Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and
Computational Mathematics (IMACM)

Preprint BUW-IMACM 16/17

T.G.J. Beelen, J.J. Dohmen, E.J.W. ter Maten, B. Tasić

Fitting generalized Gaussian distributions for Process Capability Index

September 2016

http://www.math.uni-wuppertal.de

Fitting Generalized Gaussian Distributions for Process Capability Index

Theo G.J. Beelen[1], Jos J. Dohmen[2], E. Jan W. ter Maten[1,3], and Bratislav Tasić[2]

Abstract The design process of integrated circuits (IC) aims at a high yield as well as a good IC-performance. The distribution will not be standard Gaussian anymore. In fact, the corresponding probability density function has a more flat shape than in case of standard Gaussian. In order to optimize the yield one needs a statistical model for the observed distribution. One of the promising approaches is to use the so-called Generalized Gaussian distribution function and to estimate its defining parameters. We propose a numerical fast and reliable method for computing these parameters.

1 Introduction

We assume N independent samples x_i in some given interval $[U,V]$ and based on some empirical density function. To define a quality measure index we are now interested in the 'best' fitting function within the family of Generalized Gaussian Density (GGD) distributions as shown in Fig.1 and given by the expression

$$f(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}\right), \tag{1}$$

T.G.J. Beelen, E.J.W. ter Maten
Eindhoven University of Technology, the Netherlands,
e-mail: {T.G.J.Beelen,E.J.W.ter.Maten}@tue.nl

J.J. Dohmen, B. Tasic
NXP Semiconductors, Eindhoven, the Netherlands,
e-mail: {Jos.J.Dohmen,Bratislav.Tasic}@nxp.com

E.J.W. ter Maten
Bergische Universität Wuppertal, Germany,
e-mail: Jan.ter.Maten@math.uni-wuppertal.de

where $\alpha, \beta > 0$, $\mu \in \mathbb{R}$ and $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$, for $z > 0$, is the Gamma function [8, 14]. The mean and the variance of the GGD (1) are given by μ and $\alpha^2 \Gamma(3/\beta)/\Gamma(1/\beta)$, respectively. Hence after expressing $\alpha = \sigma\sqrt{\Gamma(1/\beta)/\Gamma(3/\beta)}$ we get that, for all β, the variance is σ^2. We note that for $\beta = 2$ one has $\Gamma(1/2) = \sqrt{\pi}, \Gamma(3/2) = 0.5\sqrt{\pi}$ and then $\alpha = \sigma\sqrt{2}$; i.e., the GGD becomes the Gaussian distribution. The parameter β determines the shape. For $\beta = 1$ the GGD corresponds to a Laplacian distribution; for $\beta \to +\infty$ the pdf in (1) converges to a uniform distribution in $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$, and when $\beta \downarrow 0$ we get a degenerate distribution in $x = \mu$ (but with a finite variance). For some graphical impression, see Fig. 1. We are interested in the cases when $\beta \geq 2$.

The parameters of the 'best' fitting distribution function can be found by maximizing the logarithm of the likelihood function $L = \ln(\mathscr{L}) = \sum_{i=1}^N f(x_i)$. The necessary conditions are

$$\frac{\partial L}{\partial \alpha} = 0 : \alpha = \left(\frac{\beta}{N} \sum_{i=1}^N |x_i - \mu|^\beta \right)^{1/\beta}, \tag{2}$$

$$\frac{\partial L}{\partial \beta} = 0 : \frac{1}{\beta} + \frac{\Psi(1/\beta)}{\beta^2} - \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{\mu}}{\alpha} \right|^\beta \ln \left| \frac{x_i - \hat{\mu}}{\alpha} \right| = 0, \tag{3}$$

$$\frac{\partial L}{\partial \mu} = 0 : \sum_{x_i \geq \mu} |x_i - \mu|^{\beta - 1} - \sum_{x_i < \mu} |x_i - \mu|^{\beta - 1} = 0. \tag{4}$$

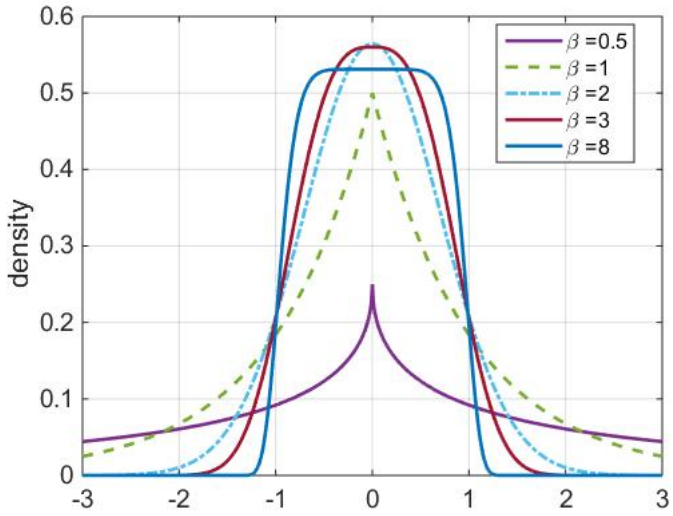

Fig. 1 Generalized Gaussian density functions with $\mu = 0$ and $\alpha = 1$.

Algorithm 1 Averaged Generalized Gaussian Distribution Fit

procedure AGGDF($\mathbf{X}$, N, M, $\hat{\mu}$)
  Determine the empirical pdf $\hat{f}(x)$ from the trimmed data $\mathbf{X}$. ▷ See Fig. 3 and [8]
  Compute the cumulative distribution function $\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(t)\mathrm{d}t$.
  for $k = 1, \ldots, M$ **do**
    Generate random values $\{x_i^k \mid i = 1, ..., N\}$ using $\hat{F}^{-1}$.
    Compute the zero $\hat{\beta}_k$ of $g(\beta) = 0$, using these x_i^k-values and $\hat{\mu}$. ▷ See (5)
    Compute $\hat{\alpha}_k$. ▷ See (2)
  end for
  Average $\hat{\beta} = \frac{1}{M}\sum_{i=1}^{M} \hat{\beta}_k$, $\hat{\alpha} = \frac{1}{M}\sum_{i=1}^{M} \hat{\alpha}_k$.
  return $\hat{\alpha}$, $\hat{\beta}$. ▷ In this Algorithm $\hat{\mu}$ is unchanged
end procedure

Here Ψ is the Digamma function $\Psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\ln(\Gamma(x)) = \Gamma'(x)/\Gamma(x)$ [14]. When we assume that $\mu = \hat{\mu}$ is known then we can ignore (4). Several papers consider estimates for α and β [2–4, 11, 12, 16] to solve the equations (2)-(3), but they assume that the sample size is large enough and/or that $\beta \leq 3$, motivated by the various application areas. We note that [6, 7] also consider the case for a small sample size. In general, the resulting estimators can be biased [8].
We exploit the explicit elimination of α in (3) after which only one additional equation remains

$$g(\beta) = g(\beta; \hat{\mu}) = 0, \tag{5}$$

in which $\mu = \hat{\mu}$ is now a given parameter. The analytical formulae for $g(\beta)$ and $g'(\beta)$ are given by (see also [1, 5, 7, 14])

$$\begin{aligned} g(\beta) =& 1 + \frac{\Psi(1/\beta)}{\beta} - \frac{\sum_{i=1}^{N} |x_i - \hat{\mu}|^{\beta} \ln|x_i - \hat{\mu}|}{\sum_{i=1}^{N} |x_i - \hat{\mu}|^{\beta}} + \frac{\ln\left(\frac{\beta}{N}\sum_{i=1}^{N} |x_i - \hat{\mu}|^{\beta}\right)}{\beta}, \\ g'(\beta) =& -\frac{\Psi(1/\beta)}{\beta^2} - \frac{\Psi'(1/\beta)}{\beta^3} + \frac{1}{\beta^2} \\ & - \frac{\sum_{i=1}^{N} |x_i - \mu|^{\beta} (\ln|x_i - \mu|)^2}{\sum_{i=1}^{N} |x_i - \mu|^{\beta}} + \left(\frac{\sum_{i=1}^{N} |x_i - \mu|^{\beta} \ln|x_i - \mu|}{\sum_{i=1}^{N} |x_i - \mu|^{\beta}}\right)^2 \\ & + \frac{\sum_{i=1}^{N} |x_i - \mu|^{\beta} \ln|x_i - \mu|}{\beta \sum_{i=1}^{N} |x_i - \mu|^{\beta}} - \frac{\ln\left(\frac{\beta}{N}\sum_{i=1}^{N} |x_i - \mu|^{\beta}\right)}{\beta^2}. \end{aligned} \tag{6}$$

Clearly, (5) can be solved by any (nonlinear) iterative method, for example by Newton's method using the expressions in (6). We have outlined our algorithm in Alg. 1. We consider the resulting density function $f(x;\, \hat{\mu}, \hat{\alpha}, \hat{\beta})$ as best fit to the measured data (see [1]). We make the following observations [1].

1. It might happen that (5) has no zero for some particular parameter choices. Numerical experiments (see Fig.2) indicate that (5) has a (unique) zero as long as $\bar{x} \neq \hat{\mu}$, and no zero in case of $\bar{x} = \hat{\mu}$;

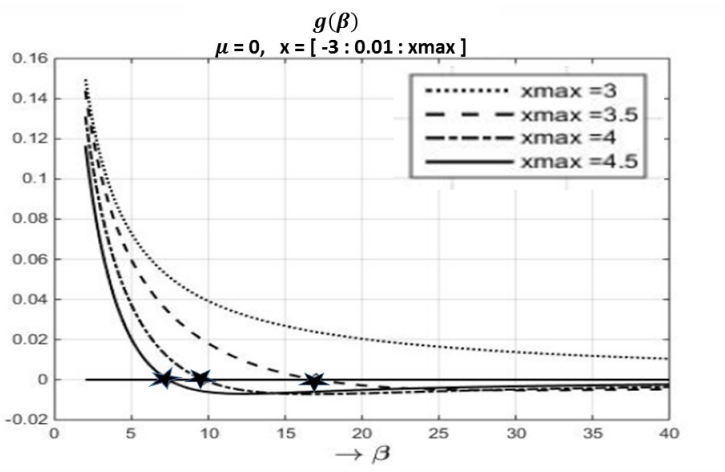

Fig. 2 Sensitivity of $g(\beta)$ with respect to the endpoint of x-interval.

2. When $M = 1$, $\hat{\beta}$ is very sensitive to the choice of the interval $[U,V]$ where the x_i are located. In fact, $\hat{\beta}$ strongly depends on the difference $|\hat{\mu} - \bar{x}|$ where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} |x_i - \hat{\mu}|$.

The first observation comes from the simplification by fixing $\mu = \hat{\mu}$, but is easy to meet.
For the second observation we introduced M-times the steps 5-7 within a loop and finally taking averages, see Alg. 1. By this, in practice, also N can be taken smaller. We observe that due to the large value of $\partial\alpha/\partial\beta$ averaging the $\hat{\alpha}_k$ gives better results for α than by using (2) on $\hat{\beta}$.

2 Numerical results

In circuit design one aims to reduce faults and to increase yield [10]. Specially added electronic control is applied to obtain narrow tails in empirical probability density functions. This process is called (electronic) 'trimming'. It has no relation to statistical techniques like Winsoring (in which one clips outliers to a boundary percentile), or Trimming (in which one simply neglects outliers). Here it is an electronic tuning, f.i., by a variable resistor. Assume that a some measurement point a circuit has a DC solution $V(R,p)$, that depends on a resistor R and an uncertain parameter p. The circuit design aims to satisfy a performance criterion $V_{\mathrm{Low}} \leq V \leq V_{\mathrm{Up}}$.

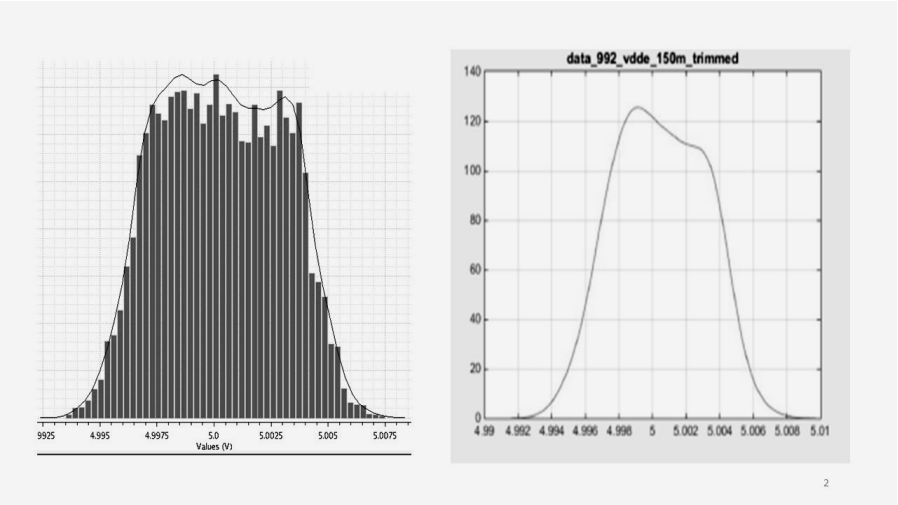

Fig. 3 Measured data (left) and the associated empirical probability density function $\hat{f}$ (right).

Now for each p we can determine how V depends on R. An optimal $R(p)$ assures that $V(R(p),p) = V_{\text{Ref}} \in [V_{\text{Low}}, V_{\text{Up}}]$. $R(p)$ can determined by some nonlinear solution techniqe, involving solving the circuit equations several times. We note that $R(p)$ can also be found by exploiting the expansion series in generalized polynomial chaos for Uncertainty Quantification using R and p as two parameters [9, 15]. The UQ facilities provide sensitivities to R and to p as library functionality in post processing. This allows that $R(p)$ can be determined quite efficiently for every realization of p. In practice a table can be made from which $R(p)$ can easily be determined or approximated.
We applied Alg. 1 (with $M = 50$) to 'trimmed' data from first NXP IC-measurements (Fig. 3). The computed values β_k and their mean $\hat{\beta}$ are shown in Fig.4-a. The computed density function f as well as the initially fitted (non-symmetrical) density function $\hat{f}$ are given in Fig. 5. Note that even the tails are very well approximated in Fig.4-b. To get an impression of the sensitivity of the computed density w.r.t. $\hat{\alpha}$ we varied the computed value of $\hat{\alpha}$ with +/- 10%, plotted the corresponding densities and computed the Mean Square Error (MSE). See Fig. 5 and [1].

3 A quality measure index for a Generalized Gaussian distribution

Assuming an underlaying distribution being standard Gaussian, the capability of a manufacturing process can be measured using some process capability indices like

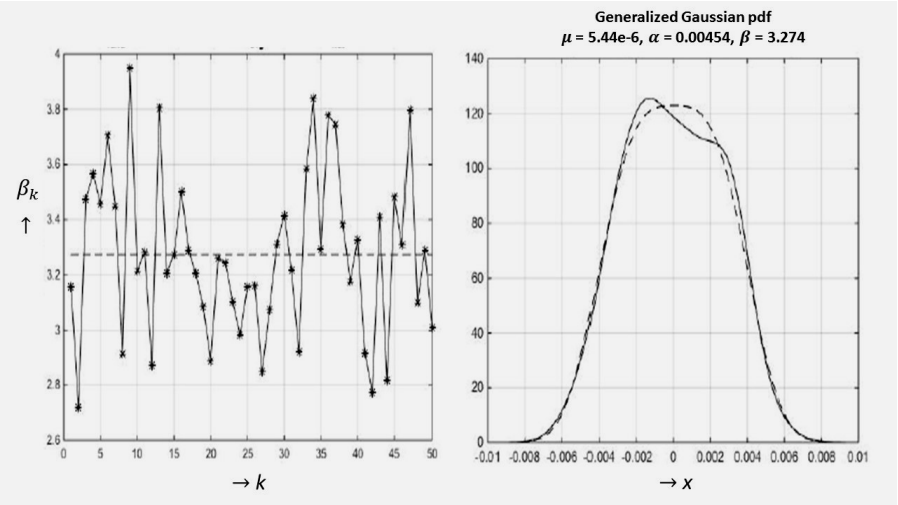

Fig. 4
(left): The computed β_k with mean $\hat{\beta} = 3.27$ and $|\beta_k - \hat{\beta}| < 20\%$.
(right): The empirical probability function (solid) and the final fitted GGD (dashed).

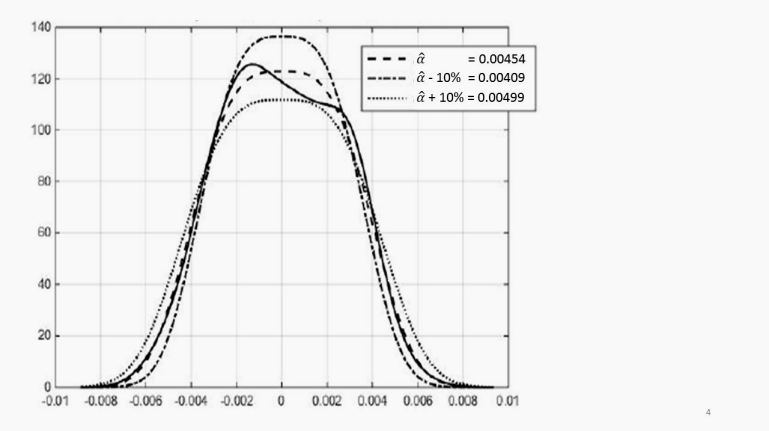

Fig. 5 Sensitivity of the density function f w.r.t. $\hat{\alpha}$.
$MSE = (14.31, 56.94, 91.95)$ for $\hat{\alpha} = (454, 499, 409) * 10^{-5}$.

$$C_p = \frac{U-L}{6\sigma} \text{ and } \mathrm{C_{pk}} = \frac{\min(\mathrm{U}-\mu, \mu-\mathrm{L})}{3\sigma}, \tag{7}$$

where $[L,U]$ is the specification interval, μ is the process mean and σ is the process standard deviation and a process is said to be capable if the process capability index exceeds a value $k \geq 1$, where usually $k = 4/3$. In case of a GGD (1) we can introduce

a capability index C_{pkg} similar to the standard Gaussian case as

$$C_{pkg} = \frac{\min(U-\mu, \mu-L)}{3\sigma}, \tag{8}$$

where $2\sigma^2 = \alpha^\beta$. L and U are the lower and upper tolerance levels, respectively. They can be determined as described below.
Notice that if $x \leq \mu$ then the cumulative distribution function $F(x)$ corresponding to the GGD (1) is given by

$$\begin{aligned} F(x) &= \frac{\beta}{2\alpha\Gamma(1/\beta)} \int_{-\infty}^{x} \exp\left(-\left(\frac{|y-\mu|}{\alpha}\right)^\beta\right) \mathrm{d}y \\ &= \frac{1}{2\Gamma(1/\beta)} \int_{((\mu-x)/\alpha)^\beta}^{\infty} \exp(-z)\mathrm{d}z. \end{aligned} \tag{9}$$

By using the Complementary Incomplete Gamma function defined by

$$\Gamma(a,x) = \int_x^\infty t^{a-1} exp(-t)\mathrm{d}t \tag{10}$$

we can rewrite (9) as

$$F(x) = \frac{\Gamma\left(1/\beta,\ \left(\frac{\mu-x}{\alpha}\right)^\beta\right)}{2\Gamma(1/\beta)} \tag{11}$$

This can be further simplified using the Upper Incomplete Gamma function [13, 14] for which standard software is available. For $x > \mu$ a similar expression holds.

4 Conclusions

We have shown that measured IC chip production data can adequately be modelled by a Generalized Gaussian distribution (GGD). We developed a new robust numerical procedure for computing the parameters of such GGD. The GGD did fit very accurately. Using the GGD a quality measure can be defined analogously to the CPK index for standard Gaussian distributions.

Acknowledgements The authors acknowledge support from the projects **CORTIF** (`http://cortif.xlim.fr/`): *Coexistence Of Radiofrequency Transmission In the Future*, a CATRENE project (Cluster for Application and Technology Research in Europe on NanoElectronics, `http://www.catrene.org/`) and **nanoCOPS** (`http://fp7-nanocops.eu/`): *Nanoelectronic COupled Problems Solutions*, FP7-ICT-2013-11/619166.

References

1. T.G.J. Beelen, J.J. Dohmen: *Parameter estimation for a generalized Gaussian distribution*. CASA Report 15-40, TU Eindhoven, 2015. Online: `http://www.win.tue.nl/analysis/reports/rana15-40.pdf`.
2. L. Bombrun, F. Pascal, J.-Y. Tourneret, Y. Berthoumieu: *Performance of the maximum likelihood estimators for the parameters of multivariate generalized Gaussian distributions*. Proc. ICASSP-2012, IEEE. Int. Conf. on Acoustics, Speech, and Signal Processing, Kyoto, Japan, 3525-3528, 2012. `https://hal.archives-ouvertes.fr/hal-00744600`.
3. G. González-Farías, J.A. Domínguez-Molina, R.M. Rodríguez-Dagnino: *Efficiency of the approximated shape parameter estimator in the generalized gaussian distribution*. IEEE Trans. on Vehicular Technology 58:8, 4214–4223, 2009.
4. K. Kokkinakis, P.C. Loizou: *Advances in modern blind separation algorithms – Theory and applications*. Synthesis Lectures on Algorithms and Software in Engineering #6, Morgan & Claypool Publishers, 2010
5. K. Kokkinakis, A.K. Nandi: *Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling*. Signal Processing 85, 1852–1858, 2005.
6. R. Krupiński: *Approximated fast estimator for the shape parameter of generalized Gaussian distribution for a small sample size*. Bull. of the Polish Academy of Sciences, Technical Sciences 63:2, 405–411, 2015.
7. R. Krupiński, J. Purczyński: *Approximated fast estimator for the shape parameter of generalized Gaussian distribution*. Signal processing 86, 205–211, 2006.
8. W.L. Martinez, A.R. Martinez: *Computational statistics handbook with Matlab*. Chapman & Hall/CRC, 2002.
9. E.J.W. ter Maten, R. Pulch, W.H.A. Schilders, H.H.J.M. Janssen: *Efficient calculation of Uncertainty Quantification*. In: M. Fontes, M. Günther, N. Marheineke (Eds): *Progress in Industrial Mathematics at ECMI 2012*. Series Mathematics in Industry Vol. 19, Springer, 361–370, 2014.
10. E.J.W. ter Maten, O. Wittich, A. Di Bucchianico, T.S. Doorn, T.G.J. Beelen: *Importance sampling for determining SRAM yield and optimization with statistical constraint*. In: B. Michielsen, J.-R. Poirier (Eds.): *Scientific Computing in Electrical Engineering SCEE 2010*. Series Mathematics in Industry Vol. 16, Springer, 39–48, 2012.
11. O.M.M. Mohamed, M. Jaïdane-Saïdane: *On the parameters estimation of the Generalized Gaussian Mixture Model*. Proc. EUSIPCO/EURASIP 2009, 17th European Signal Processing Conference, Glasgow, Scotland, 2273–2277, 2009.
12. A.A. Roenko, V.V. Lukin, I. Djurović, M. Simeunović: *Estimation of parameters for Generalized Gaussian Distribution*. Proc. ISCCSP-2014, 6th Int. Symp. on Communications, Control and Signal Processing, Athens, Greece, 376–379, 2014.
13. N.M. Temme: *Computational aspects of Incomplete Gamma Functions with large complex parameters*. Int. Series of numerical Mathematics 119, 551–562, 1994.
14. Wikipedia: *Generalized Normal Distribution*. `https://en.wikipedia.org/wiki/Generalized_normal_distribution`; *Gamma function*. `https://en.wikipedia.org/wiki/Gamma_function`; *Digamma function*. `https://en.wikipedia.org/wiki/Digamma_function`; *Incomplete Gamma Function*. `https://en.wikipedia.org/wiki/Incomplete_gamma_function`. 2016.
15. D. Xiu: *Numerical methods for stochastic computations - A spectral method approach*. Princeton Univ. Press, Princeton, NJ, USA, 2010.
16. S. Yu, A. Zhang, H. Li: *A review of estimating the shape parameter of Generalized Gaussian Distributions*. J. of Comput. Information Systems, 8:12, pp. 9055–9064, 2012.