Deconvolution of Mixture’s Components Inside Particle Size Distributions

Lorenzo Borselli*

*Instituto de Geología/fac. Ingenieria
Universidad Autónoma de San Luis Potosí
Av. M. Nava No 5, Zona Universitaria,
San Luis Potosí, 78240, Mexico

lborselli@gmail.com
http://www.lorenzo-borselli.eu

DICIM- UASLP

San Luis Potosi; Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosi
How to describe quantitatively particles size distribution in natural deposits/outcrop of sediments, soil, rock mass, or geomaterials?

Cumulative Distribution function or (CDF) or grain size cumulative distribution curve

How much is the information content inside CDFs and PDFs of Particle size distribution (PSD)?

Frequency distribution and probability density (PDF)

San Luis Potosi, Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosi
Useful formulas:

\[ \psi = 3.3219 \log_{10} X \]

\[ X = 10^{-0.301\phi} \]

\[ X = 10^{0.301\psi} \]
Experimentally derived PSD often shown multimodal shape and this characteristic is usually interpreted as a mixture of two or more populations. In geosciences the origin of these mixtures has been commonly interpreted as the result of complex processes: the different origin of the sediment ad clasts; the transport and their final deposition. The geological cycle of sediment transport, the weathering or edaphological processes may affect the final PSD in the deposit.

The basic idea is that each type of process left a fingerprint inside the granulometric distribution and in the presence of different components. In other words the whole properties of the mixture are coded in the properties of experimental PSD.

Sun et al. 2002, sedimentary geology. (mixing in eolian an hydraulic env.)

Battaglia et al. (2003), Geomorphology, Clay distribution in different erosional env.
Each PSD may be interpreted as a mixture of different phase or separate components. The knowledge of these components is important in order to understand the processes (or their combination) which produced them. Application of this concept is important in geology, engineering geology and material science.

**Tephra-Fall Deposits** from the Tefra deposits une 15, 1991, Eruption of Mount Pinatubo (*Koyaguchi, 1999*)


http://ark.cdlib.org/ark:/13030/ft6v19p151/
Total PSD in samples of Debris flow deposits at diferents distance from the source, with example of polymodality (Volcan de Fuego, Colima, MX) (Capra et al., 2010, JVGR)
DECOLOG’s OBJECTIVES

• Problem solving for separating and decoding components inside an experimental PSD.

• Use as paradigm a mixture of Lognormal and Weibull distributions.

• Automatic process by innovative non-linear multi-objectives optimization techniques.

• It does not require (as other software does) the manual preliminary selection (by User) of the peaks of each component.

WWW.DECOLOG.ORG

DECOLOG 5.4 is fully free (freeware) for scientific community (Windows 7/8.x/10 64bit)

San Luis Potosi; Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosi
DECOLOG 5.4, Win 64 bit
(last update 09/04/2016)

CONACYT:
Proyecto Ciencia Basica CB-2012/184060

What is DECOLOG
Experimentally derived particle size distribution often shown multimodal shape and this characteristic is usually interpreted as a mixture of two or more populations. The origin of these mixture has been commonly interpreted as due to the complex processes linked to the origin of the sediment clasts, to the transport and final deposition, or in other terms, the geological cycle of sediment transport and evolution, the weathering and soil process may affect the final distribution of particles present in the sampled deposit.

The basic idea is that the all the processes responsible of the deposit leave some trace of them inside the special characteristics of the mixture and their populations. We assume that the mixture mantains encoded in its the global distribution informations.

Aim of DECOLOG software is develop a solution to decode the information present in the natural mixture of particles sediments using, as paradigm, the log-normal or weibull distributions and particularly a defined mixture of these distributions.

Decolog performs this operation using innovative techniques of optimization and in automatic way, without the needs of special efforts from user, as the initial guessing of Peaks of the observed distribution ...

The easiness of use is one of the most innovative and appreciated characteristics of current version of DECOLOG.

Version 5.4
Para descargar

Para descargar

San Luis Potosi; Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosi
Graphical User Interface  DECOLOG 5.4 (2016)
Reporte Grafico en formato generado en DECOLOG 5.4

San Luis Potosi; Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosi
RESULTS

- Each distribution can be analyzed globally or by components.
- Statistical parameters of each identified sub-population.
- Global statistics by Monte Carlo “resampling” or Folk-Ward optimized.
- Analysis’s report in text or graphical formats: CSV(XLS), PDF.

---

**COMPONENTS’ OPTIMUM FITTING PARAMETERS**

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Shift (λ)</td>
<td>-8.6756</td>
<td>-3.3761</td>
<td>5.6786</td>
</tr>
<tr>
<td>Scale (α)</td>
<td>3.3375</td>
<td>3.205</td>
<td>6.9788</td>
</tr>
<tr>
<td>Shape (β)</td>
<td>2.1241</td>
<td>1.9072</td>
<td>3.4806</td>
</tr>
<tr>
<td>Fraction</td>
<td>0.4806</td>
<td>0.3444</td>
<td>0.175</td>
</tr>
</tbody>
</table>

Total minimised Multi-Objective function value: 743.1047119

---

**STATISTICAL PARAMETERS OF THE DISTRIBUTIONS AND MIXTURE**

<table>
<thead>
<tr>
<th>STATISTICS</th>
<th>1st weibull</th>
<th>2nd weibull</th>
<th>3rd weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-5.7197</td>
<td>-6.2198</td>
<td>-0.5987</td>
</tr>
<tr>
<td>Mode</td>
<td>-6.202</td>
<td>-5.547</td>
<td>-0.6531</td>
</tr>
<tr>
<td>Median</td>
<td>-5.867</td>
<td>-6.0208</td>
<td>-0.6027</td>
</tr>
<tr>
<td>Variance</td>
<td>2.1424</td>
<td>2.407</td>
<td>3.9858</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.4637</td>
<td>1.5515</td>
<td>1.9965</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.5526</td>
<td>-0.6959</td>
<td>-0.03</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.1084</td>
<td>3.3732</td>
<td>2.7122</td>
</tr>
<tr>
<td>Kurtosis (normalized)</td>
<td>0.1084</td>
<td>0.3732</td>
<td>-0.2878</td>
</tr>
</tbody>
</table>

GLOBAL STATISTICS FROM WEIGHTED COMPONENTS OF THE DERIVED MIXTURE

| Mean       | -4.9957     |
| Standard deviation | 2.59  | Very poorly sorted |
| Skewness   | 1.0352     | Positive skewed   |
| Kurtosis   | 3.9755     |
| Kurtosis (normalized) | 0.9755 | Leptokurtic |
METHODS

DECOLOG 5.4 Contains a new optimization engines that allow to consider components of Lognormal and Weibull distribution with negative skewness (left tailed distribution), a set of generalized 4 parameters distributions.

DECOLOG’s internal engine was improved a lot considering a set of multi-objective optimization genetic algorithms: “Differential evolution” (DE) y “Trigonometric differential evolution (TDE). These algorithm produced a relevant increase of convergence to global minima, significance and reproducibility of final results.


(Invited seminar Centro De Geociencias - UNAM, Queretaro (MEXICO) – May 7th 2008)

See at Presentation section, in the web site: www.lorenzo-borselli.eu
Continue distributions of 4, 3 and 2 Parameters

PDF (probability density function) 2 parameters with tail

PDF (probability density function) 3 parameters with tail, with a shift of the origin

PDF (probability density function) 4 parameters, with shift of the origin and negative skewness

Simmetrical but with negative skewed

Simmetrical but with negative skewed and different shift value
Details on Lognormal Dist. 4,3 and 2 Parameters

PDF (probability density function)

\[ f(x) = \frac{e^{-\frac{\left(\ln(k(x-\lambda))-\alpha\right)^2}{2\beta^2}}}{\sqrt{2\pi}\beta\left(k(x-\lambda)\right)} \]

CDF (Cumulative distribution function)

\[ F(x) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{k\left(\ln(k(x-\lambda))-\alpha\right)}{\beta\sqrt{2}}\right)\right) \]

Note:

with \( K=+1 \) we have a standard Lognormal 3 parameters, and with \( \lambda = 0 \) it became a standard _Lognormal_ 2 Parameters

Notes: The value \( k= +1 \) produces a classical right tailed distribution – (positive skewness), the value \( k= -1 \) produces a left tailed distribution – (negative skewness) by Aitchison and Brown (1957) (reflected distribution).\( K \) assumes only integer values.

\( \lambda \) is : the location/shift parameters related to the shifting on x axis with respect the origin of the axis. Valid values are in the interval \([-\infty, +\infty]\).

\( \alpha \) the scale parameter v. valid values are in the interval \([0, +\infty]\).

\( \beta \) is the shape parameter. Valid values are in th interval \([0, +\infty]\).
Details on Weibull Dist. 4,3 and 2 Parameters

PDF (probability density function)

\[ f(x) = \frac{\beta}{\alpha} \left( \frac{k(x - \lambda)}{\alpha} \right)^{\beta-1} e^{-\left(\frac{k(x - \lambda)}{\alpha}\right)^\beta} \]

CDF (Cumulative distribution function)

\[ F(x) = \frac{1 + k}{2k} - k e^{-\left(\frac{k(x - \lambda)}{\alpha}\right)^\beta} \]

Note: with \( K=+1 \) we have a standard weibull 3 parameters, y con \( \lambda = 0 \) it became a standard weibull 2 parámetros

Notes: The value \( k=+1 \) produces a classical weibull distribution, the value \( k=-1 \) produces ... reflected Weibull distribution Cohen(1973) – \( K \) assumes only integer values. \([-\infty, +\infty]\)

\( \lambda \) is the location/shift parameters related to the shifting on x axis with respect the origin of the axis. Valid values are in the interval \([-\infty, +\infty]\)

\( \alpha \) the scale parameter. Valid values are in the interval \([0, +\infty]\).

\( \beta \) is the shape parameter. Valid values are in the interval \([0, +\infty]\).
Deconvolution by **not linear multi-objective global optimization**

A Procedure for Global Optimization, not linear, multi-objectives has been developed (since 2004) in order to obtain a robust and efficient decoding of information inside PSD. We obtain 5 parameters identifying the fingerprint of each distribution and their relative importance in total PSD: \( \alpha_i, \beta_i, \lambda_i, k_i, w_i \)

\[
\begin{align*}
  f(x)_{\text{mix}} &= w_1 f_1(x) + w_2 f_2(x) + \ldots + w_n f_n(x) \\
  F(x)_{\text{mix}} &= w_1 F_1(x) + w_2 F_2(x) + \ldots + w_n F_n(x)
\end{align*}
\]

Mixtures of PDFs and CDFs

The multi-objective optimization we produce a concurrent fitting of observed PDF and CDF, or in other words, on PDF function and on its integral (CDF). This is more efficient than on the PDF or CDF alone (Wanga et al., 2004).

The goal can be obtained transforming a multi-objective process in single objective optimization (Andersson, 2000), using the following eqs:

\[
\begin{align*}
  \min[\text{obj}] &= \max[CEF_{cdf} + CEF_{pdf}] \\
  \text{Where} \\
  CEF_{cdf/pdf} &= 1 - EF_{cdf/pdf}
\end{align*}
\]

Based on efficiency modelling parameter (EF)

By Nash and Sutcliffe (1970)

\[
\begin{align*}
  \min[\text{obj}] &= \max[Caik_{cdf} + Caik_{pdf}] \\
  \text{where:} \\
  Caik_{cdf/pdf} &= \text{Akaike information criteria (AIC)}
\end{align*}
\]

Based on Akaike inf. criteria
Monte carlo resampling of interpolated observed CDF. Interpolation made with Convex Cubic Rational splines with tensión with C1 and C2 properties. Usually 20000 random resampling on CDF.

Global statistics calculations

San Luis Potosí; Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosí
Two example of application:

• Debris avalanche in volcano, Nevado de Toluca, Mexico
• Olistostrome formation (large Deep-sea landslide deposits), Turkey
• Debris avalanche in volcano, Nevado de Toluca, Mexico

Mt. Sat. Helens 1980, flank collapse, landslide and debris avalanche

http://www.mshslc.org/wp-content/uploads/2013/05/2_debris_avalanche_intro.jpg

https://volcaniccollapse.files.wordpress.com/2011/10/image8.png

San Luis Potosi; Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosi
lognormal 4 parámetros – 3 componentes
Nevado de Toluca avalance - global Akaike inf criteria -196.53
Weibull 4 parámetros – 3 componentes
Nevado de Toluca avalanche - global Akaike inf. criteria -229.35
Weibull 4 parámetros – 2 componentes
Nevado de Toluca avalanche - global Akaike inf criteria -205.10
lognormal 4 parámetros – 3 componentes
Nevado de Toluca  avalance  -  global Akaike inf criteria  -190.15
LogNormal 4 parámetros – 3 componentes: algoritmo TDE
Nevado de Toluca avalancha - global Akaike inf. criteria -227.43
Weibull 4 parámetros – 3 componentes : algorithm TDE
Nevado de Toluca avalanche - global Akaike inf. criteria -227.43
• Olistostrome formation (large Deep-sea landslide deposits, usually then tectonized by orogenic process), Turkey

http://www.ub.edu/ggac/images/olistostrome.jpg

http://jsedres.sepmonline.org/content/77/4/263/F10.large.jpg

San Luis Potosí; Mexico, DICIM Seminar, 18 May 2016, Universidad autonoma de San Luis Potosi
lognormal 4 parámetros – 3 componentes – algorithm: DE classic
Olistostrome formation Turkey - global Akaike inf criteria -134.13

DECLOG (rel. 5.4 - 2016) - DECONVOLUTION OF MIXTURE'S COMPONENTS
INSIDE PARTICLE SIZE DISTRIBUTIONS
HTTP://WWW.DECLOG.ORG - By L.Borselli, D.Sarocchi, UASLP(Mexico), lbbborsell@gmail.com

INPUT DATA FILE >: oil2.dat
OUTPUT REPORT >: oil2.xls

## GLOBAL FITTING STATISTICS for CDF:
Model efficiency coefficient(EF) = 0.9993700
Coefficient of Determination(R^2) = 0.9997679
Kolmogorov-Smirnoff difference(Ks) = 0.0178862

## GLOBAL FITTING STATISTICS for PDF:
Model efficiency coefficient(EF) = 0.9781112
Coefficient of Determination(R^2) = 0.9891733

## STATISTICS FROM SPLINES RESAMPLED OBSERVED CDF (MonteCarlo Method):
Mean(Phi) = -1.82910
Standard deviation(Phi) = 3.73188 Very poorly sorted
Skewness = 0.35674 Symmetrical
Kurtosis(normalized) = -1.42413 Very platykurtic
Weibull 4 parámetros – 3 componentes – algoritmo: DE classic
Olistostrome formation Turkey - global Akaike inf criteria -153.24
Weibull 4 parámetros – 3 componentes
Olistostrome formation Turkey – estrategia de optimización “Efficiency modeling”
Now two practical real time examples of application of DECOLOG:

• Debris flow deposits, Motozintla, Chiapas, Mexico

• Block and ash flow (pyroclastic flow with large blocks due to dome collapse), Colima Volcano, Colima Mexico
Conclusions

• DECOLOG is a tool able to decode part of the information usually blind inside total particle size distribution PSD).
• DECOLOG allows the easy calculation of global statistic of using the el modelo optimized folk-Ward and by new Montecarlo resampling method (much better in complex polymodal distributions).
• A lot of options for deconvolution engine applied to the PSD that is in many cases a Mixture of components (phases)
• Algorithm of Multi-objective optimization by de combined non-linear fitting of experimental PDF y CDF using self-organizing and self evolving genetic algorithm: Trigonometric differential evolution (Fan y Lampinen 2003).
• Software fully FREWARE for scientific community (WWW.DECOLOG.ORG)

REFERENCES

Many thanks for Yours attention !!!