Regression of Fluctuating System Properties: Baryonic Tully-Fisher Scaling in Disk Galaxies

Geert Verdoolaege

Abstract In various interesting physical systems, important properties or dynamics display a strongly fluctuating behavior that can best be described using probability distributions. Examples are fluid turbulence, plasma instabilities, textured images, porous media and cosmological structure. In order to quantitatively compare such phenomena, a similarity measure between distributions is needed, such as the Rao geodesic distance on the corresponding probabilistic manifold. This can form the basis for validation of theoretical models against experimental data and classification of regimes, but also for regression between fluctuating properties. This is the primary motivation for geodesic least squares (GLS) as a robust regression technique, with general applicability. In this contribution, we further clarify this motivation and we apply GLS to Tully-Fisher scaling of baryonic mass vs. rotation velocity in disk galaxies. We show that GLS is well suited to estimate the coefficients and tightness of the scaling. This is relevant for constraining galaxy formation models and for testing alternatives to the Lambda cold dark matter cosmological model.

1 Introduction

In many parametric regression problems, robustness of the estimates is an essential criterion, sometimes even more important than goodness-of-fit. Here, by 'robustness' we mean not only resilience against outliers, but also relative insensitivity to model uncertainty. A multitude of techniques, Bayesian and non-Bayesian, have been developed ensuring robustness in the presence of various departures from the regression model. However, it can be difficult for the non-expert user to make the right choices of methods and implementation details. This constitutes a major ob-

Geert Verdoolaege
Department of Applied Physics, Ghent University, B-9000 Ghent, Belgium; Laboratory for Plasma Physics, Royal Military Academy, B-1000 Brussels, Belgium. e-mail: geert.verdoolaege@ugent.be

stacle for adoption of the right techniques by practitioners in various application domains with little tradition in the data sciences.

In this paper, we advocate the use of a simple but powerful robust regression method, called *geodesic least squares* (GLS), that was previously introduced in [1] and [2]. The purpose of the present contribution is, first, to generalize to a certain extent the theoretical underpinnings of the method, and second, to compare the performance of the method in a practical application from astronomy with other methods, including a standard robust Bayesian approach.

The motivation for GLS can be explained using a simple example that essentially describes a very common situation. Imagine the turbulent flow of a fluid through a pipe with a variable cross-section. The regression task consists of finding a relation between the flow speed of the fluid (response variable) and the cross-section of the pipe (predictor variable), based on the reading from flow meters positioned at different locations along the axis of the pipe. Of course, the flow speed measured by any of these meters fluctuates in time. Thus, even when the predictor variable is held fixed, the response variable is not constant. So far this is a common regression problem, which the vast majority of practitioners from applied fields would solve by calculating time averages and performing least squares regression between these average flow speeds and the measured cross-section of the pipe.[1] A standard maximum likelihood or Bayesian solution would be possible too, basing the likelihood on the distribution of velocity fluctuations (neglecting measurement uncertainty), which we assume to be known from a previous experiment. However, in addition to the turbulent fluctuations there can be other sources of uncertainty. In the present example this could for instance be due to a variable pumping speed due to a malfunctioning pump. If the flow readings are taken sequentially, this could introduce additional uncertainty not captured by the distributional properties of the intrinsic turbulent fluctuations. This is a case of incorrectly specified uncertainties, which can be handled using various Bayesian approaches depending on the specific problem (see e.g. [3]). The solution provided by GLS is to consider, on the one hand, the distribution (likelihood) of the flow velocity that would be expected if all assumptions regarding the deterministic and probabilistic components of the regression model were correct. We call this the *modeled distribution*. On the other hand, the distribution of the data is characterized in a generic way using as few assumptions as possible, referred to here as the *observed distribution*. Then, similar to minimization by the least squares method of the sum of squared Euclidean distances between a measurement of the dependent variable and its modeled value, GLS estimates the model parameters by minimizing the sum of squared Rao geodesic distances between the observed and modeled distribution. This introduces extra flexibility ('elasticity') in the analysis, which, in practice, yields excellent robustness properties. Effectively, GLS performs regression between probability distributions on a Riemannian probabilistic manifold. It can also be characterized as a minimum distance method, generalizing likelihood-based techniques, although there are important differences with standard minimum distance estimation (MDE). Another typical example of the ap-

[1] In fact, any measurement with finite precision is an average over some smaller scale, e.g. the measurement of the cross-section of the pipe.

plication of GLS was treated in [4], relating the properties of a repetitive instability in tokamak plasmas. The distributions of two properties of the instability were determined under stationary plasma conditions and then the regression was carried out between those distributions. Indeed, GLS can take into account uncertainty on all variables (predictor and response). There are many other examples where the GLS approach is natural, e.g. involving signals, images, porous media, cosmological structure, etc., although the method itself is of general applicability.

After explaining the motivation for GLS based on regression between fluctuating system properties, in this contribution we illustrate the applicability of GLS to common regression problems by estimating a key scaling law in astrophysics: the baryonic Tully-Fisher relation. This is a remarkably tight relation between the total baryonic mass of disk galaxies and their rotational velocity, of great practical and theoretical significance in astrophysics and cosmology.

2 GLS regression: principles and motivation

In parameter estimation problems like regression analysis, the likelihood compares measured quantities with their value predicted by the model, under stationary experimental conditions, determined by fixed or stationary predictor variables. Hence, the likelihood serves as a distance measure between the measurement and the model. Maximization of the joint likelihood for all measurements is equivalent to minimization of the Kullback-Leibler divergence (KLD) between the empirical ('observed') distribution and the theoretical ('modeled') distribution of the residuals. In general, MDE techniques can be made more robust against model uncertainty by relying on similarity measures other than the KLD. The Hellinger divergence (closely related to the Bhattacharyya distance) is a common choice [5], first applied to regression in [6].

We follow a somewhat different approach, minimizing the Rao geodesic distance (GD) between the observed and modeled distributions. Consider a parametric multiple regression model involving m predictor variables ξ_j ($j = 1,\ldots,m$) and a single response variable η, all assumed to be infinitely precise. Suppose that N measurements are acquired for the predictor variables, resulting in measurements ξ_{Ij} ($I = 1,\ldots,N$). The regression model can be written as follows:

$$\eta_I = f(\xi_{I1},\ldots,\xi_{Im},\beta_1,\ldots,\beta_p) \equiv f(\{\xi_{Ij}\},\{\beta_k\}), \qquad \forall I = 1,\ldots,N. \tag{1}$$

Here, f is the regression model function, in general nonlinear and characterized by p parameters β_k ($k = 1,\ldots,p$). In regression analysis within the astronomy community, it is customary to add a noise variable to the idealized relation (1). This so-called *intrinsic scatter* serves to model the intrinsic uncertainty on the theoretical relation, i.e. uncertainty not related to the measurement process. We take another route for capturing model uncertainty, however.

In any realistic situation, we have no access to the quantities η_I and ξ_{Ij}. Instead, we assume that at each 'measurement site' I a series of n_I measurements $x_{i_I j}$, resp. y_{i_I} is collected for the noisy predictor variables x_j and the response variable y ($i_I = 1,\ldots,n_I$). In this paper, we assume that the measurement model describes fluctuation of the data around a point that lies exactly on the regression function. This need not be the case in reality, which is one of the potential causes of model uncertainty. Nevertheless, if there are multiple measurements at each measurement site, then this can provide useful information on the true distribution of the data under stationary conditions. A common situation is where, at fixed I, the $x_{i_I j}$ and y_{i_I} represent measurements of noisy stationary signals. In the remainder of the paper we will assume independent Gaussian noise, but this can be generalized to multivariate or non-Gaussian distributions. In the independent Gaussian case, we have

$$\begin{aligned} y_{i_I} &= \eta_I + \varepsilon_{y,i_I}, \qquad \varepsilon_{y,i_I} \sim \mathcal{N}(0,\sigma^2_{y,I}), \\ x_{i_I j} &= \xi_{Ij} + \varepsilon_{x,i_I j}, \quad \varepsilon_{x,i_I j} \sim \mathcal{N}(0,\sigma^2_{x,Ij}). \end{aligned} \tag{2}$$

Notice that in general the standard deviations can be different at each measurement site. For instance, in many real-world situations, such as the one discussed in this paper, there is a constant relative error on the measurements, so the standard deviation can be modeled as being proportional to the measurement itself. Of course, the noise described by the $\sigma_{y,I}$ and $\sigma_{x,Ij}$ need not be the only source of uncertainty contributing to fluctuation of the data around the regression model. This is the case of interest in this paper, where other uncertainty sources such as model uncertainty are present (cf. the intrinsic scatter mentioned before), which could even be more important than the noise at the individual measurement sites and about which little is known. For now we assume that the standard deviations $\sigma_{y,I}$ and $\sigma_{x,Ij}$ were estimated prior to the regression analysis. This may be as simple as calculating the standard deviation of the y_{i_I} and $x_{i_I j}$ at each measurement site. We also include the possibility where $n_I = 1$ for some or all I, in which case the noise variables $\sigma_{y,I}$ and $\sigma_{x,Ij}$ could be given by the error bars obtained from previous experiments or an uncertainty analysis.

In reality, the true model points $(\eta_I, \xi_{I1}, \ldots, \xi_{Im})$ from which the data are assumed to be generated are unknown, but we can estimate them by calculating averages $\bar{y}_I \equiv 1/n_I \sum_{i_I}^{n_I} y_{i_I}$ and $\bar{x}_{Ij} \equiv 1/n_I \sum_{i_I}^{n_I} x_{i_I j}$, which are expected to be distributed according to $\mathcal{N}(0,\sigma^2_{y,I}/n_I)$ and $\mathcal{N}(0,\sigma^2_{x,Ij}/n_I)$, respectively. Now suppose that the model given by (1) and (2) were exact, meaning that $\sigma_{x,Ij}$ and $\sigma_{y,I}$ would characterize the only uncertainty sources, then the joint likelihood of the average data would be given by

$$p\big(\{\bar{y}_I\},\{\bar{x}_{Ij}\}\big|C_\xi\big)=\prod_{I=1}^{N}\frac{1}{\sqrt{2\pi/n_I}\sigma_{y,I}}\exp\left\{-\frac{1}{2}\frac{\Big[\bar{y}_I-f\big(\{\xi_{Ij}\},\{\beta_k\}\big)\Big]^2}{\sigma_{y,I}^2/n_I}\right\}$$
$$\times\prod_{j=1}^{m}\frac{1}{\sqrt{2\pi/n_I}\sigma_{x,Ij}}\exp\left\{-\frac{1}{2}\frac{\Big[\bar{x}_{Ij}-\xi_{Ij}\Big]^2}{\sigma_{x,Ij}^2/n_I}\right\}. \qquad (3)$$

Here, C_ξ stands for the collection $\{\beta_k\}$, $\{\xi_{Ij}\}$, $\{\sigma_{x,Ij}\}$, $\{\sigma_{y,I}\}$, the notation $\{\bar{x}_{Ij}\}$ referring to the set of $\bar{x}_{Ij}$ for all I and j, and similar for other sets. Also, we use the same indices for summation and for indicating set members, in order not to complicate the notation. As the ξ_{Ij} are not known, they have to be marginalized over. This is usually accomplished by decomposing the line from the measurement to the unknown point on the model in a perpendicular and parallel component w.r.t. the model, and assuming a uniform prior on the coordinates along the model surface [3, 7]. For a linear model, effectively this comes down to inserting the measurement values into the model equation, and propagating the uncertainty on the predictor variables through the model. Treatment of a nonlinear model is more complicated, but can be simplified by a linear approximation of the model in the vicinity of the model point nearest to the data point. Alternatively, one can perform Gaussian error propagation to obtain an approximate normal conditional likelihood for $\{\bar{y}_I\}$:

$$p_{\mathrm{mod}}\big(\{\bar{y}_I\}\big|C_x\big)=\prod_{I=1}^{N}\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{mod},I}}\exp\left\{-\frac{1}{2}\frac{\Big[\bar{y}_I-f\big(\{\bar{x}_{Ij}\},\{\beta_k\}\big)\Big]^2}{\sigma_{\mathrm{mod},I}^2}\right\}. \qquad (4)$$

In this expression, C_x stands for the collection $\{\beta_k\}$, $\{\bar{x}_{Ij}\}$, $\{\sigma_{x,Ij}\}$, $\{\sigma_{y,I}\}$. The uncertainty on the predictor variables propagates through the function f and adds to the conditional uncertainty on the response variable, determined by $\sigma_{\mathrm{mod},I}$. For example, referring to $f\big(\{\bar{x}_{Ij}\},\{\beta_k\}\big)$ as the modeled mean $\mu_{\mathrm{mod},I}$, for a linear model we have (with relabeled β_k):

$$\mu_{\mathrm{mod},I}\equiv\beta_0+\beta_1x_{I1}+\ldots+\beta_mx_{Im},$$
$$\sigma_{\mathrm{mod},I}^2\equiv\sigma_{y,I}^2+\beta_1^2\sigma_{x,I1}^2+\ldots+\beta_m^2\sigma_{x,Im}^2.$$

In the literature, uninformative priors for the model parameters β_k have been derived as well, based on transformation invariance [8]. We use these priors for comparison of GLS with the standard Bayesian analysis.

Now, suppose for a moment that one would proceed with the maximum likelihood method to estimate the parameters β_k. From (4), one sees that this is equivalent to minimization of the sum of squared Mahalanobis distances between each observed $\bar{y}_I$ and its corresponding value $f\big(\{\bar{x}_{Ij}\},\{\beta_k\}\big)$ determined by the model

function f.[2] The Mahalanobis distance can be regarded as the distance between two univariate Gaussian clusters of points with centroids given by $\bar{y}_I$ and $f\big(\{\bar{x}_{Ij}\},\{\beta_k\}\big)$, each with the same standard deviation, in the present case $\sigma_{\mathrm{mod},I}$. Interestingly, it is also a special case of the Rao GD, namely the GD between the corresponding normal distributions with those means and common standard deviation [9]. It is therefore natural to generalize this to the case where not only the means of the distributions, but also the standard deviations are allowed to differ. One could choose to generalize the Mahalanobis distance to the Bhattacharyya distance or the Hellinger divergence, but we prefer the Rao geodesic distance owing to its solid mathematical foundations and intuitive geometric interpretation.

By allowing the standard deviation of the observed and modeled distribution to be different, the method is rendered robust, as the actual distribution of the data is allowed to deviate from the modeled distribution. So, on the one hand, we consider at each measurement site I the modeled distribution $\mathcal{N}\Big(f\big(\{\bar{x}_{Ij}\},\{\beta_k\}\big),\sigma^2_{\mathrm{mod},I}\Big)$. On the other hand, we have the observed distribution p_{obs}, which has to rely on as few assumptions as possible regarding the regression model, in an attempt to 'let the data speak for themselves'. We here only assume that it also is a Gaussian distribution, $p_{\mathrm{obs}}=\mathcal{N}(\bar{y}_I,\sigma^2_{\mathrm{obs},I})$, centered on the actually observed average $\bar{y}_I$, and with an unknown standard deviation $\sigma_{\mathrm{obs},I}$, to be estimated from the data. Although this can all be generalized, the normal distribution offers a computational advantage, as the corresponding expression for the GD has a closed form [10]. In addition, we already mentioned that, in principle, $\sigma_{\mathrm{obs},I}$ can be different at each measurement site, but in practice it is clear that we will need to introduce some sort of regularization to render the model identifiable. In this paper we either assume $\sigma_{\mathrm{obs},I}$ a constant s_{obs}, or proportional to the response variable, $\sigma_{\mathrm{obs},I}=r_{\mathrm{obs}}|\bar{y}_I|$. The parameters s_{obs} or r_{obs} have to be estimated from the data. More complicated (parametrized) relations between $\sigma_{\mathrm{obs},I}$ and the response variable or other data would be possible too, but one should be careful not to put too many restrictions on p_{obs}, thereby defeating its purpose.

GLS now proceeds by minimizing the total GD between, on the one hand, the joint observed distribution of the N values $\bar{y}_I$ and, on the other hand, the joint modeled distribution. Owing to the independence assumption in this example, we can write this in terms of products of the corresponding marginal distributions (including all dependencies and with γ_{obs} either s_{obs} or r_{obs}):

$$\Big\{\beta_k,\gamma_{\mathrm{obs}}\Big\}=\underset{\beta_k,\gamma_{\mathrm{obs}}\in\mathbb{R}}{\operatorname{argmin}}\sum_{I=1}^{N}\mathrm{GD}^2\Big[p_{\mathrm{obs}}\left(Y|\bar{y}_I,\gamma_{\mathrm{obs}}\right),p_{\mathrm{mod}}\left(Y|C_x\right)\Big]. \tag{5}$$

Here, the variable Y models the site averages. In addition, note that the parameters β_k occur both in the mean and the variance of the modeled distribution. Furthermore, in (5) we have used the property that the squared GD between products of distributions

[2] Under the assumption of symmetry of the likelihood distribution and homoscedasticity, this reduces to minimization of the sum of squared differences (Euclidean distances) between each measured $\bar{y}_I$ and predicted $f\big(\{\bar{x}_{Ij}\},\{\beta_k\}\big)$.

can be written as the sum of squared GDs between the corresponding factors [10]. Hence, the optimization procedure involves, at each measurement site, matching not only $\bar{y}_I$ with $f\big(\{\bar{x}_{Ij}\},\{\beta_k\}\big)$, but also $\sigma_{\mathrm{obs},I}$ with $\sigma_{\mathrm{mod},I}$, in a way dictated by the geometry of the likelihood distribution. As will be shown in the experiments, the result is that GLS is relatively insensitive to uncertainties in both the stochastic and deterministic components of the regression model. The same quality renders the method also robust against outliers. In the experiments below, we employed a classic active-set algorithm to carry out the optimization. Furthermore, presently the GLS method does not directly offer confidence (or credible) intervals on the estimated quantities. Future work will address this issue in more detail, but for now error estimates were derived by a bootstrap procedure.

From the conceptual point of view, GLS performs regression between points (distributions) on a Riemannian probabilistic manifold, describing the data corresponding to the response variable at each measurement site as a whole through either the observed or the modeled distribution. It is important to stress that this is quite different from treating the data at each measurement site in a pointwise way, i.e. using each individual y_{i_I}. Our method respects the intrinsic nature of the fluctuating quantity described by the variable y. For instance, if, for fixed I, y_{i_I} is a series of samples from a stationary signal, then comparing the measured signal distribution with the predicted distribution can be seen as more natural than comparing each individual sample with its predicted value. Furthermore, in MDE regression usually the data distribution is characterized using a kernel density estimate. Although this offers great flexibility, the disadvantage is that this estimate could be based on data from different measurement sites. In addition, our parametric approach can be an advantage if few measurements are available. Finally, the geometrical view on regression analysis can be illustrated by visualizing the probabilistic manifold [2].

3 Application of GLS to Tully-Fisher scaling

3.1 The baryonic Tully-Fisher relation

The baryonic Tully-Fisher relation (BTFR) between the total (stellar + gaseous) baryonic mass M_b of disk galaxies and their rotational velocity V_f is of fundamental importance in astrophysics and cosmology [11, 12]. It is a remarkably simple and tight empirical relation of the form

$$M_b = \beta_0 V_f^{\beta_1}. \tag{6}$$

Here, M_b is expressed in solar masses $M_\odot$ and V_f in $\mathrm{km\,s^{-1}}$. The BTFR not only serves as one of the tools for determining cosmic distances, but also provides constraints on galaxy formation and evolution models. In addition, it serves as a test for the Lambda cold dark matter paradigm (ΛCDM), particularly in evaluating alterna-

tives such as modified Newtonian dynamics (MOND). Indeed, whereas in ΛCDM the BTFR is a consequence of various complex processes and thus should demonstrate significant intrinsic scatter, MOND predicts a relation with zero intrinsic scatter and a well-defined exponent β_1 with a value of exactly 4.

In this scaling problem, we use data from 47 gas-rich galaxies, as detailed in [12]. The advantage of the gas-rich galaxies is that their masses can be more accurately measured than those of star-dominated galaxies, which are traditionally used to define the Tully-Fisher relation. The rotation velocity V_f is measured in the flat part of the galaxy rotation curve, determined from spectral Doppler shifts. The measurements are plotted in Figure 1(a) on the logarithmic scale and in Figure 1(b) on the original scale.

In this application clearly $n_I = 1$ for all I, so little information can be obtained regarding the distribution of the data from the single measurement at each site. However, the data in [12] also contain estimates of the observational errors, which we treat here as a single standard deviation. This suggests a measurement error on the response variable proportional to M_b, about 38%, i.e. a constant error bar on the logarithmic scale.

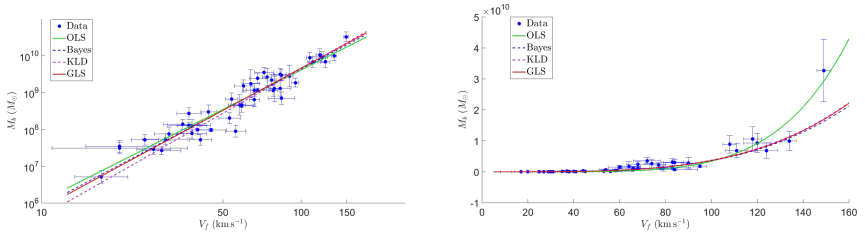
3.2 Regression analysis

Owing to the power law character of most scaling laws, they are often estimated by linear regression on a logarithmic scale. However, it is known that this may lead to unreliable estimates, as the logarithm (heavily) distorts the distribution of the data [2, 13]. This is in particular the case if the estimation is carried out using simple OLS or when there are outliers in the data. In contrast, we will show that GLS regression produces consistent results on both the logarithmic and original scales, demonstrating its robustness.

In view of the proportional error on M_b, the observed standard deviation in GLS is modeled here as a constant $\sigma_{\text{obs},I} \equiv s_{\text{obs}}$ on the logarithmic scale and as $\sigma_{\text{obs},I} = r_{\text{obs}} M_b$ on the original scale. Estimation of these parameters is of interest to get an idea of the intrinsic scatter on the BTFR.

We compare the results of GLS regression with OLS and a Bayesian approach. In the latter, uncertainty on the predictor variables was taken into account into the likelihood. In the case of nonlinear power law regression, the likelihood was approximated by a Gaussian, as the full treatment with marginalization over the model points is too computationally intensive to incorporate in an MCMC simulation [3]. Uncertainty in the specified error bars was modeled through a scale factor with a Jeffreys prior [3]. We also tested the GLS algorithm using the KLD as a similarity measure between the observed and modeled distribution, instead of the Rao GD. We will refer to this algorithm as 'Kullback-Leibler least squares', or KLS.

In order to get a feeling of the uncertainty of the estimates obtained from the optimization routines, 100 bootstrap samples were created from the data, yielding average parameter estimates and their standard deviations on the basis of the results



(a) (b)

Fig. 1 The BTFR data and estimated regression functions by OLS, a robust Bayesian method, KLS and GLS. (a) Logarithmic scale. (b) Original scale.

from OLS, KLS and GLS. Similar estimates were obtained from the MCMC chain in the robust Bayesian approach.

The parameter estimates estimated by the various methods, as well as their standard deviations, are given in Table 1. Figure 1 shows the corresponding regression curves. It is interesting to compare the results obtained by regression on the logarithmic scale, with those derived using nonlinear regression analysis. On the logarithmic scale the data follow a rather clear linear pattern, hence the estimates by the various methods are similar. However, in the nonlinear case the best fit is somewhat less clear at first sight. Although the Bayesian, KLS and GLS methods agree relatively well, the OLS parameter estimates are very different from the linear case. Most noticeably, the nonlinear OLS estimate for the exponent β_1 is heavily influenced by the point with the largest value of V_f and $M_b \approx 3 \times 10^{10} M_\odot$. The other methods are much less attracted by this point because of the large corresponding error bar on M_b. Thus, part of the danger of the logarithmic transformation is due to its influence on the error bars in the presence of model uncertainty. The differences between the parameter estimates by the other methods are much less pronounced, although the consistency appears to be best in the case of GLS. This is in agreement with the good robustness quality of GLS compared to other methods seen in previous analyses [2, 4].

It is also worth pointing out that the scale factor r_{obs} (observed relative error) was estimated by GLS to amount to roughly 63%. This is considerably larger than the value of 38% predicted by the model (and dominated by σ_{M_b}), possibly indicating that the scatter on the scaling law is not due to measurement error alone.

4 Conclusion

We have introduced and motivated geodesic least squares, a versatile and robust regression method based on regression between probability distributions describing fluctuating or otherwise uncertain system properties. Part of the strength of the method is its simplicity, allowing straightforward application by users in various

Table 1 Average regression estimates and their standard deviations for the BTFR obtained with OLS, KLS and GLS from 100 bootstrap samples. Similar results were derived by MCMC sampling with the robust Bayesian method. The units of the parameters have been left out for simplicity. (a) Logarithmic scale. (b) Original scale.

Method	β_0	β_1
OLS	360 ± 220	3.57 ± 0.15
Bayes	220 ± 220	3.72 ± 0.19
KLS	80 ± 80	3.98 ± 0.23
GLS	140 ± 82	3.80 ± 0.16

(a)

Method	β_0	β_1
OLS	$(1.0 \pm 2.3) \times 10^3$	4.94 ± 1.40
Bayes	88 ± 140	3.81 ± 0.20
KLS	120 ± 100	3.91 ± 0.19
GLS	130 ± 130	3.79 ± 0.21

(b)

application fields, without the need for parameter tuning. We have applied GLS to baryonic Tully-Fisher scaling, thereby demonstrating the robustness of the method and providing an alternative means for testing cosmological models based on the estimated intrinsic scatter.

References

1. G. Verdoolaege, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, *AIP Conference Proceedings*, vol. 1636 (Canberra, Australia, 2014)
2. G. Verdoolaege, Entropy **17**(7), 4602 (2015)
3. W. von der Linden, V. Dose, U. von Toussaint, *Bayesian Probability Theory. Applications in the Physical Sciences* (Cambridge University Press, Cambridge, 2014)
4. G. Verdoolaege, A. Shabbir, JET Contributors, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, *AIP Conference Proceedings*, vol. 1853, art. no. 100002 (8 pp.) (Ghent, Belgium, 2017)
5. R. Beran, Ann. Stat. **5**(3), 445 (1977)
6. R. Pak, Stat. Probab. Lett. **26**(3), 263 (1996)
7. M. Werman, D. Keren, IEEE Trans. Pattern Anal. Mach. Intell. **23**(5), 528 (2001)
8. R. Preuss, V. Dose, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, *AIP Conference Proceedings*, vol. 803 (Melville, NY, 2005)
9. C. Rao, in *Differential Geometry in Statistical Inference* (Institute of Mathematical Statistics, Hayward, CA, 1987)
10. J. Burbea, C. Rao, J. Multivariate Anal. **12**(4), 575 (1982)
11. R. Tully, J. Fisher, Astron. Astrophys. **54**(3), 661 (1977)
12. S. McGaugh, Astron. J. **143**(2), 40 (15 pp.) (2012)
13. X. Xiao, et al., Ecology **92**(10), 1887 (2011)