

dImmc: Dynamical linear model regression for atmospheric time-series analysis

Justin A. Alsing^{1, 2, 3}

1 Oskar Klein Center for Cosmoparticle Physics, Stockholm University, Stockholm **2** Center for Computational Astrophysics, Flatiron Institute, New York **3** Imperial Centre for Inference and Cosmology, Imperial College London, London

DOI: [10.21105/joss.01157](https://doi.org/10.21105/joss.01157)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 21 December 2018

Published: 03 May 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

Regression analysis of atmospheric time-series observations is a key endeavour for identifying long-term trends and studying underlying drivers of variability in the data. Multiple linear regression (MLR) has been a standard tool for climate and atmospheric time-series data analysis. However, MLR has a number of well-known shortcomings that can, in the worst case, lead to biased scientific inferences being drawn from the data. Dynamical linear modeling (DLM) provides a more flexible regression framework that addresses a number of the key issues faced by MLR (reviewed below), providing an attractive and more robust alternative.

DLMMC provides a suite of DLM models for atmospheric time-series analysis, with a user friendly python interface, for use by the community. Models are implemented in stan (Carpenter et al., 2017), and use a combination of efficient Hamiltonian Monte Carlo (HMC) sampling and Kalman filtering to explore the (Bayesian) posterior distribution of the model parameters given the data.

The basic set of DLM models are based on (Laine, Latva-Pukkila, & Kyrölä, 2014) and have four main components: a dynamic seasonal cycle (with 6- and 12-month components), a smooth non-linear background trend, forcings from any number of user defined proxy variables (eg., solar activity, quasi-biennial oscillations, the El-Nino southern oscillation, etc), and an auto-regressive (AR) process. These models were originally built with stratospheric research in mind, but can be applied or modified for other climate/atmospheric time-series analysis (or more general time-series analysis problems).

The DLM models in this suite can be readily extended to more complex models with additional components; DLMMC provides a general-purpose and efficient HMC sampling-Kalman filtering framework for DLM regression analysis.

The code has already been used in a number of scientific publications, including (Ball et al., 2017) and (Ball et al., 2018). For an introduction to DLM in the context of stratospheric time-series analysis, see (Laine et al., 2014) (on which this code was inspired), or see (Durbin & Koopman, 2012) for a more comprehensive review of time-series analysis with state-space models.

Advantages of DLM over traditional MLR methods

Dynamical linear models provide a rich framework for time-series regression analysis. Let us briefly review some of the key advantages offered by DLM over commonly used MLR approaches (in the context of stratospheric time-series analysis):

Flexible non-linear background trends DLM uses a flexible (non-parametric) model for the non-linear background trend, where the degree of non-linearity is included as a free parameter that is fit along with the rest of the regression model. That way, the data are free to choose how rapidly (in time) the background trend is allowed to vary. In contrast, independent-linear trends (ILT) or piecewise-linear trends (PWLT) that are commonplace in MLR trend analyses provide very restrictive models for the background trend. For example, typically the positions of any inflection point(s) in the ILT/PWLT must be fixed in advance even though they are not necessarily known a priori, and the resulting trends are sensitive to these choices. Such restrictive trend-model assumptions such as ILT or PWLT severely hamper our ability to let the data speak for themselves in terms of identifying long-term background trends, particularly when those trends might be surprising relative to our prior expectations.

Dynamical seasonal and regressor modulation DLM allows the amplitudes of the seasonal cycle and forcings via proxy variables (such as solar activity, quasi-biennial oscillations, the El-Nino southern oscillation, etc) to vary dynamically with time, whereas MLR has fixed (in time) regression coefficients. This extra flexibility in the model enables DLM to capture more of the variability in the data due to the evolving state of the atmosphere, or evolution in the observing conditions and/or sampling that can lead to changes in the distributional properties of the data with time.

Principled treatment of auto-regressive processes In many situations, phenomenological regression models (such as DLM and MLR) only capture some fraction of the total variability in the data, ie., there are correlated processes left in the residuals that are not captured by the main features of the regression model. This is often mitigated by including an auto-regressive (AR) process as a surrogate for the un-modelled variability. DLM infers an auto-regressive process simultaneously with the rest of the model parameters, and carefully propagates uncertainties by formally marginalizing over uncertainties in the AR-process when reporting errors on the recovered trend, seasonal cycle etc. In contrast, MLR typically applies a post-hoc correction to the recovered regression coefficients and error bars to account for auto-regressive correlations (eg., the Cochrane–Orcutt correction); this is often approximate, and less principled in its propagation of uncertainties.

Principled treatment of time-varying uncertainties (heteroscedasticity) Finally, many implementations of MLR in the literature use ordinary least squares (OLS) estimation for the regression coefficients. This approach relies on the assumption that the error distribution on observations is constant in time (homoscedasticity), and can lead to biased parameter estimates when this condition is not met (heteroscedasticity). DLM implements heteroscedastic time-varying uncertainties as standard.

Acknowledgements

I thank Will Ball, Daniel Mortlock, Marko Laine and Sean Davis for useful discussions and comments on the code.

References

Ball, W. T., Alsing, J., Mortlock, D. J., Rozanov, E. V., Tummon, F., & Haigh, J. D. (2017). Reconciling differences in stratospheric ozone composites. *Atmospheric Chemistry and Physics*, 17(20), 12269–12302. doi:[10.5194/acp-17-12269-2017](https://doi.org/10.5194/acp-17-12269-2017)

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., et al. (2018). Evidence for a continuous decline in lower stratospheric ozone offsetting

ozone layer recovery. *Atmospheric Chemistry and Physics*, 18(2), 1379–1394. doi:[10.5194/acp-18-1379-2018](https://doi.org/10.5194/acp-18-1379-2018)

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., et al. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1). doi:[10.18637/jss.v076.i01](https://doi.org/10.18637/jss.v076.i01)

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). Oxford University Press. doi:[10.1093/acprof:oso/9780199641178.001.0001](https://doi.org/10.1093/acprof:oso/9780199641178.001.0001)

Laine, M., Latva-Pukkila, N., & Kyrölä, E. (2014). Analysing time-varying trends in stratospheric ozone time series using the state space approach. *Atmospheric Chemistry and Physics*, 14(18), 9707–9725. doi:[10.5194/acp-14-9707-2014](https://doi.org/10.5194/acp-14-9707-2014)