Data from fitting Gaussian process models to various data sets using eight Gaussian process software packages

Erickson, Collin B.; Ankenman, Bruce E.; Sanchez, Susan M.

Elsevier

Erickson, Collin B., Bruce E. Ankenman, and Susan M. Sanchez. "Data from fitting Gaussian process models to various data sets using eight Gaussian process software packages." Data in brief 18 (2018): 684-687.
http://hdl.handle.net/10945/60927

Downloaded from NPS Archive: Calhoun
Data Article

Data from fitting Gaussian process models to various data sets using eight Gaussian process software packages

Collin B. Erickson⁎, Bruce E. Ankenman, Susan M. Sanchez
Northwestern University, Naval Postgraduate School, United States

A R T I C L E   I N F O

Article history:
Received 12 October 2017
Accepted 5 December 2017
Available online 12 December 2017

A B S T R A C T

This data article provides the summary data from tests comparing various Gaussian process software packages. Each spreadsheet represents a single function or type of function using a particular input sample size. In each spreadsheet, a row gives the results for a particular replication using a single package. Within each spreadsheet there are the results from eight Gaussian process model-fitting packages on five replicates of the surface. There is also one spreadsheet comparing the results from two packages performing stochastic kriging. These data enable comparisons between the packages to determine which package will give users the best results.

Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications Table

Subject area Operations research
More specific subject area Gaussian process modeling
Type of data Tables

DOI of original article: https://doi.org/10.1016/j.ejor.2017.10.002
⁎ Corresponding author.
E-mail address: collinerickson@u.northwestern.edu (C.B. Erickson).
How data was acquired
Simulation

Data format
Analyzed

Experimental factors
Software package, function type, number of input points

Experimental features
Summary measures of fit are provided for each software package on each function type for various input sample sizes.

Data source location
Evanston, IL, USA

Data accessibility
Available with article

Value of the data

- This data set contains the differences between various software implementations of Gaussian process modeling that use the same set of fitting equations and input data.
- This data set can be used as a benchmark to compare new software.
- Practitioners may use the data to determine which software package they use in their research.

1. Data

This Data in Brief is associated with the paper Comparison of Gaussian process modeling software [1]. That paper provides details about the collection of the data and analysis of the results. The data consists of 10 spreadsheets. Each spreadsheet contains the data for all packages on a particular input function and a particular input sample size.

2. Experimental design, materials and methods

The packages used are the R packages DiceKriging, GPfit, laGP, and mlegp; the Python modules GPy and scikit-learn; the Matlab toolbox DACE; and JMP. Some of these are used with multiple configurations of the fitting options, and thus have different designations. For example, mlegp is used

| Designation | Package | Settings used |
|-------------|---------|---------------|
| DiceM52     | DiceKriging | The nugget is set to zero and the Matern 5/2 correlation function is used. |
| Dice0       | DiceKriging | The nugget is set to zero. |
| GPfit1.95   | GPfit   | The nugget is set to a stable number, and the exponent in the correlation function is 1.95. |
| GPfit2      | GPfit   | The nugget is set to a stable number, and the exponent in the correlation function is 2. |
| laGP6       | laGP    | The nugget is set to 1e-6. |
| laGPE       | laGP    | The nugget is estimated. |
| mlegp0      | mlegp   | The nugget is set to zero. |
| mlegpE      | mlegp   | The nugget is estimated. |
| JMP0        | JMP     | The nugget is set to zero. |
| JMPE        | JMP     | The nugget is estimated. |
| DACE        | DACE    | The nugget is set a small value. |
| GPy         | GPy     | The nugget is estimated. |
| sklearn     | sklearn | The nugget is set to a small value. |
both with the nugget set to zero (designated mlegp0), and with the nugget estimated (designated mlegpE). All of the designations are described in Table 1 and further explained in the paper. In addition, the spreadsheets also include the results from fitting a linear model (LM), a quadratic model (QM), and a model that just predicts the mean of the input values (PredictMean). These can be used to evaluate whether it is worthwhile to use a Gaussian process model instead of simpler models.

We perform tests using four functions: the borehole function, the OTL circuit function, the Dette and Pepelyshev function, and the Morris function. We include results for two input sample sizes for each of these. For each function, we perform five replicates.

The input space is a unit cube where the range of each dimension is zero to one. The input values are scaled appropriately to match the function domains. For each replicate, the input data set is selected by taking points from a maximin Latin hypercube, which helps ensure the data is spread over the input space. The number of points used in the input data set is selected to allow for reasonable fitting of the surfaces based on our experience. After each package fits a model to the input data, the model is used to make predictions for the mean and variance at 2000 points. The testing points are also taken from a maximin Latin hypercube. Thus the packages can be compared on each replicate since they use the exact same data.

The output table in each spreadsheet contains the following columns:

- **Fit**: The designation for the package and settings used to fit the data. This is explained further in the paper.
- **EMRMSE**: The empirical model root-mean-square error of all 2000 predictions over the surface. Smaller values mean the predictions are more accurate.
- **PMRMSE**: The predicted model root-mean-square error of all 2000 predictions over the surface. This should be close to the EMRMSE if the model predicted variances are accurate.
- **POARMSE**: The predicted RMSE over actual RMSE, or PMRMSE/EMRMSE. Values near 1 are preferred. Values smaller than 1 mean that the predictions are overconfident, and values larger than 1 mean that the predictions are conservative. This is not mentioned in the paper but can be a useful tool for exploring the data.
- **PWBRMSE**: The proportion that the EMRMSE is worse than the best EMRMSE on that replicate from all packages. The best fit on the replicate will have a PWBRMSE of 0, a PWBRMSE of 1 means that the replicate has an EMRMSE twice as large as the best EMRMSE on that replicate. Smaller values are better. This is also not mentioned in the paper but also helps make comparisons between packages easy.
- **xi**: The EMRMSE divided by the EMRMSE of the linear model on that replicate. Small values mean the Gaussian process does much better than a linear model.
- **pi**: The PMRMSE divided by the EMRMSE of the linear model on that replicate. This value should be near xi if the model predicted variances are accurate.
- **Rep**: The number of the replicate, between 1 and 5.
- **RunTime**: The time in seconds the entire fitting and prediction process took.

The spreadsheets SK-MM1_D1_SS5_100_R5_SupplementaryTable.csv and SK-MM1_D1_SS5_200_R5_SupplementaryTable.csv are different from the rest since they perform stochastic kriging, which allows for noisy data. The only two packages that are able to perform stochastic kriging are DiceKriging and mlegp, so we only include these two along with a linear model in the data. For each of the five replicates, the data is gathered in two stages. In the first stage, 5 replicates are collected at each of the design points: 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. In the second stage, 100 or 200 points are allocated to these design points according to their sample variances in the first stage. The model is fit to all the data from the first two stages. The predictions are tested at 300 points equally spaced from 0.3 to 0.9. The data in these two spreadsheets is not the same data as shown in [1] since the random seeds were lost. However, they were generated the same way and show similar results. The output tables in these spreadsheets are formatted similarly to those in the other spreadsheets.
Acknowledgements

U.S. Department of Defense Distribution Statement: Approved for public release; distribution is unlimited. The views expressed in this document are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. This work was supported in part by the Office of Naval Research via NPS’s CRUSER initiative and the Naval Supply Systems Command Fleet Logistics grant number N00244-15-2-0004.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.12.012.

Reference

[1] C.B. Erickson, B.E. Ankenman, S.M. Sanchez, Comparison of Gaussian process modeling software, European Journal of Operational Research 266 (1) (2018) 179–192.