**Solid Earth**

# The stochastic quantization method and its application to the numerical simulation of volcanic conduit dynamics under random conditions

**E. Peruzzo[1,*], M. Barsanti[2,3], F. Flandoli[2], and P. Papale[3]**

[1]Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126, Pisa, Italy
[2]Dipartimento di Matematica Applicata, Università di Pisa, Via Buonarroti 1c, 56127, Pisa, Italy
[3]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Via della Faggiola 32, 56126, Pisa, Italy
[*]*Invited contribution by E. Peruzzo, one of the EGU, Young Scientists Outstanding Poster Paper Award winners 2009*

**Abstract.** Stochastic Quantization (SQ) is a method for the approximation of a continuous probability distribution with a discrete one. The proposal made in this paper is to apply this technique to reduce the number of numerical simulations for systems with uncertain inputs, when estimates of the output distribution are needed. This question is relevant in volcanology, where realistic simulations are very expensive and uncertainty is always present. We show the results of a benchmark test based on a one-dimensional steady model of magma flow in a volcanic conduit.

## 1 Introduction

Since the demand for eruption scenario forecast in the world is pressing, there is a strong need for using complex physical models and numerical codes in order to get information about the possible eruptive conditions at many hazardous volcanoes (e.g., Sparks, 2003; Neri et al., 2007). Such models can describe volcanic processes thoroughly, but this ability often results in high computational costs: a single simulation can require a time of the order of days to weeks to be completed.

On the other hand, since volcanic systems are largely inaccessible to direct observation, the models which describe them often involve intrinsically uncertain quantities. As a consequence, some of the input data required by the numerical codes should be considered as random variables rather than as fixed parameters. Therefore, the most direct way of obtaining information about the probability of possible eruptive scenarios would be to implement a Monte Carlo (MC) method. As a drawback, this would require a large number of numerical simulations (e.g. $10^4$), while often a number of the order of 10 simulations cannot be exceeded.

It is thus fundamental to be able to choose the input data values in such a way that the simulations provide the maximum amount of information possible about the output quantities. Furthermore, the selection of the "best" sets of values of input data should be guided by fairly general principles, which could be applied to a large class of models and numerical codes and are not devised for a particular situation. The strategy presented in this paper is indeed general, since the choice of the optimal sets of input data does not involve the numerical code at all. In this respect, a different (and somewhat complementary) approach to the problem would be that of using the simulations to construct a function which is a reasonably good approximation to the complex code and, at the same time, can be evaluated with a low computational effort (Sacks et al., 1989; Currin et al., 1991). The simplified function produced could then be used in a Monte Carlo simulation. In this case, the choice of the sets of input data used

to find the approximating function is less critical, while the focus is on the approximation of the numerical code. A possibility that has been explored in the mathematical literature is that of employing Bayesian inference to select a function with the required properties (Kennedy and O'Hagan, 2001). Our approach, on the other hand, aims at defining an optimal set of input data, that sufficiently describes the output distribution. Future developments may involve a mixed approach, whereby some gross properties of the numerical code are exploited to guide the choice of the optimal sets of input data and corresponding output distribution.

In the first section of this paper we present our approach to the problem, showing that the stochastic quantization (SQ) method provides a possible solution. In Sects. 2 and 3 we briefly sketch the basic principles of one-dimensional and multi-dimensional quantization. In Sect. 4 we show the application of our strategy to an ideal case, while in Sect. 5 we turn to a more realistic situation, involving the one-dimensional steady model of magma flow in a volcanic conduit described in (Papale, 2001). We show that, given a maximum number $N$ of simulations, the results obtained using the SQ method to choose $N$ sets of values of input data are better than those produced by $N$ MC simulations.

The origins of the theory of quantization of probability distributions date back to the three fundamental articles (Oliver et al., 1948; Bennett, 1948; Panter and Dite, 1951); the field of application was that of signal processing, in particular modulation and analog-to-digital conversion. See (Gray and Neuhoff, 1998) for a survey of the method and of its engineering applications, besides a huge list of references (mainly in the field of engineering); the book (Graf and Luschgy, 2000) gives a mathematically rigorous treatment of the subject.

## 2   Outline of the strategy

In order to illustrate our strategy, consider a practical situation: suppose that a numerical code for the simulation of some volcanic processes has the random variable $X$ among its input data. Likewise, $X$ can be a collection of input random variables $(X_1, \ldots, X_d)$, i.e. a $d$-dimensional random vector. Let $Y$ be one relevant output quantity of the numerical code. We denote by $f(x_1, \ldots, x_d)$ and $g(y)$ the probability density functions of $X$ and $Y$, respectively; $f$ is assumed to be known, while nothing is known about $g$. We also suppose that the numerical code has such a high degree of complexity that the maximum number $N$ of affordable simulations is very small, of the order of 10. It is thus not possible to collect information about $g$ using a MC method.

Our strategy (see Fig. 1) consists in three main steps and an optional fourth step.

1. Find $N$ values of the random vector $X$,

$$\left(x_1^{(1)}, \ldots, x_d^{(1)}\right), \ldots, \left(x_1^{(N)}, \ldots, x_d^{(N)}\right)$$

and $N$ corresponding weights

$$\left(w^{(1)}, \ldots, w^{(N)}\right),$$

with $\sum_{i=1}^{N} w^{(i)} = 1$, such that the discrete probability distribution $\hat{f}$ which assigns the weight $w^{(i)}$ to the point $\left(x_1^{(i)}, \ldots, x_d^{(i)}\right)$ (for $i = 1, \ldots, N$) is the optimal approximation of $f$, among all the discrete probability distributions concentrated in $N$ points. The meaning of optimality is to be precised later.

2. Perform $N$ numerical simulations to compute the $N$ corresponding values $y^{(1)}, \ldots, y^{(N)}$ of the random variable $Y$. We represent the action of the numerical code on the input data through a function $\varphi$, so that

$$y^{(i)} = \varphi\left(x_1^{(i)}, \ldots, x_d^{(i)}\right) \text{ for } i = 1, \ldots, N.$$

3. Build a discrete approximation $\hat{g}$ of $g$ by assigning the weight $w^{(i)}$ to the point $y^{(i)}$ for $i = 1, \ldots, N$.

4. Use $\hat{g}$ to build a continuous probability distribution.

The core of the problem is thus to find the "best" discretization of a fixed continuous probability distribution $f$ (step 1). Stochastic quantization is a mathematical theory which allows to accomplish this task by giving a definition of the optimal discretization and providing an algorithm to find it. Once this is done, a discrete approximation of the unknown output probability distribution is automatically produced (steps 2 and 3). Concerning step 3, it is a rigorous fact that the transformation of a discrete probability distribution by a function $\varphi$ is the discrete probability distribution having the same weigths on the image points. Thus step 3 is the most natural choice. If we would know the output density $g$, a better choice would be the weigths given by the SQ algorithm applied to $g$ and the given output points. But we do not know $g$. It is an open problem to improve step 3 in this direction.

Though the discrete probability distribution $\hat{g}$ can be used to estimate the values of some parameters of $g$ (e.g. its mean, its variance, some quantiles), its graphical representation gives poor insight into the main qualitative features of $g$. This is one of the motivations of step 4, which can be carried out through a kernel smoothing algorithm (e.g., Wand and Jones, 1995). Moreover, it may be useful for other purposes, like random number generation, to have a continuous distribution in output. The main idea of kernel smoothing algorithms is to smear out each weight $w^{(i)}$ around the corresponding point $y^{(i)}$ according to a fixed rule and then to sum up all the contributions. This leads to the continuous probability distribution

$$\hat{g}_{KS}(y) = \frac{1}{N} h \sum_{i=1}^{N} K\left(\frac{y - y^{(i)}}{h}\right), \tag{1}$$

(a) Known input distribution.



(b) Unknown output distribution.



(c) Input discretization, with $N = 10$.
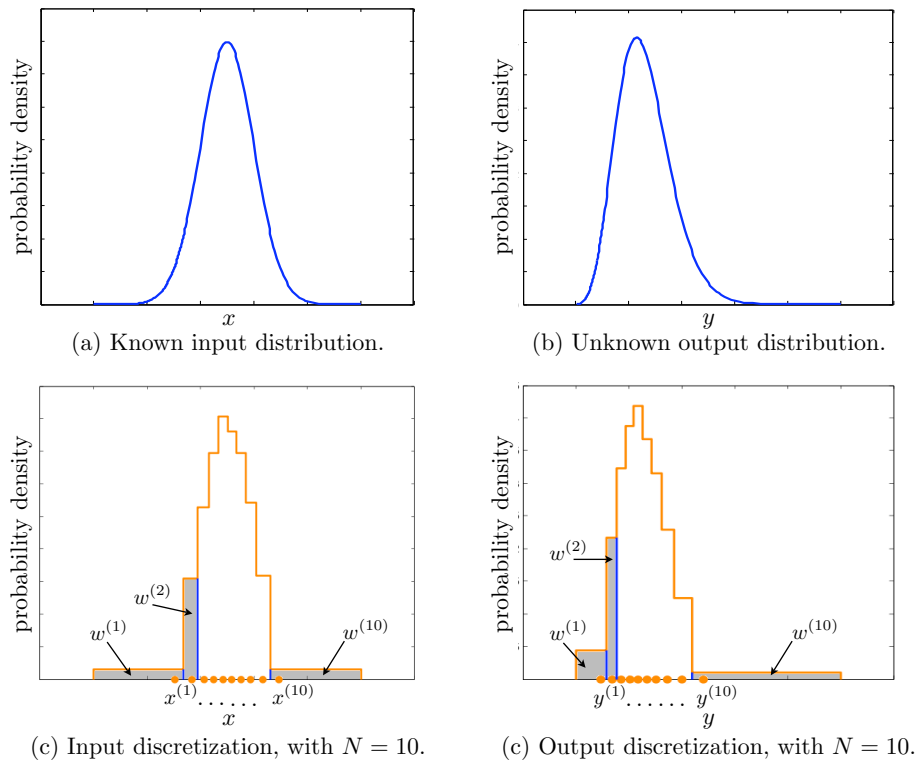


(c) Output discretization, with $N = 10$.

**Fig. 1.** Graphical representation of the first three steps of our strategy, with $d=1$ and $N=10$. The lower graphs represent the discrete approximations of the input and output probability distributions produced by the SQ method. The weight $w^{(i)}$ is the area of the rectangle associated to the $i$-th point for $i = 1, \ldots, N$.

where $K$ is a smooth probability distribution and $h$ is a measure of the width of the interval over which each weight is spread. $h$ is chosen according to an optimality criterion, based on the minimization of some kind of error resulting from the substitution of $g$ with $\hat{g}_{KS}$. $K$ is usually chosen to be a unimodal probability density symmetric about zero, but its exact expression does not affect very much the result.

## 3   Quantization of univariate probability distributions ($d=1$)

As a first step, in order to give a precise meaning to the expression "best approximation", we introduce a distance between probability distributions. Consequently, the optimal discretization of $f$ can be defined as the discrete probability distribution $\hat{f}$ which has the minimum distance from $f$.

Since we have to compare discrete and continuous probability distributions, it is easier to rely on the cumulative distribution functions, especially in the case $d=1$, in which there is only one input random variable $X$. Let $F$ be the cumulative distribution function associated with the density $f$ and $\hat{F}$ the one associated with $\hat{f}$: namely (see Fig. 2),

$$F(x) = \int_{x_{\min}}^{x} f(t)dt,$$

$$\hat{F}(x) = \sum_{x^{(i)} \leq x} \hat{f}\left(x^{(i)}\right) = \sum_{x^{(i)} \leq x} w^{(i)},$$

for $x_{\min} \leq x \leq x_{\max}$,

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of $X$, $x^{(i)}$ are the points in which $\hat{f}$ is concentrated and $w^{(i)}$ are the corresponding weights.

For instance, we can define a distance between $f$ and $\hat{f}$ as

$$d(f, \hat{f}) = \int_{x_{\min}}^{x_{\max}} \left| F(x) - \hat{F}(x) \right| dx \qquad (2)$$

(see Fig. 3). Therefore, the procedure consists in searching for the discrete probability distribution $\hat{f}$ which minimizes the quantity $d(f, \hat{f})$, among all the discrete distributions concentrated in $N$ points.

## 4   Quantization of multivariate probability distributions ($d>1$)

A more complicated situation arises when there are several random variables $X_1, \ldots, X_d$ in input, or equivalently a single random vector $X=(X_1, \ldots, X_d)$. If $X_1, \ldots, X_d$ are independent and $f_1(x_1), \ldots, f_d(x_d)$ are their probability density functions, then $f(x_1, \ldots, x_d) = f_1(x_1) \times \ldots \times f_d(x_d)$ is the
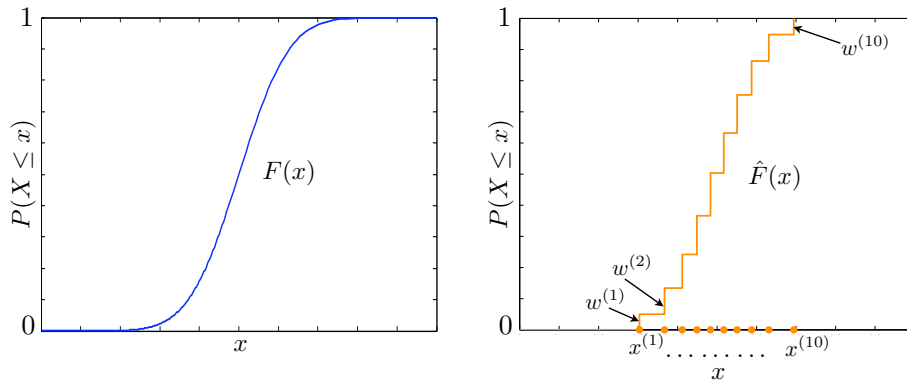
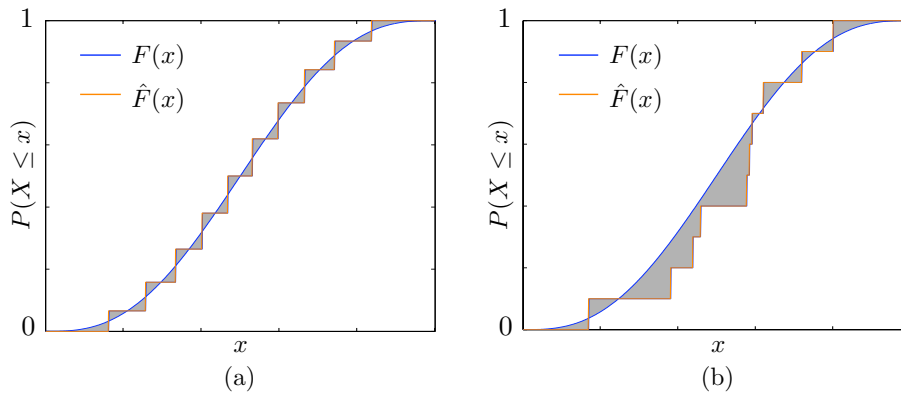**Fig. 2.** Continuous and discrete cumulative distribution functions.



**Fig. 3.** The distance between $f$ and $\hat{f}$ is defined as the shaded region area in the graphs. Optimal discrete distributions look like the one in graph **(a)**, while graph **(b)** represents a sub-optimal discrete distribution.

probability density function of $X$. If $X_1,\ldots,X_d$ are not independent, the probability density function $f(x_1,\ldots,x_d)$ of $X$ is not related to $f_1(x_1),\ldots,f_d(x_d)$ in an obvious way. However, the independency hypothesis is not necessary for the SQ algorithm.

A definition of distance similar to the one in Eq. 2 could be given, but it is easier and more appropriate to use another definition of distance, based on the random variables rather than on their cumulative distribution functions.

Let $\hat{X}$ be a discrete random vector with probability distribution $\hat{f}$, approximating the continuous random vector $X$; it seems quite natural to choose, as a measure of the distance between $f$ and $\hat{f}$, the mean value of the error $|X-\hat{X}|$ resulting from the substitution of $X$ with $\hat{X}$:

$$d(f,\hat{f}) = \mathbb{E}\left[|X - \hat{X}|\right]. \qquad (3)$$

The distance defined by Eq. (3) could be computed numerically, but the calculation is easier and faster (especially in high dimension) via a MC method; Appendix A shows in detail how it can be performed. The MC simulations involved in the algorithm use only samples of $X$ and $\hat{X}$, which can be generated easily, and not samples of $Y$, whose generation

is out of reach. The randomness on the value of $d(f,\hat{f})$ (and, as a consequence, on the values of the "optimal" points $\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(N)}$) due to the choice of a stochastic algorithm can be made negligible, provided that the number $M$ of MC simulations is large enough. Moreover, the stochastic algorithm easily provides the weights $w^{(1)},\ldots,w^{(N)}$ (see Appendix A).

In appendix A we also show that, for univariate distributions, the distances in Eqs. (2) and (3) lead to the same optimal discretization: therefore the definition given by Eq. (3), besides having an intuitive motivation, is a natural generalization of that given by Eq. (2).

## 5   Testing SQ in a simple case

In order to assess the quality of the approximation of the output probability distribution given by the SQ, we first consider a simple test case.

– There is a 2-dimensional random vector $X=(X_1,X_2)$ in input; $X_1$ and $X_2$ are independent and their probability distributions are both gaussians, truncated at 0 and at 1; the distribution of $X_1$ has mean 0.5 and standard
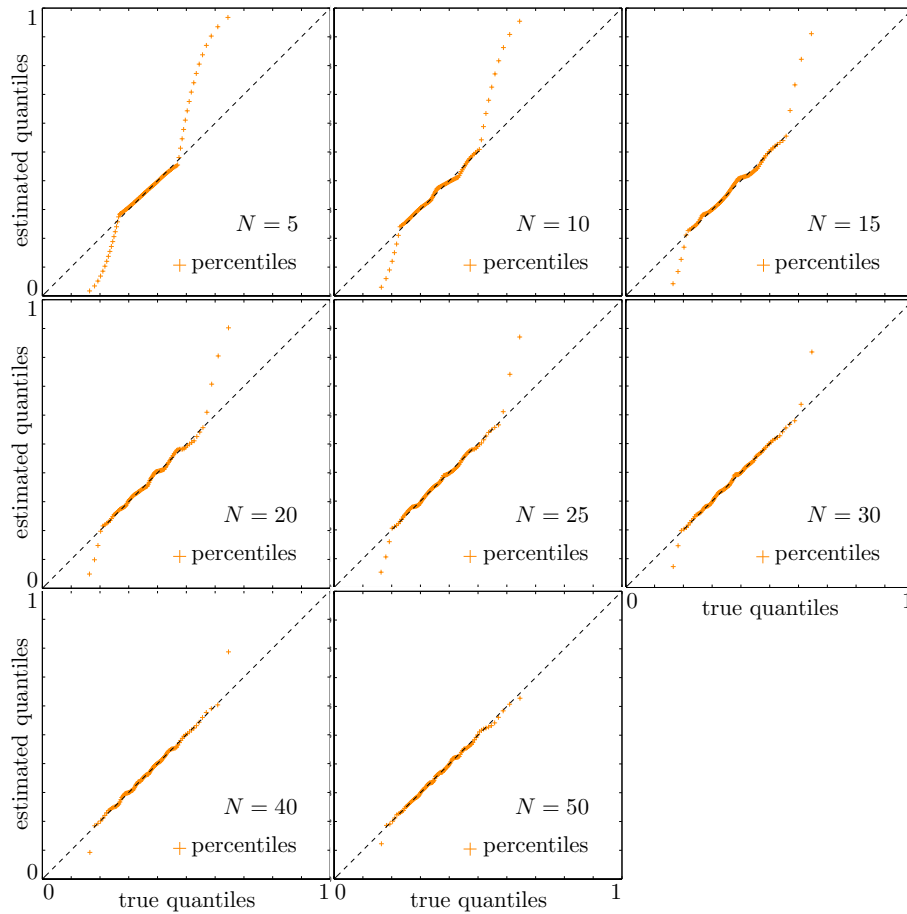
**Fig. 4.** Quantile-quantile plots. On the x-axis there are the true values of the quantiles of $g$, on the y-axis their estimates given by the SQ. The orange daggers represent the percentiles, from 1 to 99. If a dagger is close to the dashed line $y=x$, the SQ gives a good estimate of the corresponding percentile. Different graphs refer to different numerosities of the SQ.

deviation 0.2, that of $X_2$ has mean 0.6 and standard deviation 0.1.

– The relation between $X$ and the output random variable $Y$ is known and has a simple analytical expression:

$$Y = \varphi(X_1, X_2)$$
$$= \frac{1}{8}(X_1^2 + X_1)(X_2^2 + X_2) + \frac{1}{4}(X_1 + X_2). \quad (4)$$

In such a simple case it is possible to implement a MC method with very high numerosity (e.g. $10^5$), producing an estimate of $g$ which can be considered exact for practical purposes. This version of $g$ can be directly compared to the results produced by the SQ method and by MC methods with variable numerosity, in order to establish which one gives the best approximation of $g$: for instance, the SQ method can be compared to a MC with the same numerosity, or to other low numerosity MC simulations. For each numerosity, the MC can be repeated several times, thanks to the simplicity of the

function $\varphi$; consequently, it is possible to estimate the probability that the performances of the SQ are better than those of the MC (see Fig. 5).

Figure 4 compares some quantiles of $g$, whose values are very well known thanks to the high numerosity MC, to their estimates produced by the SQ method. These estimates are obtained via a linear interpolation of the cumulative distribution associated to $\hat{g}$. Figure 4 shows that, as the numerosity $N$ of the SQ grows, the estimates of the percentiles of $g$ globally improve. $N=15$ is a first good compromise; then the improvement is not so strong, until $N=40$ or 50, where the result is almost perfect. On the other hand, the central quantiles do not become better and better but fluctuate around the true values in an unpredictable manner. This is the reason why the estimate of a parameter sometimes gets worse even if $N$ increases, as is shown in Fig. 5: for instance, the estimate of the median of $g$ (red curve) gets worse as $N$ passes from 5 to 10. This makes it possible for MC simulations to rapidly achieve better estimates of the median than 10 points SQ as their numerosity increases.
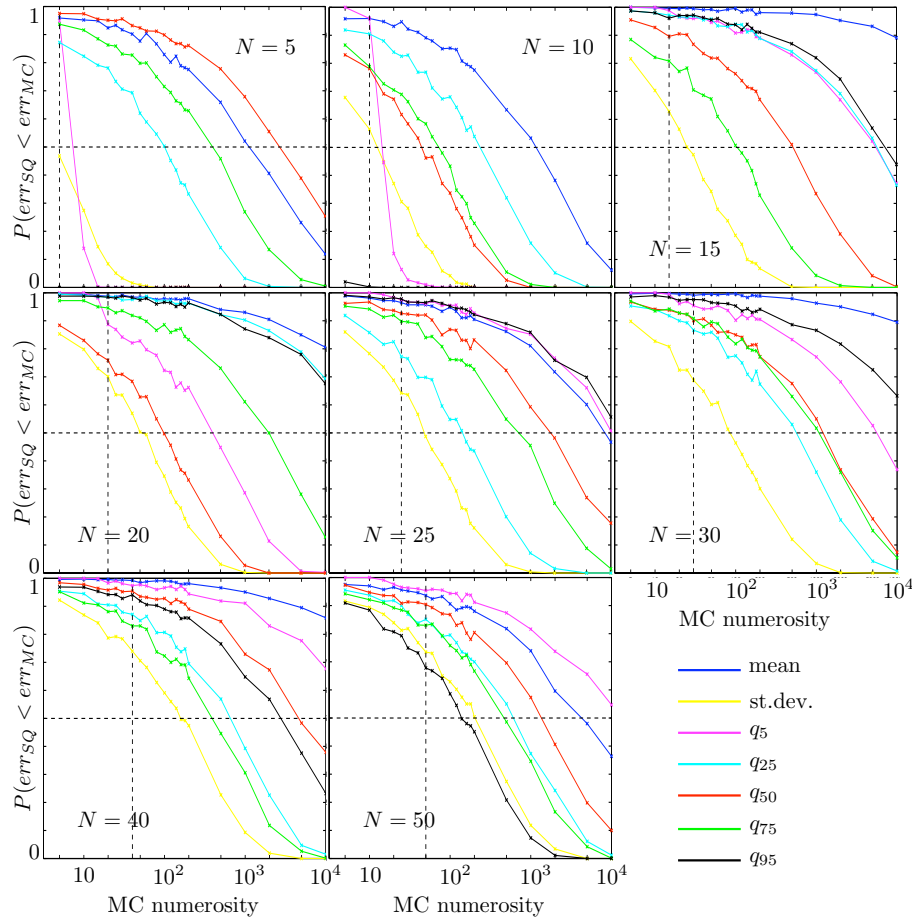
**Fig. 5.** Comparison between the SQ and MC simulations. On the $x$ axis there is the numerosity of the MC, in logarithmic scale; on the $y$ axis there is the probability that $\mathrm{err}_{SQ} < \mathrm{err}_{MC}$, where $\mathrm{err}_{SQ} = \left| p - p_{SQ} \right|$ and $\mathrm{err}_{MC} = \left| p - p_{MC} \right|$, $p$ is the real value of a parameter of the output probability distribution $g$, $p_{SQ}$ and $p_{MC}$ are its estimates given by SQ and MC, respectively. The curves represent the results for various parameters: mean, standard deviation and the 5%, 25%, 50%, 75%, 95% quantiles. The horizontal dashed line marks the value 0.5 of probability. Different graphs refer to different values of the numerosity $N$ of the SQ. In each graph, a vertical dashed line is drawn in correspondance of the MC numerosity equal to $N$.

However, Fig. 5 also shows that there is a general tendency to improvement as $N$ grows. The whole bundle of curves, in fact, moves gradually towards the upper right corner of the graph: this means that the probability of the estimates given by the SQ being better than those given by the MC is generally increasing as $N$ grows. Furthermore, it is evident that, if $N$ is sufficiently high ($N > 10$ in this case), the SQ method gives better results than MC simulations with the same numerosity with probability greater than 0.5.

## 6  Application of SQ to volcanic conduit dynamics

As a test application of the SQ approach to a volcanologically relevant case, we consider a one dimensional steady model of magma flow in a cilindrical conduit with fixed diameter and uniform temperature (Papale, 2001). This model is ideal for

testing SQ, since it provides a set of volcanologically relevant, strongly non-linear equations relating input and output distributions in a complex, unpredictable way, despite keeping the computational time small enough (order of minutes for each simulation) to allow a MC simulation with $N = 10^3$. The output distribution given by this MC is reasonably close to the exact one and can be used for comparison with SQ. Hence, this first application of SQ to a volcanologically relevant case is also a further test of the method.

Among the several input quantities that are intrinsically uncertain we choose two of them, namely, the diameter $D$ of the conduit and the total mass fraction of water $w_{H_2O}$. These two quantities are known to largely control the conduit flow dynamics and the associated mass flow-rate (e.g., Wilson et al., 1980; Papale et al., 1998). $D$ and $w_{H_2O}$ are therefore considered as random variables. We assign to each of them a probability distribution (truncated gaussian) and we study the
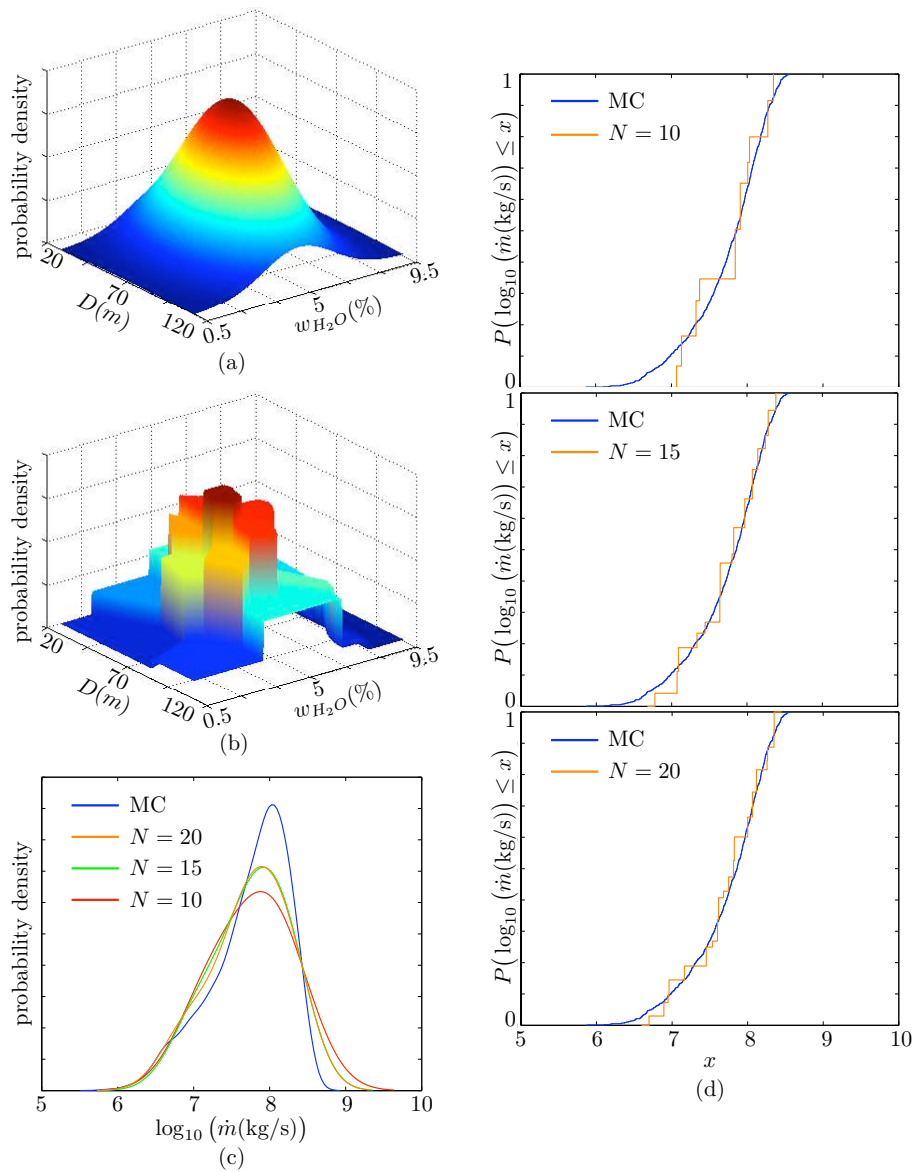
**Fig. 6.** Graph **(a)** represents the probability distribution of the input random vector $\left(D, w_{H_2O}\right)$, while in graph **(b)** there is its discretization produced by the SQ with $N=20$. Graph **(c)** shows a comparison between the approximation of $g$ resulting from $10^3$ MC simulations and $\hat{g}_{KS}$, which is obtained through the application of a kernel smoothing algorithm to the discrete distribution $\hat{g}$ given by the SQ with $N=10, 15, 20$. Graph **(d)** represents a comparison between the cumulative distribution function resulting from $10^3$ MC simulations and those resulting from the SQ with $N=10, 15, 20$, without application of kernel smoothing.

corresponding probability distribution of the logarithm of the mass-flow rate $\dot{m}$. The latter is a volcanologically relevant quantity, as it defines the intensity of an eruption and as it largely affects the impact on the surroundings (Valentine and Wohletz, 1989; Todesco et al., 2002).

Figure 6a shows the assumed continuous distribution of the two random variables $w_{H_2O}$ and $D$, while Fig. 6b illustrates the result of the application of the SQ method to discretize the distribution in 20 $w_{H_2O}$-$D$ pairs. In order to test the SQ method, the discretization has been done also with

15 and 10 $w_{H_2O}$-$D$ pairs. The results have been compared, in terms of output probability density (Fig. 6c) and cumulative probability distribution (Fig. 6d).

The three SQ cases with $N=10$, 15 and 20 reproduce well the mode of the distribution at a mass flow-rate of about $10^8$ kg/s, but fail in predicting correctly the shape of the distribution, resulting in a larger and less skewed curve with respect to the MC case. The larger values of density predicted by SQ at the high mass flow-rate tail largely explain the lower values of the mode with respect to the MC case.
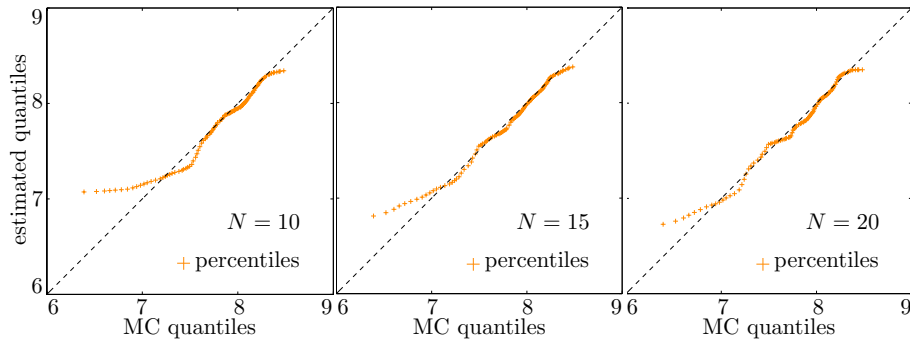
**Fig. 7.** Quantile-quantile plots, analogous to Fig. 4. On the x-axis there are the estimates of the quantiles of $g$ given by a MC simulation with $N = 10^3$, which are close to the true values.
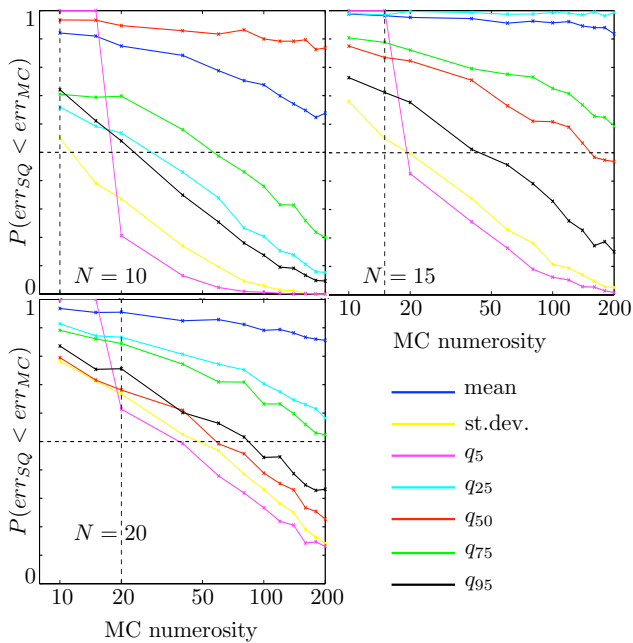


**Fig. 8.** Comparison between the SQ and MC simulations, analogous to Fig. 5.

**Table 1.** Comparison between the estimates of some parameters of the output distribution given by the SQ with $N = 10, 15, 20$ points and by a MC with $10^3$ simulations. The considered parameters are the mean, the standard deviation and the 5%, 25%, 50%, 75%, 95% quantiles.

|              | mean | st. dev. | $q_5$ | $q_{25}$ | $q_{50}$ | $q_{75}$ | $q_{95}$ |
|--------------|------|----------|-------|----------|----------|----------|----------|
| MC           | 7.74 | 0.50     | 6.71  | 7.45     | 7.86     | 8.12     | 8.38     |
| SQ, $N=10$   | 7.75 | 0.41     | 7.09  | 7.32     | 7.86     | 8.07     | 8.32     |
| SQ, $N=15$   | 7.74 | 0.44     | 6.94  | 7.46     | 7.82     | 8.10     | 8.33     |
| SQ, $N=20$   | 7.73 | 0.46     | 6.85  | 7.49     | 7.79     | 8.09     | 8.35     |

On the contrary, the left tail of the distribution, corresponding to the minimum mass flow-rates, is predicted accurately by the SQ. The improvement due to increased numerosity of SQ from 10 to 20 is clearly visible from the cumulative plots in Fig. 6d.

Figure 7 compares the estimates of the quantiles of the output distribution of mass flow-rates given by MC and SQ with $N=10$, 15 and 20. The quantiles are obtained by means of a linear interpolation of the cumulative distributions in Fig. 6d. As for the analogous Fig. 4 (referring to the polynomial map at Eq. 4), the bulk of the distribution is predicted well by the SQ method, but the tails of the distribution are not. While a significant improvement clearly emerges from $N=10$ to $N=15$, there is no significant gain in accuracy when moving

from $N=15$ to $N=20$. Table 1 shows the same results numerically for a few selected quantiles.

As for the ideal case discussed above, it is possible to compare the performances of the SQ with those of some low numerosity MC simulations (Fig. 8). This time, the discrete distributions generated by the SQ and by the low numerosity MC simulations are both compared to the one obtained from the MC simulation with $N=10^3$. In all cases, and for any quantity investigated, the SQ provides a better approximation of the output distribution than the MC with equal numerosity. For many quantities a MC with at least hundreds of simulations is required in order to exceed the accuracy given by SQ simulations with numerosity up to 20.

# 7   Conclusions

The performance of the SQ method has been analyzed both for an artificial polynomial map with random input and for a more complex set of non-linear equations. The latter case also represents an application of the SQ method to the volcanologically relevant case of steady multiphase magma flow along a volcanic conduit. Our analysis includes both a comparison between SQ and pure MC (Monte Carlo) method, and the accuracy of SQ in itself. In both cases the SQ method

provides substantially better estimates of output distributions than the MC method with the same number of simulations. This property is already clear with values of $N$ around 15–20, and it becomes stronger with the increase of $N$. The results are less definite for very small $N$, like $N=5$ and sometimes 10, where further ideas and research are needed. In general, MC gives results comparable to SQ only when the number of MC simulations is much higher, sometimes by one or more orders of magnitude, than that of SQ. Therefore, the SQ method results in a considerable computational saving for the same degree of accuracy of the estimates. With values of $N$ of 15 or 20, in general the estimates obtained by SQ are very close to the true ones (or to our better estimates of the true ones).

In conclusion, the SQ method allows the introduction of uncertainties in the deterministic approach without requiring exceeding CPU time. This result is promising for the capability of estimating future volcanic scenarios and volcanic hazards by means of a merged deterministic-probabilistic approach, whereby complex deterministic models are employed by taking into account the intrinsic uncertainties involved in the definition of the conditions characterizing the volcanic systems.

## Appendix A

### The numerical algorithm

This appendix describes in detail how the distance $d(f, \hat{f})$, defined in Eq. (3), is approximately computed via a stochastic algorithm. Morevorer, it shows that the optimal discretization of $f$ is found by moving the points $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$ in such a way that $d(f, \hat{f})$ is minimized, i.e. by minimizing a function of $N$ $d$-dimensional vectors (function $h$ in Eq. A3 below).

The starting point is represented by a fundamental result of the SQ theory (Graf and Luschgy, 2000, Lemma 3.1), which states that, if the $N$ possible values

$$\left(x_1^{(1)}, \ldots, x_d^{(1)}\right), \ldots, \left(x_1^{(N)}, \ldots, x_d^{(N)}\right)$$

of the random vector $\hat{X}$, i.e. the $N$ points in which $\hat{f}$ is concentrated, are fixed, then the corresponding optimal weights $w^{(1)}, \ldots, w^{(N)}$ are uniquely determined.

More precisely, the weights are defined as follows.

 – For $i=1, \ldots, N$, let $V_i$ be the region of the $d$-dimensional space such that

$$\boldsymbol{x} \in V_i \Longleftrightarrow \left|\boldsymbol{x} - \boldsymbol{x}^{(i)}\right| = \min_{k=1,\ldots,N} \left|\boldsymbol{x} - \boldsymbol{x}^{(k)}\right|,$$

where $\boldsymbol{x}^{(k)} = \left(x_1^{(k)}, \ldots, x_d^{(k)}\right)$ for $k=1, \ldots, N$. $V_i$ is called the Voronoi region of $\boldsymbol{x}^{(i)}$ with respect to the set $\left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\right\}$; it contains the points which are closer to $\boldsymbol{x}^{(i)}$ than to any other element of the set $\left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\right\}$ (see figure A1).

 – The optimal approximation $\hat{X}$ of the random vector $X$ is defined as follows: $\hat{X} = \boldsymbol{x}^{(i)}$ if and only if the value of $X$ belongs to $V_i$. Namely, $\hat{X}$ is obtained by rounding off $X$ to the nearest vector among $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$.

 – Correspondingly, $w^{(i)}$ (i.e. the probability that $\hat{X} = \boldsymbol{x}^{(i)}$) is the weight assigned to $V_i$ by the probability distribution $f$:

$$w^{(i)} = \int_{V_i} f(\boldsymbol{x}) d\boldsymbol{x}.$$

If $\hat{X}$ is defined as just described, it can be shown (Graf and Luschgy, 2000, Lemma 3.4) that, in the case $d=1$,

$$\mathbb{E}\left[|X - \hat{X}|\right] = \int_{x_{\min}}^{x_{\max}} \left|F(x) - \hat{F}(x)\right| dx;$$

moreover, if $\hat{X}'$ is another random variable, with the same possible values $x^{(1)}, \ldots, x^{(N)}$ but defined in whatever way, and $\hat{F}'$ is its cumulative distribution function, it can be shown that

$$\mathbb{E}\left[|X - \hat{X}'|\right] \geq \int_{x_{\min}}^{x_{\max}} \left|F(x) - \hat{F}'(x)\right| dx \qquad (A1)$$

$$\geq \int_{x_{\min}}^{x_{\max}} \left|F(x) - \hat{F}(x)\right| dx = \mathbb{E}\left[|X - \hat{X}|\right].$$

This means that, in the case $d=1$, minimizing $\int_{x_{\min}}^{x_{\max}} \left|F(x) - \hat{F}(x)\right| dx$ is the same as minimizing $\mathbb{E}\left[|X - \hat{X}|\right]$, so that the criterion used when there are several parameters in input is indeed a generalization of that used when there is only one parameter.

If the points $\left(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\right)$ are fixed, an approximate calculation of the minimum value of the distance in Eq. (3) and of the corresponding optimal weights $\left(w^{(1)}, \ldots, w^{(N)}\right)$ can be carried out through the following steps (see Fig. A1):

1. generate a large number $M$ (e.g. $M=10^5$) of $d$-dimensional random vectors $z_1, \ldots, z_M$ with probability distribution $f$;

2. for each vector $z_j$, select the index $i_j$ such that $z_j$ belongs to $V_{i_j}$, i.e.

$$\left|z_j - \boldsymbol{x}^{(i_j)}\right| = \min_{k=1,\ldots,N} \left|z_j - \boldsymbol{x}^{(k)}\right|;$$

3. for $k=1, \ldots, N$, assign to $\boldsymbol{x}^{(k)}$ the weight
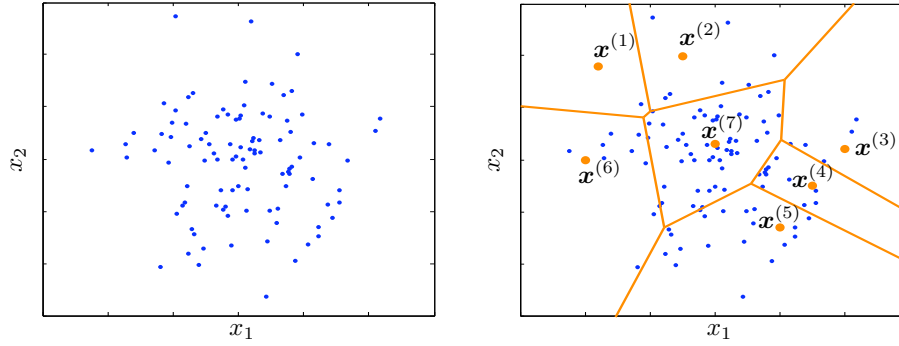
$$w^{(k)} = \frac{m^{(k)}}{M},$$

**Fig. A1.** Implementation of the stochastic quantization method with $d = 2$ and $N=7$. The blue points are a sample of the input random vector $\boldsymbol{X} = (X_1, X_2)$; the orange points $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(7)}$ are the possible values of $\hat{\boldsymbol{X}}$, which is a discrete approximation of $\boldsymbol{X}$. The orange lines define the Voronoi regions generated by the set $\left\{ \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(7)} \right\}$: the region associated to $\boldsymbol{x}^{(i)}$ contains the sample points which are closer to $\boldsymbol{x}^{(i)}$ than to any other of the orange points.

where $m^{(k)}$ is the number of vectors $\boldsymbol{z}_j$ into the Voronoi region $V_k$, i.e. the number of indexes $j$ such that $i_j = k$;

4. calculate

$$d(f, \hat{f}) = \mathbb{E}\left[ \left| \boldsymbol{X} - \hat{\boldsymbol{X}} \right| \right] \approx \frac{1}{M} \sum_{j=1}^{M} \left| \boldsymbol{z}_j - \boldsymbol{x}^{(i_j)} \right|. \quad \text{(A2)}$$

In our situation, in which the points $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)})$ are not fixed, the function

$$h\left( \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)} \right) = \frac{1}{M} \sum_{j=1}^{M} \left| \boldsymbol{z}_j - \boldsymbol{x}^{(i_j)} \right| \quad \text{(A3)}$$

must be minimized. The minimization is performed using Powell's method (Powell, 1964; Press et al., 2001), which moves the points $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}$, starting from an initial guess; for each new choice of $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}$, the algorithm evaluates $h\left( \boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)} \right)$ going through the steps 1–4 above. The set of points which produces the minimum value of $h$ is just the optimal set of points we are searching for. In order to minimize the risk of finding local minima, the minimization is repeated 10 times, varying the initial guesses, and the lowest minimum is taken as the best estimate of the true minimum.

Note that the error in the estimate (Eq. A2) of $d(f, \hat{f})$ is proportional to $\frac{1}{\sqrt{M}}$, so that, for sufficiently high values of $M$, it becomes negligible and minimizing

$$\frac{1}{M} \sum_{j=1}^{M} \left| \boldsymbol{z}_j - \boldsymbol{x}^{(i_j)} \right|$$

is the same as minimizing $d(f, \hat{f})$.

Edited by: R. Moretti

**References**

Bennett, W. R.: Spectra of quantized signals, Bell Syst. Tech. J., 27, 446–472, 1948.

Currin, C., Mitchell, T. J., Morris, M., and Ylvisacker, D.: Bayesian prediction of deterministic functions with applications to the design and analysis of computer experiments, J. Am. Stat. Assoc., 86, 953–963, 1991.

Graf, S. and Luschgy, H.: Foundations of quantization for probability distributions, Springer-Verlag, Berlin-Heidelberg, 2000.

Gray, R. M. and Neuhoff, D. L.: Quantization, IEEE T. Inform. Theory, 44(6), 2325–2383, 1998.

Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, J. Roy. Stat. Soc. B, 63, 425–464, 2001.

Neri, A., Esposti Ongaro, T., Menconi, G., De' Michieli Vitturi, M., Cavazzoni, C., Erbacci, G., and Baxter, P. J.: 4D simulation of explosive eruption dynamics at Vesuvius, Geophys. Res. Lett., 34, L04309, doi:10.1029/2006GL028597, 2007.

Oliver, B. M., Pierce, J. and Shannon, C. E.: The philosophy of PCM, P. IRE, 36, 1324-1331, 1948.

Panter, P. F. and Dite, W.: Quantizing distortion in pulse-count modulation with nonuniform spacing of levels, P. IRE, 39, 44–48, 1951.

Papale, P., Neri, A., and Macedonio, G.: The role of magma composition and water content in explosive eruptions, 1. Conduit ascent dynamics, J. Volcanol. Geoth. Res., 87, 75–93, 1998.

Papale, P.: Dynamics of magma flow in volcanic conduits with variable fragmentation efficiency and nonequilibrium pumice degassing, J. Geophys. Res., 106, 11043–11065, 2001.

Powell, M. J. D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives, Comput. J., 7, 155–162, 1964.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes in Fortran 77, Cambridge University Press, Cambridge, 2001.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P.: Design and analysis of computer experiments, Stat. Sci., 4, 409–435, 1989.

Sparks, R. S. J.: Forecasting volcanic eruptions, Earth Planet. Sc. Lett., 210, 1–15, 2003.

Todesco, M., Neri, A., Esposti Ongaro, T., Papale, P., Macedonio, G., Santacroce, R., and Longo, A.: Pyroclastic flow hazard assessment at Vesuvius (Italy) by using numerical modelling: I. Large-scale dynamics, B. Volcanol., 64, 155–177, 2002.

Valentine, G. A. and Wohletz, K. H.: Numerical models of Plinian eruption columns and pyroclastic flows, J. Geophys. Res., 94, 1867–1887, 1989.

Wand, M. P. and Jones, M. C.: Kernel smoothing, Chapman and Hall, Boca Raton, 1995.

Wilson, L., Sparks, R. S. J., and Walker, G. P. L.: Explosive volcanic eruptions: IV. The control of magma properties and conduit geometry on eruption column behaviour, Geophys. J. Roy. Astr. S., 63, 117–148, 1980.