The Journal of Open Source Software

hetGP4cast: adventures in ecological forecasting

Maike Holthuijzen [1*], **R. Quinn Thomas**[1*], **Cayelan C. Carey**[1*], **and Robert M. Gramacy**[1¶]

1 Virginia Tech, USA ¶ Corresponding author * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 14 January 2024
Published: unpublished

In partnership with

AMERICAN ASTRONOMICAL SOCIETY

This article and software are linked with research article DOI NA, published in the NA.

Summary

Climatological models are useful for forecasting ecological phenomena, as they can capture long-term trends and tendencies. Climatological models also provide important baseline forecasts for researchers developing and testing new ecological forecasting models; for example, climatological models can generate 'null model' forecasts that can be used to determine the relative improvement in skill of new models. Many ecological variables that are increasingly being forecasted are characterized by heteroscedasticity over time and other dimensions. However, climatological models typically do not account for input-dependent variability, limiting their use for ecological forecasting. The incorporation of a non-constant variance over at least one dimension would improve uncertainty quantification (UQ) of climatological forecast models and could dramatically improve the prediction accuracy and quality of UQ of baseline forecasts. Heteroscedastic Gaussian process models (hetGPs) are ideal for constructing climatological models; however, while hetGPs can be fitted with existing software packages (i.e. the `hetGP` package Binois & Gramacy (2021) in R), they are not as easily accessible to non-statisticians. In response to this need, we developed the `hetGP4cast` package to fit hetGPs to data targeted for environmental applications.

Statement of need

Climatological forecasts are often used as baseline or reference models for ecological forecasting, but none currently account for heteroscedasticity. Typically, the mean and standard deviation of climatological forecasts used for ecological forecasting are calculated directly as statistics from long-term climatologies. Empirical means and standard deviations are suboptimal for forecasting, because unequal sample sizes used in calculations are ignored, making UQ less robust. There is a critical need for a more sophisticated statistical modeling approach that incorporates a non-constant variance. Such an approach would involve fitting a hetGP model to long-term climatological records. A hetGP climatological model would result in more accurate predictions and, importantly, much more reliable UQ, resulting in better quality of baseline forecasts used for ecological forecasting. The `hetGP4cast` package was specifically designed for this purpose. Unlike existing tools (i.e., the `hetGP` package by Binois & Gramacy (2021)), the `hetGP4cast` R package empowers ecological forecasters, especially those without an extensive background in Gaussian process modeling, to easily fit, generate predictions, and plot results from heteroscedastic Gaussian process models. By adopting `hetGP4cast`, ecologists and other non-statisticians can enhance the accuracy of their climatological models, resulting in more reliable predictions and, crucially, enhanced UQ. Looking ahead, the `hetGP4cast` package represents a valuable resource for the ecological forecasting challenge community. We believe that its intuitive interface and powerful capabilities will prove useful to ecologists aiming to construct robust climatological models. The package bridges the gap between statistical sophistication and accessibility, ensuring that even non-statisticians can contribute meaningfully to advancing cutting edge research in ecological forecasting. To see the full capabilities of the

hetGP4cast package, please see the vignette here https://rpubs.com/dutchie/vignette.

Acknowledgements

We acknowledge the Ecological Forecasting Project team (especially Mary Lofton, Freya Olsson, and Austin Delany) for helpful comments and feedback. We acknowledge funding from the U.S. National Science Foundation grants 1933016, 2327030, and 2318861.

References

(**article?**){binois2021hetgp, title={hetgp: Heteroskedastic Gaussian process modeling and sequential design in R}, author={Binois, Micka{"e}l and Gramacy, Robert B}, volue={98}, pages={1–44}, year={2021} }

Binois, M., & Gramacy, R. B. (2021). Hetgp: Heteroskedastic gaussian process modeling and sequential design in r. *Journal of Statistical Software*, *98*, 1–44.