Spatio-temporal Analysis of Extreme Values from Lichenometric Studies and their Relationships to Climate

Daniel Cooley
Department of Applied Mathematics
University of Colorado at Boulder, USA
cooleyd@colorado.edu

Vincent Jomelli
Laboratoire de Geographie Physique
CNRS, Meudon-Bellevue, France

Philippe Naveau
Laboratiore des Sciences du Climat et de l'Environnement
IPSL-CNRS, Gif-sur-Yvette, France
Department of Applied Mathematics
University of Colorado at Boulder, USA

Keywords: lichenometry, extremes, climate, Generalized Extreme Value distribution, EM algorithm

Abstract

Arctic and alpine regions are very important to understand the effects of climate change and other geophysical phenomena. The lack of relevant time series in such environments gave rise to lichenometry, the study of lichen growth for the purpose of dating rock features such as glacial moraines. Although lichenometry has been practiced for years, it has lacked a solid statistical basis. The statistical challenge is to propose a spatio-temporal model for extreme lichen diameters. As a first step, we develop a temporal bivariate model (lichen sizes and their associated dates) based on extreme value theory. The flexibility of our statistical model allows us to integrate the error associated with the dating process (i.e. estimating the age of each moraine). To validate our statistical methodology, simulated examples were analyzed and tested. The proposed techniques are applied to data from Bolivia. Finally, we propose a few directions to extend our model into a spatial framework.

1. Introduction

The climate of the Little Ice Age is well studied in Europe through historical and proxy records. However, in areas of the globe where fewer climatological records are available, it is not well understood to what extent (if any) the Little Ice Age occurred. With no historical record of climate in these areas, we must look to indirect sources to gain an understanding of the climate change. One source of information comes from glaciers, which may be sensitive to climate change. By determining the ages of the moraines formed by glaciers, we can gain an understanding of when periods of relatively cooler climate occurred.

Lichenometry is a method for dating glacial landforms. It is well suited for alpine environments where the sparse vegetation makes dating techniques like dendrochronolgy impossible, and is best at dating features from recent centuries where Carbon-14 techniques have low precision. The basic

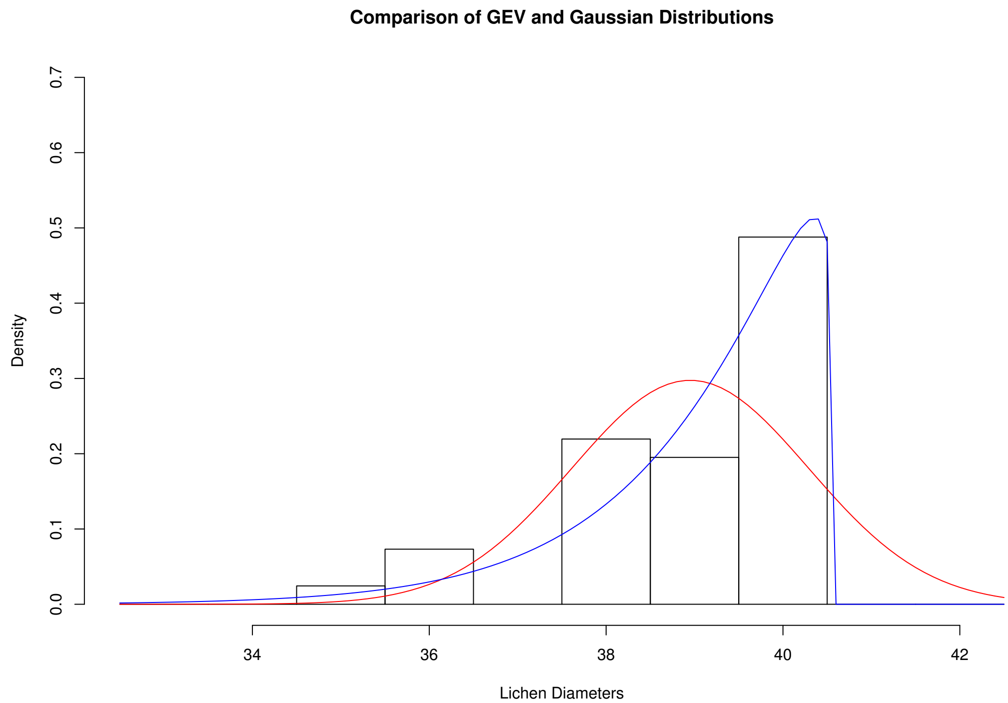

Figure 1: *An example of distribution for maximum lichen diameters. The black line corresponds to the histogram of the data. The red line shows the fit by a Gaussian density distribution. The blue line indicates the fit by a Generalized Extreme Value (GEV) density distribution.*

premise of lichenometry is that the size of the largest lichens on the landform are proportional to the age of the landform. Typically, lichen measurements from undated land features are compared to a known growth curve of lichen measurements verses age of feature. The growth curves are usually constructed from lichen measurements on features with known dates (man-made structures or features dated historically or by other methods) (Innes J. L., 1985). Growth curves are unique to each locality as lichen growth rates are effected by climate (Benedict J.B., 1967, 1990, 1991), lithology (Rodbell D.T. , 1992), and exposition (Pentecost A., 1979).

Lichenometry has suffered from a lack of a statistical model for the ages of the glacial features and the lichen measurements obtained from these features. To our knowledge, no comprehensive statistical model has been proposed for use in lichenometry. Any model should be based on the specificity of the data of lichenometry. Although there is disagreement on the best field procedure (Locke et al., 1979, Mathews 1974; 1975; 1977, Innes 1983; 1985) , all of the methods analyze only the largest lichen measurements for use in lichenometric dating. Since the raw material of lichenometry is the measure of the largest lichens, it seems only natural to apply extreme value theory to the measurements. However, the statistics of extremes has been all but absent in lichenometry. Another statistical weakness in past lichenometry studies is that an account of uncertainty has rarely been included. A few past studies (O'Neal M. and Schoenenberger K. R., 2003) associated confidence intervals and/or p-values to their specific growth curves. But the hypothesis to derive these confidence intervals was formulated as: *maximum diameter* $=$ *f(age)* $+$ *noise* where f is the growth curve and the noise Gaussian. We will illustrate that this last assumption is not satisfied (See Figure 1).

We propose a bivariate statistical model for the lichen measurements and the dates of the associated features. The model is based on extreme value theory, and allows one to compute small confidence intervals for the inferred moraine dates. In addition, it offers three other advantages: (1) a global statistical model, all the data (from both dated features and the undated moraines of interest) are analyzed with a one-step procedure, (2) a theoretical framework, the maximum lichen distribution is derived from a statistical theory, (3) a flexibility, different types of growing curves can be investigated and their fits compared.

The probabilistic and statistical study of extremes is well-formulated. There are many excellent books on the subject including Embrechts et al. 1998 which gives a comprehensive mathematical background of the theory, and Coles 2001 which focuses on application and data analysis. Since the pioneering work of Gumbel 1958, extreme value theory has been applied in numerous fields. Much of the early work was done in hydrology, from which we also get the canonical example of an extreme value statistic, the so-called "100-year flood". Other areas of application have included insurance and finance, and more recently climate. Given any data set, if one wants to describe the behavior of the largest values of the data, one should resort to the methods based in the theory of extremes.

Like the familiar Central Limit Theorem (CLT), extreme value theory arises from an asymptotic result. However, instead of studying the mean, we wish to study the maximum of an iid sequence of random variables, $M_n = \max(X_1, X_2, ..., X_n)$. The primary theorem of extreme value theory describes the distribution of M_n as $n \to \infty$. If there exist sequences a_n and b_n such that the renormalized maximum $(M_n - b_n)/a_n$ has a non-degenerative distribution, then as $n \to \infty$ this distribution converges to the Generalized Extreme Value (GEV) distribution. The GEV distribution has three parameters, μ, $\sigma > 0$, and ξ, and its cumulative distribution function is given by

$$G(x; \mu, \sigma, \xi) = \left\{ \begin{array}{ll} \exp\left\{ - \left[1 + \xi \frac{x-\mu}{\sigma} \right]_+^{-1/\xi} \right\}, & \text{when } \xi \neq 0 \text{ and } a_+ = \max(0, a), \\ \exp\left\{ - \exp\left(- \frac{x-\mu}{\sigma} \right) \right\}, & \text{when } \xi = 0, \end{array} \right\}. \quad (1)$$

The shape parameter, ξ, describes the tail behavior of the distribution of the original X_i. If ξ is negative, the tail is bounded; if ξ is zero the tail is "light" and decreases exponentially; and if ξ is positive the tail is "heavy" and the distribution of X_i has a finite number of moments. Although the GEV distribution is remarkably flexible in its ability to account for any type of tail behavior, it is important to recognize that the distribution of the maximum does not resemble the familiar Gaussian distribution of the CLT. (Figure 1)

The remainder of this article is structured as follows. We discuss our work on lichenometry for a single glacier in section 2. Included in the discussion is a description of our model, results from tests of our model with simulated data, and results from the actual data gathered in Bolivia. For a more detailed description of our work on a single glacier we refer to Naveau, et al. 2004. Section 3 is a brief discussion of a planned model which could incorporate spatial effects.

2. Lichenometry for a Single Glacier

2.1. Model and Method

There are two sources of uncertainty in lichenometry: that which comes from modeling the size of lichen diameters given the age of the feature, and that which comes from associating a date with a feature. Because of these two sources of uncertainty, we choose to model a bivariate random vector, $(\mathbf{X_i}, T_i)$, where T_i is the date associated with the feature, and $\mathbf{X_i} = (X_{i,1}, X_{i,2}, ...X_{i,n_i})$ are the maximums of the lichen diameters from the n_i blocks on moraine i. Because the $X_{i,j}$ are block maximums, we model them with a GEV distribution (See Equation (1)), with parameters μ_i, σ_i, and xi_i dependent on the moraine i. The variable T_i is assumed to be a Gaussian random variable

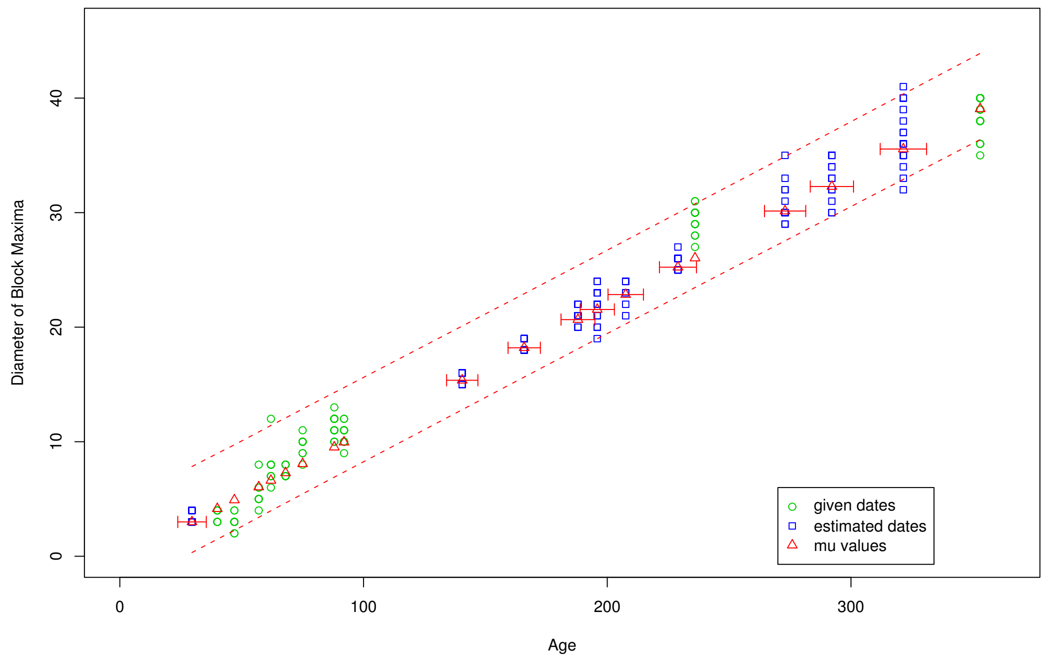

Figure 2: *A new type of growth curve plot. This shows each feature's GEV parameter μ vs the feature's age (given or estimated by the procedure). Also plotted are the maximum diameters from each block for each feature (blue and green), the 95% confidence intervals for the maximum diameter values (red dashed lines), and the 95% confidence intervals for the ages of the undated moraines (red bars).*

with mean α_i, and variance β_i. To link the two random variables, we assume that the diameters are independent in distribution from the age of the feature, but that the Gaussian parameter α_i, is a function of the GEV parameters μ_i. This yields the joint density function:

$$\begin{aligned} f(x_{i,j}, t_i) &= \exp\left\{-\left[1+\xi\left(\frac{x_{i,j}-\mu_i}{\sigma_i}\right)\right]^{-\frac{1}{\xi}}\right\} \times \left\{\frac{1}{\sigma_i}\left[1+\xi_i\left(\frac{x_{i,j}-\mu_i}{\sigma_i}\right)\right]^{-\frac{1}{\xi_i}-1}\right\} \\ &\quad \times \frac{1}{\sqrt{2\pi\beta^2}} \exp\left[-\frac{1}{2\beta^2}(t_i-\alpha)^2\right] \text{ where } \alpha = g(\mu) \text{ and } \beta^2 = h(\mu) \end{aligned}$$

Using this model, the notion of a "growth curve" takes on a new meaning (Figure 2). Rather than being a simple plot between the age of the feature and the lichen diameters directly, the growth curve is now a relationship between the GEV parameter μ and the mean of the age distribution α. This model is also quite flexible, allowing one the freedom to choose the type of growth curve simply by defining the function $\alpha_i = f(\mu_i)$. For the Bolivian data, a simple linear relationship appeared to be adequate, so we let $\alpha_i = a + b\mu_i$. We further reduced the number of parameters by assuming that $\xi_i = \xi$ since we assume that the shape parameter (which describes tail behavior of lichen diameters) does not change in time.

Given the joint density (See Equation (2)) and the random vectors $(\mathbf{X_i}, T_i)$, we could estimate the parameters μ_i, σ, ξ, a, and b using numerical maximum likelihood methods. However, the variable T_i is only observed for the features which have already been dated, and this random variable must be estimated for the moraines which we have no dates. To deal with the missing dates, we resort to the EM algorithm. The recursive algorithm uses the current parameter estimates to find the expected values of the missing dates, and then obtains the best parameter estimates given those dates via maximum likelihood (Dempster, et al. 1977, McLachlan and Thriyambakam 1997). The algorithm is allowed to run until the likelihood converges, yielding both point estimates for the parameters and the ages of the undated moraines. Confidence intervals for the parameters can be obtained using the delta method (Coles, S., 2001). To obtain confidence intervals for the ages, we use the conditional variance formula

$$\mathrm{var}(\hat{\alpha}_i) = \mathrm{var}(\mathrm{E}(\hat{\alpha}_i|\hat{\mu}_i)) + \mathrm{E}(\mathrm{var}(\hat{\alpha}_i)|\hat{\mu}_i)). \tag{2}$$

For a complete formulation of the confidence intervals, see Naveau et al. (2004).

2.2. Results from both Simulated Data and Bolivian Glacier Data

To test our method, many data simulations were run. The simulation whose results are shown below was for twelve dated features with a varying number of blocks. Dates for four of the features were then deleted and the method described above was used to estimate parameter values and dates for the undated features.

	Parameters			
	$\log(\sigma)$	ξ	a	b
Simulated Value	0.5	-0.2	-20	10
Point Estimates	0.574	-0.223	-22.556	10.087
0.95 Conf Intvl	(0.491,0.657)	(-0.279,-0.168)	(-27.4, -17.7)	(9.84, 10.33)
	Moraine Ages			
	Moraine 9	Moraine 10	Moraine 11	Moraine 12
Simulated Age	100	30	230	90
Point Estimates	96.0	23.3	235.8	90.5
0.95 Conf Intvl	(85.7,106.4)	(13.5,33.2)	(221.8,249.9)	(80.3,100.8)

The method was then applied to the field data gathered in Bolivia. Lichen measurements from ten dated features were combined with measurements from ten undated moraines of the Charquini glacier.

	Parameters			
	$\log(\sigma)$	ξ	a	b
Point Estimates	0.368	-0.184	2.69	8.97
0.95 Conf Intvl	(0.29, 0.44)	(-0.23, -0.13)	(-0.17, 5.56)	(8.84,9.09)

	Moraine Ages				
	M1	M2	M3	M4	M5
Point Estimates	321.5	292.1	272.9	229.0	207.6
0.95 Conf Intvl	(315.9, 327.1)	(288.5, 295.8)	(270.1, 275.78)	(227.0, 231.0)	(205.6, 209.5)
	M6	M7	M8	M9	M10
Point Estimates	195.8	188.0	165.9	140.5	29.6
0.95 Conf Intvl	(193.9, 197.7)	(186.0, 189.9)	(163.9, 167.9)	(138.8, 142.2,)	(29.2, 29.9)

Certainly there is more investigation that can be done in the single glacier case. Functions different from our linear one relating the GEV parameters and dates could be proposed and tested using likelihood values and quantile plots. These new functions could also incorporate the GEV scale parameter σ. Relationships could also be given between the GEV parameters and the variable β in the Gaussian distribution of the dates. All of the above tests would be relatively easy given the flexibility of the above model.

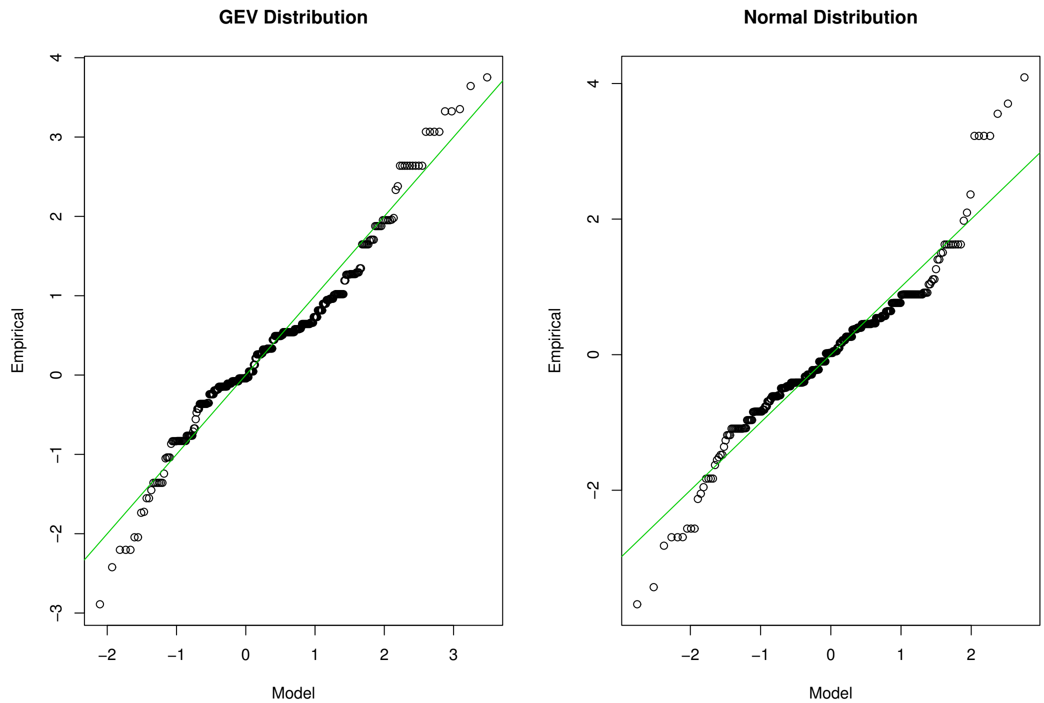

Figure 3: *A comparison of the quantile plots of the model based on the GEV distribution to a model based on the normal distribution.*

3. Future Directions to Add the Spatial Component to Lichenometry

Having a working method for a single glacier, we have now turned our attention toward trying to integrate spatial variability into our model. We have data from other nearby and not-so-nearby glaciers, and would like to integrate all of the data into a global model.

One possibility for a spatial model for extremes was proposed by Coles and Casson 1999. They propose a model which the parameters of an extreme value distribution are dependent on the location from which the data come. The spatial structure is related through a regression model with parameters of its own. To apply their ideas with our model, we may have to abandon our maximum likelihood approach in favor of a Markov Chain Monte Carlo method. This would change the nature of the unknown dates from what we proposed in single glacier model above. Rather than being unseen random variables, the dates would have to be treated as parameters. Of course, all of the parameters are treated as random variables in the MCMC method, so this is not necessarily a drawback. And although we foresee that the MCMC method would be computationally intensive, the recursive maximum likelihood method described in the single glacier case was computationally intensive as well.

References

Benedict J.B. (1967). Recent glacial history of an alpine area in the colorado front range, usa. i establishing a lichen growth curve. *Journal of Glaciology*, 6:817–832.

Benedict J.B. (1990). Lichen mortality due to late-lying snow, results of a transplant study. *Arctic and Alpine Research*, 22:81–99.

Benedict J.B. (1991). Experiments on lichen growth ii. effects of a seasonal snow cover. *Arctic and Alpine Research*, 23:189–199.

Casson, E. and Coles, S. (1999). Spatial regression models for extremes. *Extremes*, 1:449–468.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer.

Dempster A.P., Laird N.M., Rubin D.B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–21.

Embrechts, P. Kluppelberg, C., Mikosch T. (1998). *Modelling Extremal Events for Insurance and Finance.* Springer Verlag.

Gumbel, E.J. (1958). *Statistics of Extremes.* Columbia University Press.

Innes J. L. (1983). Size frequency distributions as a lichenometric technique, an assessment. *Arctic and Alpine Research*, 15, 3:285–294.

Innes J. L. (1985). Lichenometry. *Progress in Physical Geography*, 9:187 – 254.

Locke W. W., Andrews J. T., and Webber P. J. (1979). A manual for lichenometry. *BGRG Techn. Bull.*, 26:48pp.

Mathews J.A. (1974). Families of lichenometric dating curves from the storbreen gletschervorfeld, jotunheimen, norway. *Norsk geografisk Tidskrift*, 28:215–235.

Mathews J.A. (1975). Experiments on the reproducibility and reliability on lichenometric dates, storbreen gletshervorfeld, jotunheimen, norway. *Norsk geografisk Tidskrift*, 31:129–136.

Mathews J.A. (1977). A lichenometric test of the 1750 end-moraine hypothesis, storbreen gletshervorfeld, jotunheimen, norway. *Norsk geografisk Tidskrift*, 31:129–136.

McLachlan G.J., Thriyambakam, K. (1997). *The EM Algorithm and Extensions.* Wiley.

Naveau, P., Jomelli, V., Cooley D., Rabatel, A. (2004). Modeling uncertainties in lichenometry studies with an application: the tropical andes. *Submitted for Publication.*

O'Neal M. and Schoenenberger K. R. (2003). A rhizocarpon geographicum growth curve for the cascade range of washington and northern oregon, usa. *Quaternary Research*, pages 233–241.

Pentecost A. (1979). Aspect and slope preferences in a saxicolous lichen community. *Lichenologist*, 11:81–83.

Rodbell D.T. (1992). Lichenometric and radiocarbon dating of holocene glaciation, cordillera blanca, peru. *The Holocene*, 2:19–29.