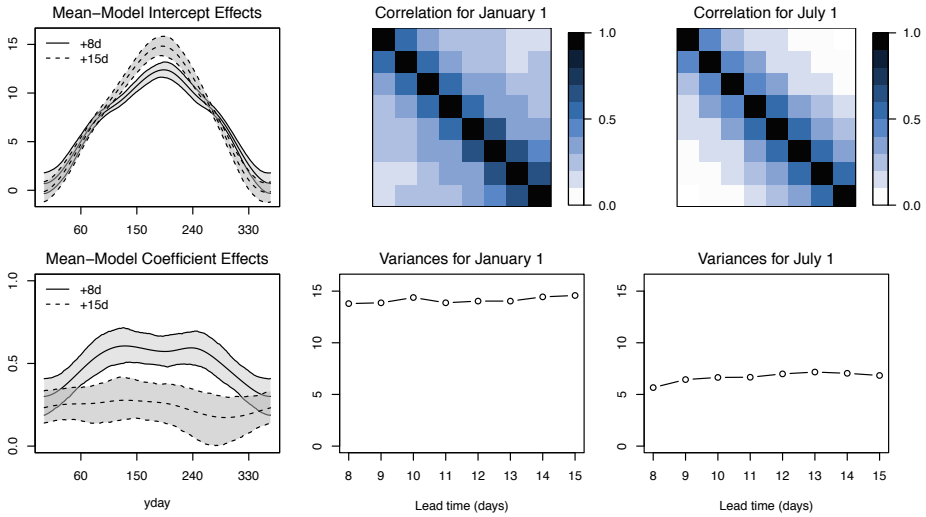

FIGURE 2. **Left column:** Estimated mean-model effects for $\beta_{0,i} + f_{0,i}(\texttt{yday})$ in Eq. 4 (top) and $\beta_{1,i} + f_{1,i}(\texttt{yday})$ (bottom). **Center column:** Correlation matrix (top) and marginal variances (bottom) calculated from the Cholesky factor estimated for January 1. **Right column:** Correlation and variances for July 1.

strongly correlated (Fig. 2). This is the benefit of the proposed multivariate Gaussian regression method: flexible mean and covariance estimation, while ensuring positive-definiteness and enabling data-driven regularization.

Acknowledgments: This project was funded by the Austrian Science Fund (FWF, grant no. P 31836). We thank the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) for providing the observational data.

References

Hamill, Bates, Whitaker et al. (2013). *NOAA's second-generation global medium-range ensemble reforecast dataset*. B. Am. Meteorol. Soc., **94(10)**, 1553–1565. doi: 10.1175/BAMS-D-12-00014.1.

Klein, Kneib, Klasen and Lang (2015). *Bayesian structured additive distributional regression for multivariate responses*. J. Roy. Stat. Soc. C, **64(4)**, 569–591. doi: 10.1111/rssc.12090.

Pourahmadi (2011). *Covariance estimation: The GLM and regularization perspectives*. Statistical Science, **26(3)**, 369–387. doi: 10.1214/11-STS358

Umlauf, Klein and Zeileis (2018). *BAMLSS: Bayesian additive models for location, scale, and shape (and beyond)*. J. Comput. Graph. Stat, **3**, 612–627. doi: 10.1080/10618600.2017.1407325.