

Geodesic Least Squares Regression for Scaling Studies in Magnetic Confinement Fusion

Geert Verdoolaege

*Department of Applied Physics, Ghent University, Ghent, Belgium
Laboratory for Plasma Physics, Royal Military Academy, Brussels, Belgium*

Abstract. In regression analyses for deriving scaling laws that occur in various scientific disciplines, usually standard regression methods have been applied, of which ordinary least squares (OLS) is the most popular. However, concerns have been raised with respect to several assumptions underlying OLS in its application to scaling laws. We here discuss a new regression method that is robust in the presence of significant uncertainty on both the data and the regression model. The method, which we call *geodesic least squares* regression (GLS), is based on minimization of the Rao geodesic distance on a probabilistic manifold. We demonstrate the superiority of the method using synthetic data and we present an application to the scaling law for the power threshold for the transition to the high confinement regime in magnetic confinement fusion devices.

Keywords: regression, information geometry, scaling laws, nuclear fusion

PACS: 02.50.Cw, 02.40.Ky, 52.55.Dy

INTRODUCTION

Scaling laws are used in various branches of science, such as astronomy, biology and geology, to characterize the underlying mechanisms at work in complex systems. In nuclear fusion experiments based on magnetic confinement of a hot hydrogen plasma, scaling laws are crucial for predicting the performance of future fusion reactors, which will have a larger size, magnetic field, plasma density, etc., compared to present-day experimental devices. These scaling laws can be estimated from data sets spanning a significant part of the parameter space. Ordinary least squares (OLS) regression combined with frequentist theory is the statistical workhorse that is employed for this purpose in the vast majority of cases. However, often there is considerable uncertainty in the experimental data, including predictor variables, and in the model equations (regression model). As such, OLS regression is often unsuitable [1, 2], and many scientific fields could benefit greatly from a unified regression methodology that is flexible and robust and yet is relatively simple to implement. For these reasons we have developed a new regression method, called *geodesic least squares* regression (GLS). We briefly introduce the method in this paper, showing its good performance, and we present an application to an important scaling law in magnetic confinement fusion.

GEODESIC LEAST SQUARES

The GLS regression method starts from the premise that the probability distribution underlying experimental measurements is the fundamental object resulting from the measurement. As such, GLS does not perform regression based on data points in a Euclidean space, but rather operates on probability distributions lying on a probabilistic manifold [3]. This introduces additional flexibility that renders the method robust in the presence of large uncertainties, as will be demonstrated in the experiments.

The principles of GLS regression with a single scalar response (dependent) variable were introduced in [4] and [5]. Briefly, the idea is to consider two different proposals for the distribution of the dependent variable y , conditional on the predictor (independent) variables. On the one hand, there is the distribution that one would expect if all assumptions were correct regarding the deterministic component of the regression model (regression function) and the stochastic component. We call this the *modeled distribution*. On the other hand, we try to capture the conditional distribution of y by not relying on the model assumptions, but directly on the measurements of y . For this we will use the term *observed distribution*, although in the context of the generalized linear model it is also known as the *saturated model*. In another sense GLS is similar to the ideas behind a range of parameter estimation methods that are collectively known in the statistics community as *minimum distance estimation* [6], with the Hellinger distance as a popular similarity measure [7], which was first applied to regression in [8]. The differences are that we use a parametric rather than a nonparametric estimate of the observed distribution, we explicitly model all parameters of the modeled distribution and, finally, we use the *Rao geodesic distance* (GD) as a similarity measure, based on the Fisher information as a Riemannian metric [3, 9].

As a simple example that we will use also in the experiments, consider a linear relation $\eta = b\xi$ between a predictor variable ξ and a response variable η , with b a constant. In accordance with the discussion above, we explicitly wish to allow for the challenging case of uncertainty on the predictor variable ξ . Therefore we assume that, in reality, n samples of a stochastic (noisy) variable x are observed, together with n samples of a stochastic response variable y . We take the simple case of normally distributed (Gaussian) noise, with $\mathcal{N}(\mu, \sigma^2)$ denoting the normal probability distribution with mean μ and standard deviation σ :

$$y = \eta + \varepsilon_y = b\xi + \varepsilon_y, \quad \varepsilon_y \sim \mathcal{N}(0, \sigma_y^2), \quad (1)$$

$$x = \xi + \varepsilon_x, \quad \varepsilon_x \sim \mathcal{N}(0, \sigma_x^2). \quad (2)$$

The observations x_i ($i = 1, \dots, n$) are taken as mutually independent and so are the y_i . σ_x and σ_y are assumed to be known and the same for all measurements (homoscedasticity). According to the regression model, conditionally on x_i each measurement y_i has a normal distribution:

$$p_{\text{mod}}(y|x_i) = \mathcal{N}(bx_i, \sigma_{\text{mod}}^2), \quad \text{where} \quad \sigma_{\text{mod}} \equiv \sigma_y^2 + b^2\sigma_x^2, \quad (3)$$

with the subscript ‘mod’ referring to the modeled distribution. In our simple example, (3) follows from standard Gaussian error propagation rules. However, for nonlinear

regression laws the conditional distribution for y has to be obtained by marginalizing the unknown true values ξ_i . Nevertheless, the Gaussian error propagation laws may be used in the nonlinear case as well, to *approximate* the conditional distribution $p(y|x_i)$ by a normal distribution, as will be shown in the experiments.

The observed distribution corresponding to each data point is next defined as the normal distribution $\mathcal{N}(y_i, \sigma_{\text{obs}}^2)$, with σ_{obs} to be estimated from the data. This extra parameter gives the method added flexibility, since σ_{obs} is not *a priori* required to equal σ_{mod} . As a result, GLS is less sensitive to incorrect model assumptions. Note that in this example we have taken the observed distribution from the same model (Gaussian) as the modeled distribution. Also, σ_{mod} is a fixed value for all measurements and so is σ_{obs} . These assumptions can of course be relaxed, leading to a more general method. However, for ease of computation we will continue working with these simplifications.

GLS now proceeds by minimizing the total GD between, on the one hand, the product of modeled distributions and, on the other hand, the product of observed distributions:

$$\hat{b} = \underset{b \in \mathbb{R}}{\operatorname{argmin}} \operatorname{GD} \left[\prod_{i=1}^n p_{\text{mod}}(y|x_i), \prod_{i=1}^n p_{\text{obs}}(y|y_i) \right].$$

This can be simplified since the squared GD between products of distributions can be written as the sum of squared GDs between the corresponding factors [10]. Hence, the optimization procedure involves matching not only y_i with bx_i , but also σ_{obs}^2 with $\sigma_y^2 + b^2\sigma_x^2$. Note that the parameter b also occurs in the variance of the modeled distribution. In the present work the minimization was performed using a standard sequential quadratic programming method implemented in Matlab.

The manifold of Gaussian distributions amounts to hyperbolic geometry and an analytic expression for the geodesic distance can be calculated. Indeed, for two univariate normal distributions $p_1(x|\mu_1, \sigma_1^2)$ and $p_2(x|\mu_2, \sigma_2^2)$ the GD is given by [10]

$$\operatorname{GD}(p_1, p_2) = \sqrt{2} \ln \frac{1 + \delta}{1 - \delta} = 2\sqrt{2} \tanh^{-1} \delta, \quad \delta \equiv \left[\frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2} \right]^{1/2}.$$

Various models exist to visualize hyperbolic geometry; see e.g. [11]. An intuitive model is the two-dimensional surface called the *pseudosphere* (tractroid), depicted in Figure 1a together with two example geodesics. It is only valid for $\sigma > 1$ and it is periodic in μ . Every point on this surface corresponds to a Gaussian distribution. In Figure 1a two geodesic curves are illustrated between some specific points (distributions) on the manifold. With the standard deviation increasing in the upward direction, it can be understood intuitively how a geodesic can minimize its length by making a ‘detour’ into regions of increased standard deviation and thus stronger curvature. Interestingly, similar arguments enable a deeper insight into the operation of GLS regression, as will be shown below.

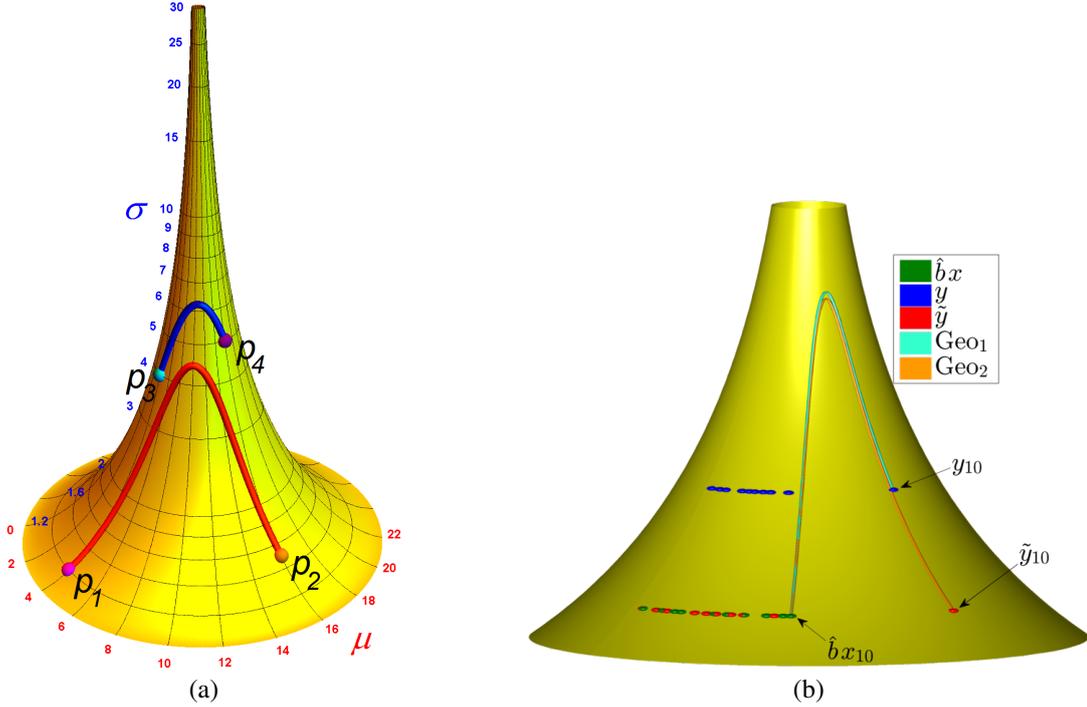


FIGURE 1. (a) The pseudosphere as a model for the univariate normal manifold. Meridians represent lines of constant mean, while circles of latitude have a constant standard deviation. The distributions $p_1(x|4, 1.2^2)$, $p_2(x|16, 1.5^2)$, $p_3(x|4, 4.0^2)$ and $p_4(x|16, 5.0^2)$ are indicated, together with the geodesics between p_1 and p_2 , and p_3 and p_4 . (b) A portion of the pseudosphere together with the regression results on synthetic data with an outlier, as described in the main text.

NUMERICAL SIMULATIONS

Effect of outliers

We first demonstrate the robustness of GLS in the presence of outliers, an advantage that has also been noted in the classic literature of minimum distance estimation [7]. We concentrate on estimating the slope of a regression line with a single independent variable. To this end, a data set was generated consisting of ten points labeled by coordinates ξ_i and η_i ($i = 1, \dots, 10$), with the ξ_i chosen unevenly between 0 and 50 and $\eta_i = 3\xi_i$. Then, Gaussian noise was added to all coordinates according to (1) and (2), with $\sigma_y = 2.0$ and $\sigma_x = 0.5$. Finally, an outlier was created by doubling the value of y_j , with j chosen uniformly among the indices 8, 9 and 10.

We next estimated b by means of GLS and compared the estimates with those obtained by OLS, maximum likelihood estimation (MLE) using the model in Equation (3), total least squares (TLS) [12], which is a typical errors-in-variables technique, and a robust method (ROB) based on iteratively reweighted least squares (bisquare weighting) [13]. In all cases we assumed knowledge of the values of σ_x and σ_y . In order to get an idea of the variability of the estimates, Monte Carlo sampling of the data-generating distributions was performed and the estimation was carried out 100 times.

TABLE 1. Monte Carlo estimates of the mean and standard deviation for the slope parameter in linear regression with errors on both variables and one outlier.

Original	GLS	OLS	MLE	TLS	ROB
$b = 3.00$	3.031 ± 0.035	3.528 ± 0.038	3.696 ± 0.049	4.61 ± 0.11	2.992 ± 0.041

The results are given in Table 1, mentioning the sample average and standard deviation of b over the 100 runs for each of the methods. GLS is seen to perform similar to the robust method. The average σ_{obs} was 5.43 with a standard deviation of 0.24. On the other hand, the modeled value of the standard deviation in the conditional distribution for y_i was $\sqrt{\sigma_y^2 + 9\sigma_x^2} = 2.5$. Hence, GLS succeeds in ignoring the outlier by increasing the estimated variability of the data. As mentioned before, this can be understood in terms of the pseudosphere as a geometrical model for the normal distribution. To see this, we refer to Figure 1b, where several sets of points (distributions) were drawn on a portion of the surface of the pseudosphere for one particular data set generated as described above. First, the modeled distributions were plotted with mean $\hat{b}x_i$ and standard deviation $\sigma_{\text{mod}} = 2.5$, using the average estimate $\hat{b} = 3.03$ obtained by GLS. In this particular data set the index of the outlier was $j = 10$, so the point $\hat{b}x_{10}$ is indicated individually. Then come the observed distributions with mean y_i and standard deviation $\sigma_{\text{obs}} = 5.43 > \sigma_{\text{mod}}$. The outlier in y_{10} can clearly be observed. Finally, the y_i were plotted again (labeled \tilde{y}), but this time at the lower standard deviation σ_{mod} , which would have been expected according to the model. Again, the outlier (labeled \tilde{y}_{10}) can be seen, but now it becomes clear why GLS can compensate for the outlier by increasing its estimate of σ_{obs} w.r.t. σ_{mod} . Indeed, the result is that the geodesic between the points $(\hat{b}x_{10}, \sigma_{\text{mod}})$ and $(y_{10}, \sigma_{\text{obs}})$ (labeled Geo₁) is actually *shorter* than the geodesic between $(\hat{b}x_{10}, \sigma_{\text{mod}})$ and $(y_{10}, \sigma_{\text{mod}})$ (labeled Geo₂): when calculating the GD one finds 2.4 for the former geodesic and 2.8 for the latter.

Effect of logarithmic transformation

We next tested the effect of a logarithmic transformation, which is often used to transform a power-law regression model into a linear form. However, the logarithm alters the data distribution, which may lead to misguided inferences from OLS [1, 2]. Therefore the flexibility offered by GLS is expected to be beneficial in this case, as it allows the observed distribution to deviate from the modeled distribution. To this end, we performed a regression experiment with a power law deterministic model and additive Gaussian noise on all variables. In accordance with the typical situation of fitting fusion scaling laws to multimachine data, the noise standard deviation was taken proportional to the simulated measurements, corresponding to a given set of relative error bars. As a result, in the logarithmic space the distributions were only approximately Gaussian, with the standard deviation given by the constant relative error on the original measurement (homoscedasticity). Ten points were chosen with independent coordinates ξ_i unevenly

TABLE 2. Monte Carlo estimates of the mean and standard deviation for the parameters in a log-linear regression experiment with proportional additive noise on both variables.

Parameter	Original	GLS	OLS	MLE	TLS	ROB
b_0	0.80	0.94 ± 0.47	2.2 ± 2.3	1.75 ± 0.58	0.99 ± 0.70	2.72 ± 0.77
b_1	1.40	1.39 ± 0.11	1.19 ± 0.16	1.21 ± 0.10	1.41 ± 0.14	1.17 ± 0.11

spread between 0 and 60. A power law was proposed to relate the unobserved ξ_i and η_i :

$$\eta_i = b_0 \xi_i^{b_1}, \quad i = 1, \dots, 10.$$

Then, Gaussian noise was added to both coordinates, corresponding to a substantial relative error of 40%. We finally took the natural logarithm of all observed values x_i and y_i , enabling application of the same linear regression methods that were used in the previous experiment. In this particular experiment we chose $b_0 = 0.8$ and $b_1 = 1.4$, but we found that other values yield similar conclusions. Again, 100 data replications were generated, allowing calculation of Monte Carlo averages.

The averages and standard deviations over all 100 runs are given in Table 2. Again, the results show that GLS is robust against the flawed model assumptions, performing similar to TLS.

POWER THRESHOLD SCALING

One of the most important scaling relations in fusion science concerns the threshold P_{thr} for the heating power that is required for the plasma to make the transition to a desired regime of high energy confinement (H-mode) in the next-step fusion device ITER [14]. To a good approximation this power threshold depends on the electron density in the plasma \bar{n}_e (in 10^{20} m^{-3}), the main magnetic field B_t (in T) and the total surface area S of the confined plasma (in m^2). This is usually expressed by means of the following scaling relation:

$$P_{\text{thr}} = b_0 \bar{n}_e^{b_1} B_t^{b_2} S^{b_3}. \quad (4)$$

To estimate the coefficients in this relation, we employed data from seven fusion devices, yielding 645 measurements of power, density, magnetic field and surface area (subset IAEA02 [14]).

Linear scaling

We first followed the standard practice of transforming to the logarithmic scale to estimate the coefficients b_i ($i = 0, \dots, 3$) via linear regression. In the GLS method we introduced additional parameters, approximately describing the relative errors for the power threshold (one for each device), similar to the parameter σ_{obs} in the example above. The estimation results using OLS and GLS are shown in Table 3, with the relative errors on P_{thr} varying between 21% and 48%, as estimated by GLS through the σ_{obs} parameters. The estimates of the coefficients by GLS are rather different from those

TABLE 3. Estimates of regression coefficients b_i and predictions for ITER (with 1σ and 95% confidence intervals (CI) in case of OLS) in log-linear scaling for the H-mode threshold power.

Method	b_0	b_1	b_2	b_3	$\hat{P}_{\text{thr},0.5}$ (MW)	CI 1σ	CI 95%	$\hat{P}_{\text{thr},1.0}$ (MW)	CI 1σ	CI 95%
OLS	0.059	0.73	0.71	0.92	48	+3.7 -3.5	+7.6 -6.6	80	+7.4 -6.8	+15 -12
GLS	0.065	0.93	0.64	1.02	62	-	-	117	-	-

obtained by OLS, particularly in the density dependence. The predictions for ITER are also shown, for two typical densities (0.5 and $1.0 \times 10^{20} \text{ m}^{-3}$). Whereas confidence intervals for the predictions can be readily obtained via OLS in the standard way, in the case of GLS confidence intervals should be estimated from the residuals using suitable approximations or by means of bootstrapping methods. This will be addressed in future developments, but note from Table 3 that the predictions by GLS do fall out of the 1σ and 95% confidence intervals provided by OLS.

Nonlinear scaling

Finally, we show the results of nonlinear regression in the original data space, i.e. without logarithmic transformation. Whereas this prevents an analytic solution using OLS, the advantage is that the distribution of the data is left undistorted [1, 2], while the implementation of both OLS and GLS is not significantly more complex. Indeed, the distribution of the right-hand side in (4) can be approximated by a Gaussian with mean $\mu_{\text{mod}} = b_0 \bar{n}_e^{b_1} B_t^{b_2} S^{b_3}$ and standard deviation σ_{mod} , given by

$$\sigma_{\text{mod}}^2 = \sigma_{P_{\text{thr}}}^2 + \mu_{\text{mod}}^2 \left[b_1^2 \left(\frac{\sigma_{\bar{n}_e}}{\bar{n}_e} \right)^2 + b_2^2 \left(\frac{\sigma_{B_t}}{B_t} \right)^2 + b_3^2 \left(\frac{\sigma_S}{S} \right)^2 \right].$$

Hence, the error bars depend on the measurements (heteroscedasticity). Nevertheless, we introduced an approximation assuming constant error bars for all measurements from a single machine. This assumption may be relaxed in the future. The results of the scaling and predictions are given in Table 4. It is interesting to note that, in contrast with OLS, the results for GLS are similar to those derived on the logarithmic scale (Table 3), indicating that, indeed, GLS is less susceptible to flawed model assumptions. Furthermore, the results for OLS and GLS are now in the same range, with slightly lower predictions by OLS. Nevertheless, both methods suggest higher power thresholds than those obtained in earlier studies in the same database ($\hat{P}_{\text{thr},0.5} = 44 \text{ MW}$ [14]).

CONCLUSION

Regression and scaling laws represent crucial tools in science in general and in the analysis of complex physical systems in particular. We have presented geodesic least squares

TABLE 4. Fits and predictions for nonlinear scaling.

Method	b_0	b_1	b_2	b_3	$\hat{P}_{\text{thr},0.5}$ (MW)	CI 1σ	CI 95%	$\hat{P}_{\text{thr},1.0}$ (MW)	CI 1σ	CI 95%
OLS	0.051	0.85	0.70	1.00	62	± 5	± 10	111	± 12	± 23
GLS	0.048	0.96	0.59	1.05	64	–	–	124	–	–

regression as a method that is able to handle large uncertainties on the deterministic and stochastic components of the regression model. Operating on a manifold of probability distributions, the results can be easily visualized in the case of the univariate Gaussian distribution. However, GLS is sufficiently general to allow tackling much more general regression problems within the same framework. We have shown the robustness of the method in some specific examples on synthetic data. Next, we have addressed the scaling of the power threshold in magnetically confined fusion plasmas, yielding significantly higher estimates of the threshold for ITER.

In future work we intend to develop a Bayesian estimation method for the regression parameters, by describing the distribution of the data probability distributions directly on the probabilistic manifold. This will allow to provide reliable uncertainty estimates on the parameters.

REFERENCES

1. D. McDonald, et al., *Plasma Phys. Control. Fusion* **48**, A439–A447 (2006).
2. X. Xiao, et al., *Ecology* **92**, 1887–1894 (2011).
3. S. Amari, and H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society, New York, 2000.
4. G. Verdoolaege, “Geodesic Least Squares Regression on Information Manifolds,” in *33rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Canberra, Australia, 2013.
5. G. Verdoolaege, *Rev. Sci. Instrum.* **85**, 11E810 (2014).
6. J. Wolfowitz, *Ann. Math. Statist.* **28**, 75–88 (1957).
7. R. Beran, *Ann. Stat.* **5**, 445–463 (1977).
8. R. Pak, *Stat. Probab. Lett.* **26**, 263–269 (1996).
9. C. Rao, *Differential Geometry in Statistical Inference*, Institute of Mathematical Statistics, Hayward, CA, 1987, chap. Differential metrics in probability spaces.
10. J. Burbea, and C. Rao, *J. Multivariate Anal.* **12**, 575–596 (1982).
11. F. Nielsen, and R. Nock, “Visualizing hyperbolic Voronoi diagrams,” in *Proceedings of the 30th Annual Symposium on Computational Geometry (SOCG’14)*, Kyoto, Japan, 2014, p. 90, URL <https://www.youtube.com/watch?v=i9IUzNxeH4o>.
12. I. Markovskiy, and S. Van Huffel, *Signal Process.* **87**, 2283–2302 (2007).
13. R. Maronna, D. Martin, and V. Yohai, *Robust Statistics: Theory and Methods*, Wiley, New York, 2006.
14. J. Snipes, et al., “Multi-Machine Global Confinement and H-mode Threshold Analysis,” in *Proceedings of the 19th IAEA Fusion Energy Conference*, CT/P-04, Lyon, France, 2002.