Merging deterministic and probabilistic approaches to forecast volcanic scenarios

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

[1] *Scuola Normale Superiore, Pisa, Italy*

[2] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Italy*

[3] *Dipartimento di Matematica Applicata, Università di Pisa, Italy*

e-mail:e.peruzzo@sns.it

UNIVERSITÀ DI PISA

Istituto Nazionale di Geofisica e Vulcanologia

Abstract

We present the stochastic quantization (SQ) method for the approximation of a continuous probability density function with a discrete one. This technique reduces the number of numerical simulations required to get a reasonably complete picture of the possible eruptive conditions at a considered volcano. Finally we show the results of a test using a one-dimensional steady model of magma flow [1] as a benchmark.

Rationale 1

- The demand for eruption scenario forecast is pressing.
- Volcanic systems are largely out of direct observation.
- Some of the quantities which determine volcanic processes are uncertain.
- Taking into account these uncertainties in predictive models can be exceedingly demanding.
- We are therefore developing the application of stochastic methods to largely reduce the computational costs.

What is stochastic quantization? 2

A practical situation:

- the random vector $X = (X_1, \dots, X_d)$ is part of the input data of a numerical code φ and the random variable Y is one relevant model output;
- the probability density function $f(x)$ of X is assumed to be known;
- there is a maximum number N of affordable simulations.

Strategy ⇒ stochastic quantization method:

- find N values of X,
  $$\left(x_1^{(1)}, \dots, x_d^{(1)}\right), \dots, \left(x_1^{(N)}, \dots, x_d^{(N)}\right),$$
  and N corresponding weights, $\left(w^{(1)}, \dots, w^{(N)}\right)$, with $\sum_{i=1}^{N} w^{(i)} = 1$, so that the resulting discrete distribution is the "best" approximation of $f(x)$;
- for $i = 1, \dots, N$ compute
  $$y^{(i)} = \varphi\left(x_1^{(i)}, \dots, x_d^{(i)}\right)$$
  and give it the weight $w^{(i)}$; the resulting discrete distribution is an approximation of the unknown distribution of Y.

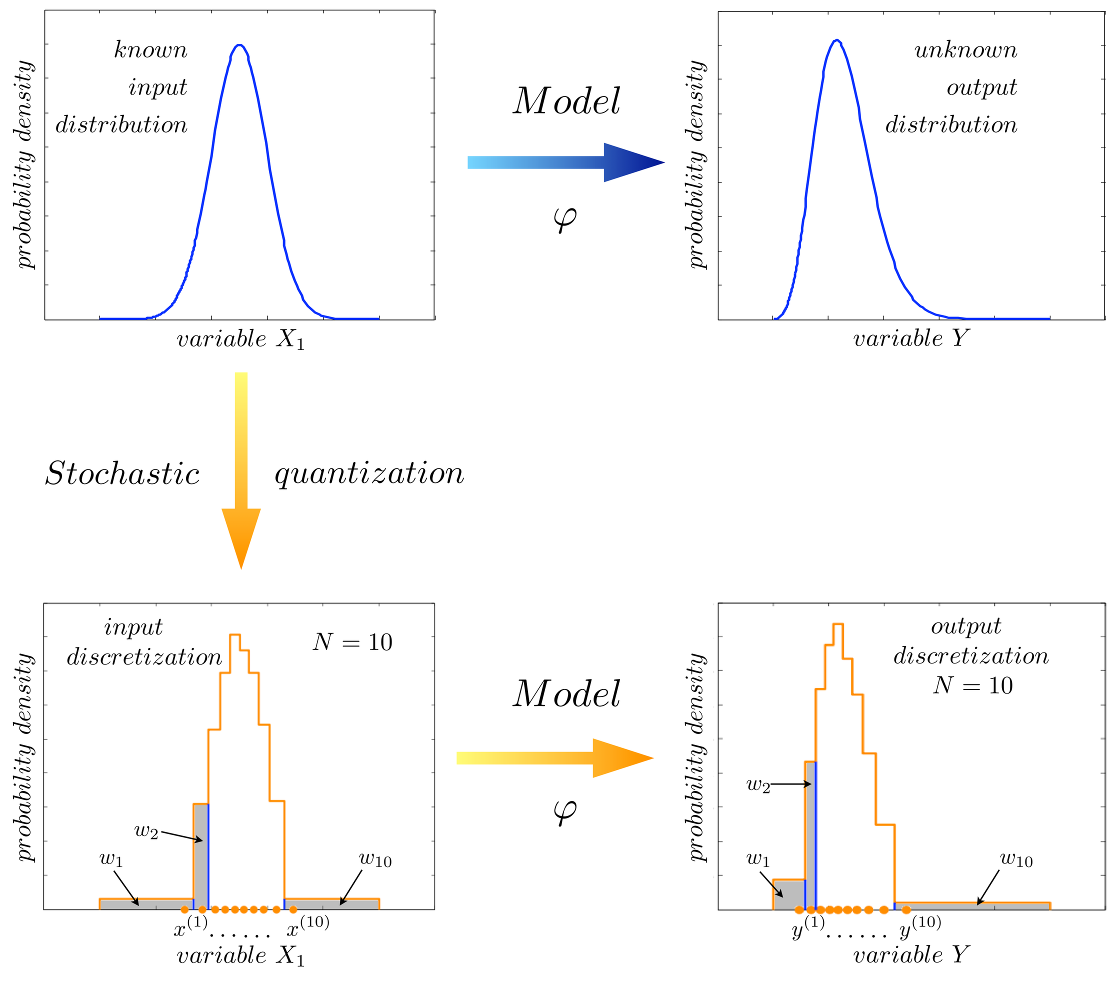


Figure 1: approximation of input and output distributions. The orange arrows represent performed computations, while the blue one represents a computation often out of reach in real situations.

The single parameter input case, d=1 3

- Introduce a distance between two probability distributions, in the case in which X is a scalar quantity: if $F(x)$ is the cumulative distribution function associated with the density $f(x)$ and $\hat{F}(x)$ is the one associated with its discretization $\hat{f}(x)$, we define the distance between f and $\hat{f}$ as follows:
  $$d(f, \hat{f}) = \int_{x_{min}}^{x_{max}} \left|F(x) - \hat{F}(x)\right| dx, \qquad (1)$$
  where x_{min} and x_{max} are the minimum and maximum possible values of X.
- The procedure consists in searching for N points
  $$\left(x^{(1)}, \dots, x^{(N)}\right)$$
  and N corresponding weights
  $$\left(w^{(1)}, \dots, w^{(N)}\right)$$
  that minimize the quantity $d(f, \hat{f})$.

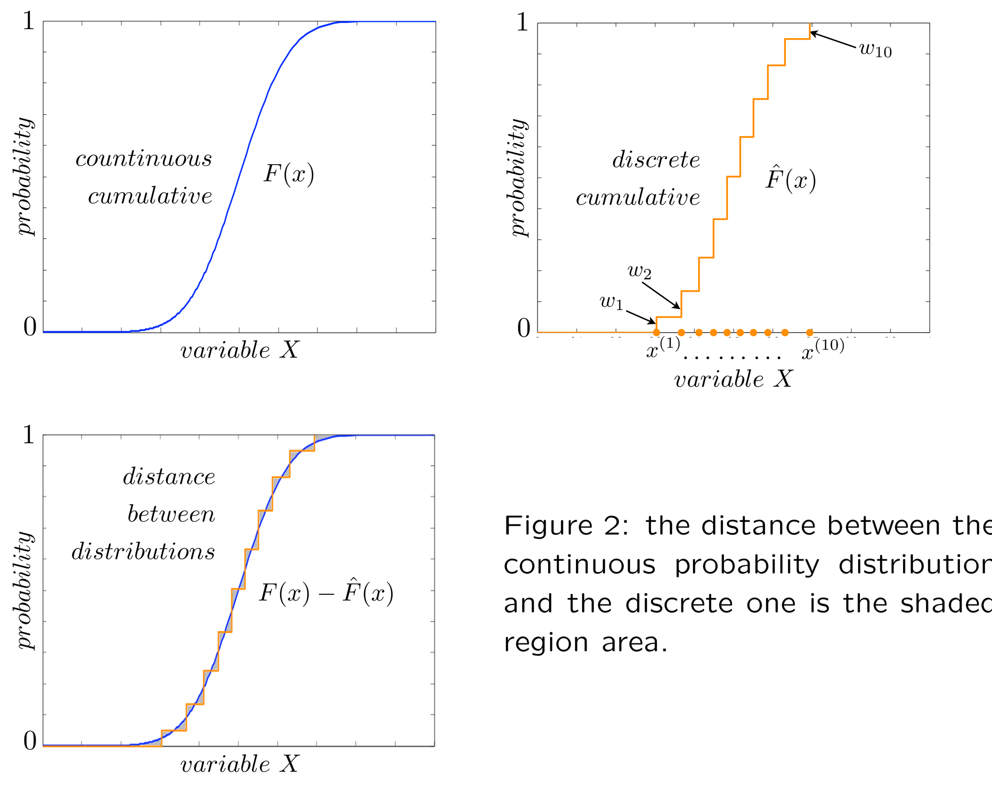


Figure 2: the distance between the continuous probability distribution and the discrete one is the shaded region area.

The multi-parameter input case, d>1 4

- When X is a d-dimensional vector quantity, a different definition of distance is more appropriate.
- Let $\hat{X}$ be a discrete random vector with probability distribution $\hat{f}$, approximating a continuous random vector X. The distance between f and $\hat{f}$ can be defined as the mean value of the error $\left|X - \hat{X}\right|$ resulting from the substitution of X with $\hat{X}$. We thus minimize
  $$d(f, \hat{f}) = \mathbb{E}\left[|X - \hat{X}|\right].$$
- It can be shown that, in the case $d = 1$,
  $$\mathbb{E}\left[|X - \hat{X}|\right] = \int_{x_{min}}^{x_{max}} \left|F(x) - \hat{F}(x)\right| dx;$$
  hence, the criterion for the multidimensional problem is a generalization of that used in the one-dimensional case.
- $d(f, \hat{f})$ is calculated through a Monte Carlo method which involves the concept of Voronoi partitions.
- The procedure consists in searching for the discrete random vector $\hat{X}$ that minimizes $\mathbb{E}\left[|X - \hat{X}|\right]$; the possible values $x^{(1)}, \dots, x^{(N)}$ of $\hat{X}$ and the corresponding weights $w^{(1)}, \dots, w^{(N)}$ generate the discrete approximation $\hat{f}$ of the density f.

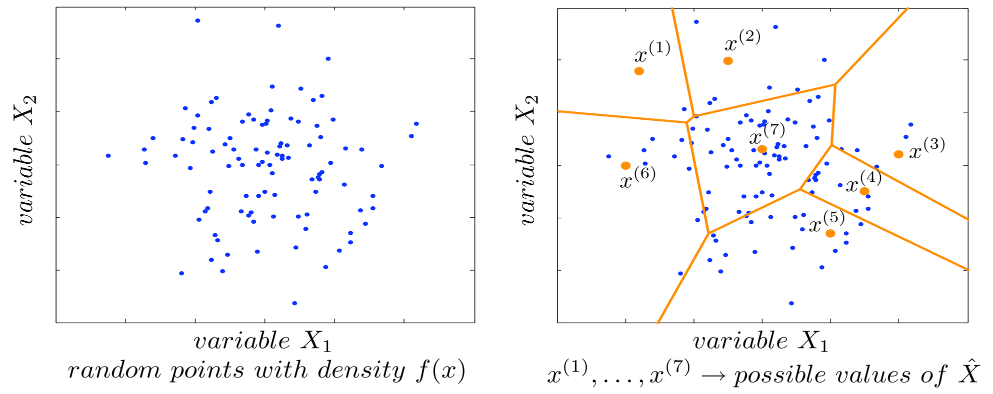
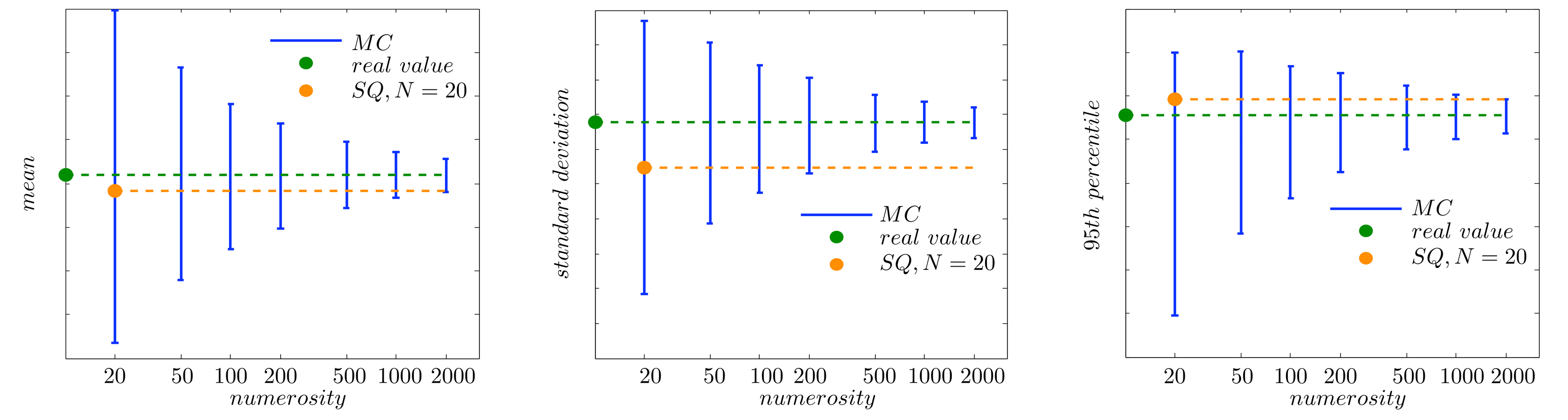


Figure 3: implementation of the multi-parameter input case. The blue points are a sample of $X = (X_1, X_2)$; the orange points $x^{(1)}, \dots, x^{(7)}$ are the possible values of $\hat{X}$, which is a discrete approximation of X. The orange lines define the Voronoi regions generated by the set of points $x^{(1)}, \dots, x^{(7)}$: the region associated to $x^{(i)}$ contains the blue points which are closer to $x^{(i)}$ than to any other of the orange points.

Testing stochastic quantization in simple cases 5



- Let $\varphi(X)$ be a known analytical function. The probability distribution of the output random variable Y can be calculated exactly and compared with the approximations produced by SQ and by Monte Carlo (MC) methods with variable numerosity.
- The case in the figure refers to
  $$\varphi(X_1, X_2) = X_1^2 X_2^2.$$

Figure 4: with the SQ method and only $N = 20$ simulations, we approximate the true values at a confidence level corresponding to $N = 2000$ MC simulations for the mean or $N = 200$ MC simulations for the standard deviation.

Application of SQ to volcanic conduit dynamics 6

- Application of SQ to a situation in which the output probability distribution cannot be explicitly calculated, but quite complete statistical information about it can be obtained through MC simulations.
- One-dimensional steady model of magma flow in a cilindrical conduit with fixed diameter and uniform temperature [1].
- Random input quantities: diameter D of the conduit and total mass fraction w_{H_2O} of water. Random output quantity: logarithm of the mass flow rate $\dot{m}$.

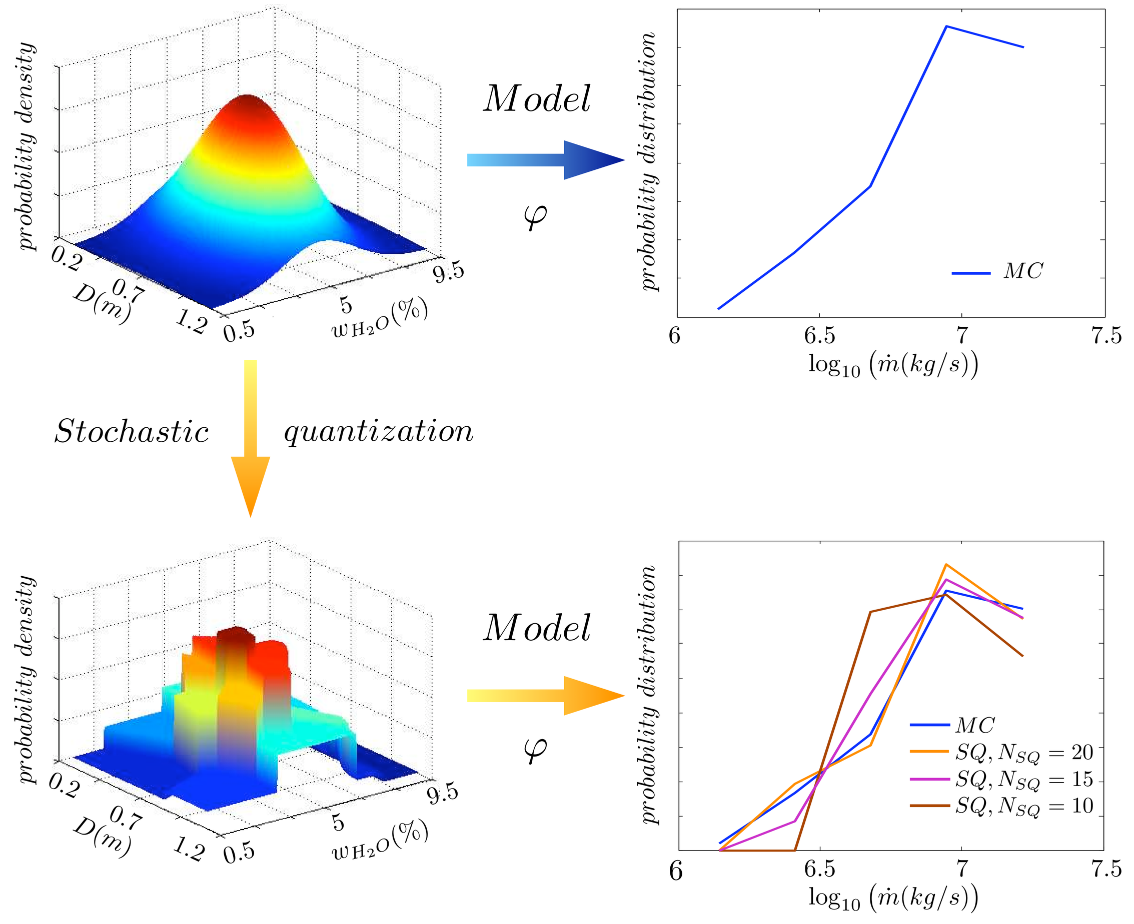


Figure 5: the correspondence between the distribution of mass flow rate found with 1000 MC simulations and SQ method is fully satisfactory when $N_{SQ} = 20$.

CONCLUSIONS

The SQ method allows the introduction of uncertainties in the deterministic approach without requiring exceeding CPU time. As a consequence, volcanic scenarios can be estimated in the future by means of complex deterministic models and taking into account the intrinsic uncertainties involved in the definition of volcanic systems.

References

[1] P. Papale, *Dynamics of magma flow in volcanic conduits with variable fragmentation efficiency and nonequilibrium pumice degassing*, J. Geophys. Res., 106, 11043-11065, 2001

[2] S. Graf, H. Luschgy, *Foundations of quantization for probability distributions*, Springer-Verlag, 2000

This poster participates in **YSOPP**

Young Scientists' Outstanding Poster Paper Contest