

Machine Learning Project

BENJAMIN MEYER*

University of Basel
benja.meyer@unibas.ch

August 3, 2022

I. INTRODUCTION

In the original paper [Zwe+21] shallow landslides in Switzerland and their causal factors are investigated. They are using the group lasso algorithm [YL04] for estimating coefficients over a set of causal factors (or features). The evaluated coefficients are interpreted as an indicator of the importance for each feature to get an understanding which features drive land slides. To get a confidence interval, different model instances of the group lasso are trained by bootstrapping the data.

In this study project we want to investigate the same data but with a bayesian statistics approach. Rather than frequentist statistic confidence intervals we want to build bayesian statistic credible intervals. Instead of bootstrapping the group lasso, we use the Bayesian Group Lasso algorithm [XG] to approximate the distribution over the coefficients directly. The distribution, or samples from its approximation, can then be used to get the desired credible intervals.

For the Bayesian Group Lasso the software package MBSGS [Ben19] was chosen. Later this turned out to be a mistake, as this software package only supports Bayesian Group Lasso for normal distributed outputs and not for binomial distributed outputs. In other words we are solving a least square regression problem rather than a log loss classification problem. Applying linear regression on binary class labels is equivalent to the LDA algorithm, therefore we simply ignored this mistake. Note however that all here mentioned Bayesian re-

sults are Bayesian LDA with a group selection laplace prior model rather than the originally planned Bayesian Group Lasso!

Finally during the study project a bug in the original code was found and fixed.

The rest of this document is structured as follows. First we discuss the bug found in the original code (see section II). Then we discuss the reproduction of parts from the original result (see section III). Next we discuss the software package used for the Bayesian Group Lasso in more details (see section IV) as well as our results (see section V). Finally, we conclude this study project (see section VI).

II. BUG IN THE ORIGINAL PAPER

During investigating the code of [Zwe+21], we found a bug in accumulating the model brier score results of the bootstrap runs. The variable for accumulating the results of the cross validation (inner loop) is not reset each bootstrap iteration (outer loop), which leads to the results of the first of n bootstrap iteration being added n times, the results of the second bootstrap iteration being added $n - 1$ times and so on. Therefore, some results are counted multiple times as individual results leading to a too confident confidence interval in the brier score in figure 9 in [Zwe+21]. The bug is quite evident when investigating the file sizes of the bootstrap runs, as shown in figure 1.

In Figure 2 we show the confident intervals of the brier score for each site as in the original paper, our reproduction with the bug as well as our reproduction without the bug.

The bug has no effect on the coefficient figure

*A thank you or further information

R	Arosa.model.1.Gestkl...rpreg_grLasso_1.rda	19 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_2.rda	36 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_3.rda	52 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_4.rda	69 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_5.rda	84 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_6.rda	101 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_7.rda	115 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_8.rda	134 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_9.rda	149 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_10.rda	166 KB
R	Arosa.model.1.Gestkl...rpreg_grLasso_11.rda	182 KB

Figure 1: File sizes of bootstrap run results. Red marked are the index of the bootstrap runs. We would expect that all file sizes are equal regardless of the index as all files (should) contain the result of a single bootstrap iteration.

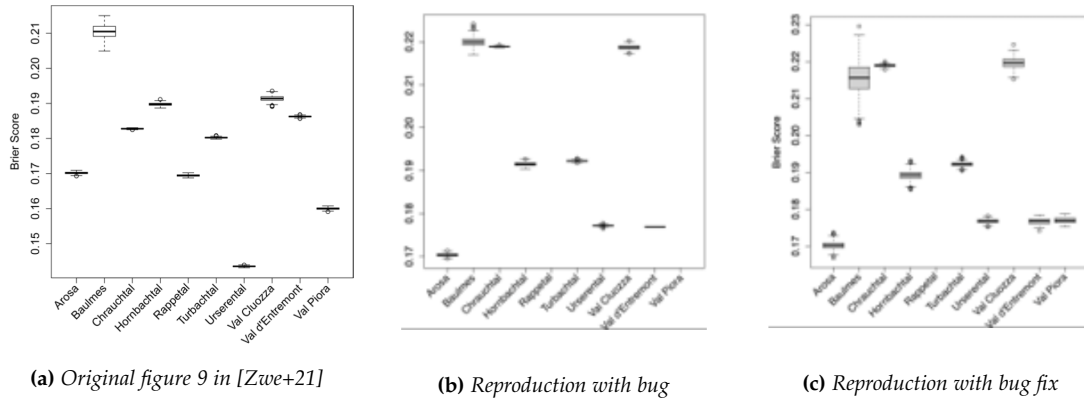


Figure 2: (a) original paper (b) reproduction without bugfix. (c) reproduction with bugfix. The bug caused duplication of samples therefore the confidence intervals was to narrow in the original paper. Some results differ in our reproduction (e.g Chrauchtal), the cause for this was not found.

7 in [Zwe+21], as the accumulating error was only done for the results used in figure 9.

III. REPRODUCING RESULTS OF ORIGINAL PAPER

In figure 3 we see the plot of the confidence intervals of the original paper as well as our reproduction. In the reproduction the first dummy variable of each categorical feature is removed. Otherwise, the reproduction is the same as the original paper and seems to be fine.

The reproduction was not straightforward as some used software packages had breaking changes in the meantime [Iss] and the preprocessing script had to be manually edited for each site.

IV. METHODS

i. Lasso

The Lasso is a Linear Regression with a L1-Regularization on the weights. The cost function is therefore the following:

$$\text{MSE}(\vec{y}, \vec{x}\vec{\beta}) + \|\vec{\beta}\|_1 \quad (1)$$

where:

$$\text{MSE} = \text{Mean Squared Error} \quad (2)$$

$$\vec{x} = \text{Input} \quad (3)$$

$$\vec{\beta} = \text{Weights} \quad (4)$$

$$\|\vec{\beta}\|_1 = \text{L1 Norm} \quad (5)$$

ii. Group Lasso

The Group Lasso puts features into groups. The idea being that features in a group are selected all together or not at all.

This is achieved by first calculating the L2 norm of the features inside a group and then calculating the L1 norm across all those group or rather L2 norms.

Note that the L2 norm has no inherit selection mechanism whereas the L1 norm does. Therefore, we achieve what we set out to do: We select features on the group level.

Written mathematically in an unusual style where we take the norms over sets, this looks like this:

$$\text{MSE}(\vec{y}, \vec{x}\vec{\beta}) + \|\{\|\beta_g\|_2 \mid \forall \beta_g \in G\}\|_1 \quad (6)$$

where:

$$G = \text{Partition of elements in } \beta \quad (7)$$

$$(8)$$

iii. Bayesian Group Lasso with Spike and Slab prior

The Bayesian treatment of the Lasso is simply a Laplace Prior on the weights.

The Bayesian treatment of the Group Lasso is the Bayesian Group Lasso. For the Bayesian Group Lasso the features are again grouped.

The prior per group β_g is then simply a Multivariate Laplacian:

$$\pi(\beta_g) \propto \exp\left(-\frac{\lambda}{\sigma} \|\beta_g\|_2\right) \quad (9)$$

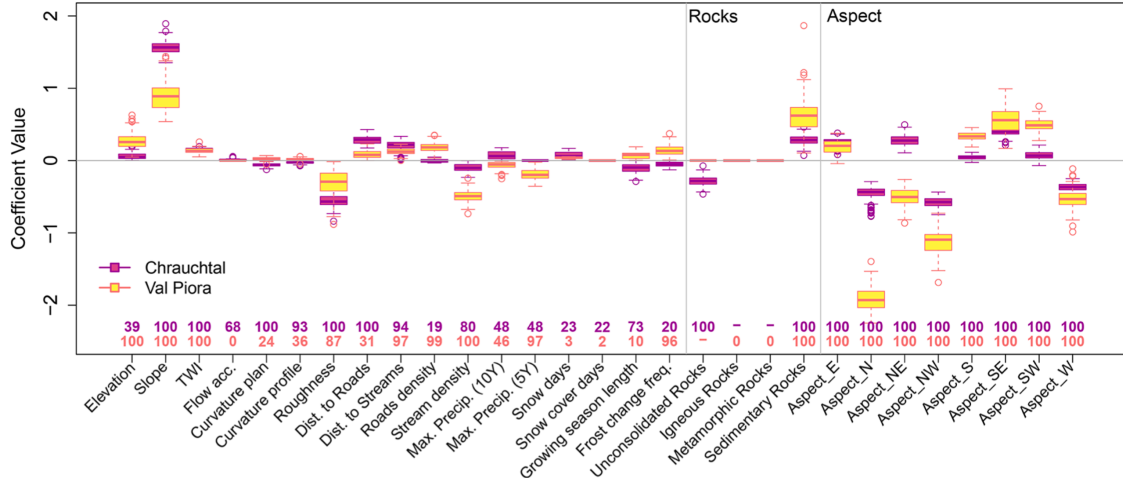
And the prior for the group partition G is simply the product of each priors:

$$\pi(G) = \prod_{\beta_g \in G} \pi(\beta_g) \quad (10)$$

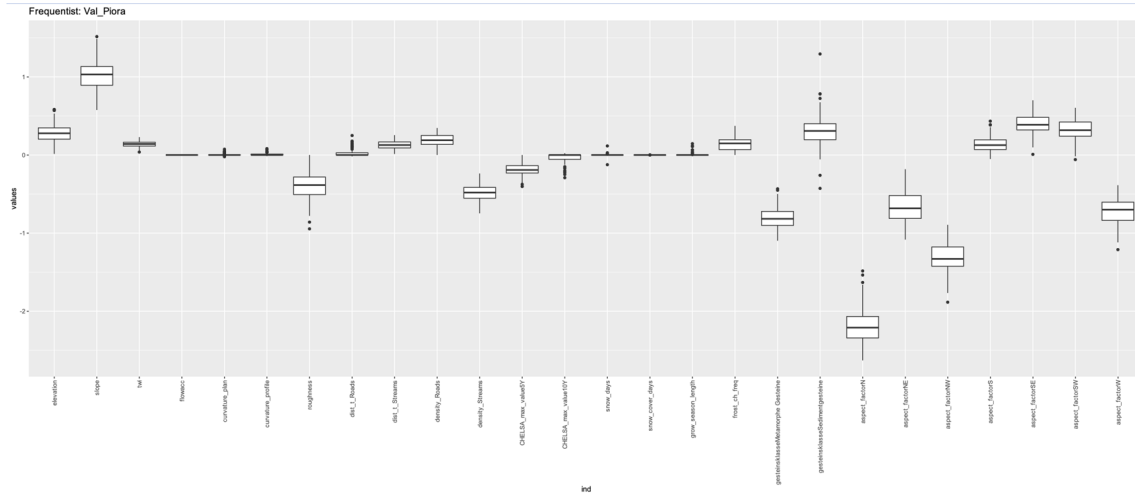
To actually work with those priors, we must rewrite them in a more useful form as described in section 2 of [XG]. Rather than doing it ourselves, we use the *BGLSS* method of the software package *MBSGS* [Ben19] based on [XG].

iv. BGLSS: Bayesian Group Lasso with Spike and Slab prior

We are using the *BGLSS* method of the software package *MBSGS* [Ben19]. The method



(a) Figure 7 from [Zwe+21]



(b) Reproduction of coefficients from site Val Piora (yellow boxplots from above figure)

Figure 3: (a) Show the original plot from [Zwe+21] as well as our (b) reproduction. Note that the reproduction seems to be correct. Only the categorical variables differ, which is explained as we drop the first dummy variable.

runs a gibbs sampler for sampling 10000 samples of the posterior distribution with a burn in phase of 5000 steps. The regularization strength parameter is chosen by the Monte Carlo EM algorithm [Cas01], so it must not be specified. In the group lasso of the original paper a nested cross validation for this hyper parameter selection is run.

V. RESULTS

In figure 4 we show the confidence and credible intervals for the site Val Piora. It shows the confidence we have in the value of the coefficient for a particular feature. The confidence intervals are created by the bootstrapped results, whereas the confidence intervals are created by the samples of the posterior distribution.

The credible intervals are similar to the confidence intervals. For features that are never or not often selected in the original approach (e.g snow dasys), the new approach seems to have more outliers. This could be due to the different regularization strength, which adds a different selection threshold. The scale of the coefficients differ, this could be due to the different cost function used in the new approach (LDA) as well as a different regularization strength, which causes a different shrinkage behavior.

VI. CONCLUSION

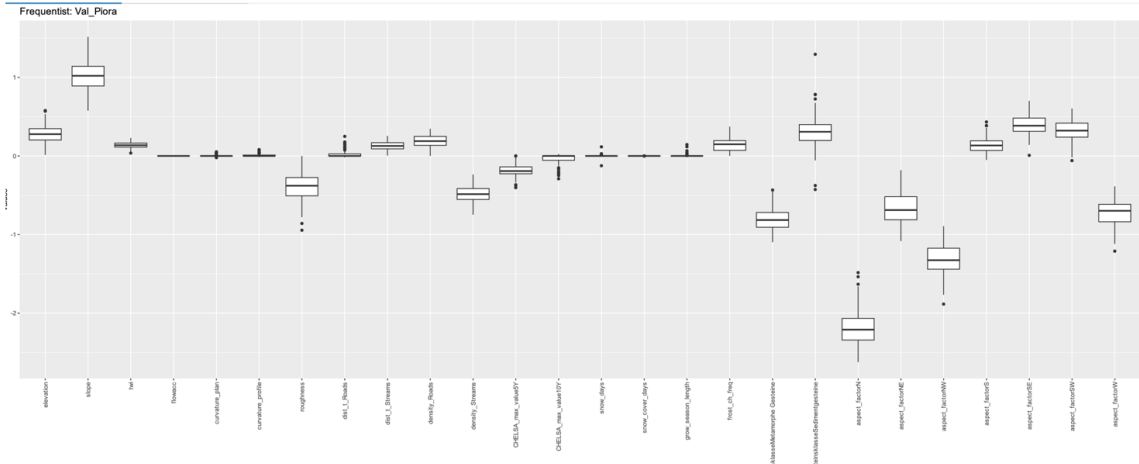
I wanted to do this study project to further my understanding of the bayesian approach covered in the machine learning lecture. In my daily work I often use frequentist models like Neural Networks or Gradient Boosting and I wanted to have a better understanding of how to use Bayesian models in practice.

In my opinion this was partly successful. I understand the Bayesian approach better due to reading papers, using the above-mentioned software package as well as analysing the results of the model. On the other hand the lack of in my opinion good software packages made the use of Bayesian models for me more complicated than expected. Initially I imagined

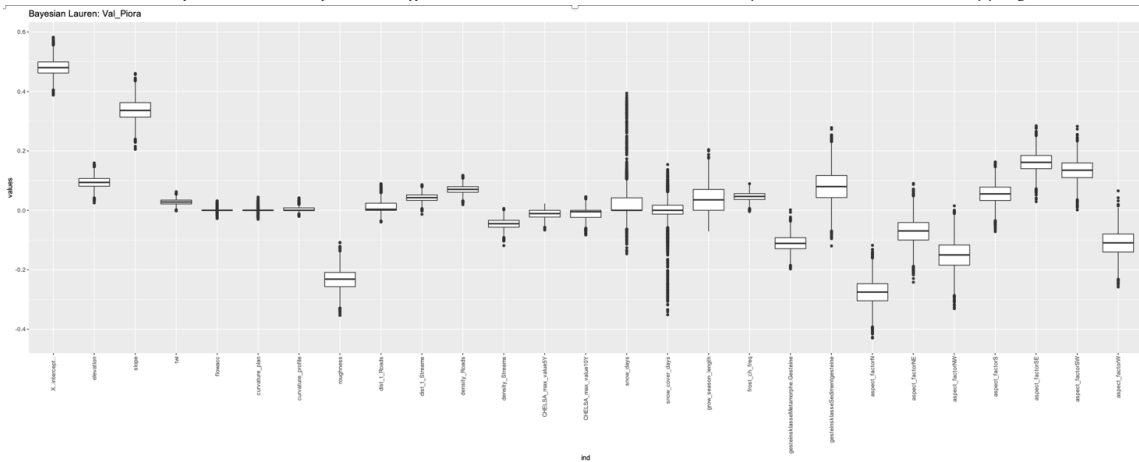
a software package for building and training Bayesian Nets, which would be, once learned, quit easy to apply to new problems (something like Tensorflow or PyTorch for Neural Networks, but for BayesNets).

I was also surprised at how much problems I had with the programming language R. As a software engineer that is quit fluid in multiple programing languages I had a hard time picking up R and run in many problems with data types and cryptic error messages. Driven by this frustration I even translated the hole preprocessing to Python, yet switched back to R as I did not find a implementation of the Bayesian Group Lasso in Python.

I want to thank Professor Volker Roth for the interesting technical inputs and Maxim for the guidance and pleasant communication throughout the project.



(a) Confidence interval for the coefficients at the site Val Piora with Group Lasso as well as bootstrapping



(b) Credible intervals for the coefficients at the site Val Piora with Bayesian LDA with group selection laplace prior.

Figure 4: *Results for the site Val Piora*

REFERENCES

- [Cas01] George Casella. *Empirical Bayes Gibbs sampling*. Dec. 2001. URL: <https://academic.oup.com/biostatistics/article/2/4/485/321497>.
- [YL04] Ming Yuan and Yi Lin. 2004. URL: https://www.researchgate.net/publication/4993325_Model_Selection_and_Estimation_in_Regression_With_Grouped_Variables.
- [Ben19] Matthew Sutton Benoit Lique. *MB-SGS: Multivariate bayesian sparse group selection with Spike and slab version 1.1.0 from cran*. May 2019. URL: <https://rdrr.io/cran/MBSGS/>.
- [Zwe+21] Lauren Zweifel et al. *Investigating causal factors of shallow landslides in grassland regions of Switzerland*. Nov. 2021. URL: <https://nhess.copernicus.org/articles/21/3421/2021/nhess-21-3421-2021-discussion.html>.
- [Iss] Github Issue. *'spatialblock' fails for data in EPSG:3035 · issue 21 · rvalavi/BLOCKCV*. URL: <https://github.com/rvalavi/blockCV/issues/21#issuecomment-1031667857>.
- [XG] Xiaofan Xu and Malay Ghosh. *Bayesian variable selection and estimation for group Lasso*. URL: <https://projecteuclid.org/journals/bayesian-analysis/volume-10/issue-4/Bayesian-Variable-Selection-and-Estimation-for-Group-Lasso/10.1214/14-BA929.full>.