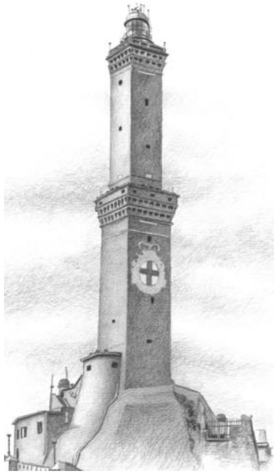


An exact framework for Uncertainty Quantification in Monte Carlo simulation

Paolo Saracco, Maria Grazia Pia
INFN Genova, Italy

CHEP 2013

Amsterdam, 14-18 October 2013

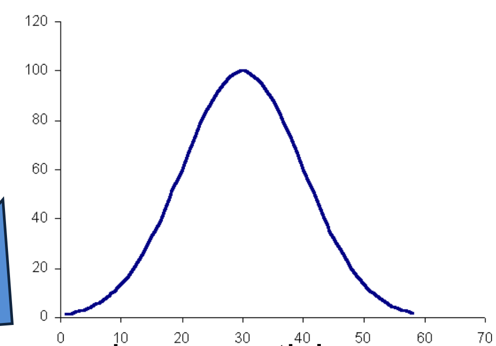
Monte Carlo in HEP

cross sections,
branching ratios,
physics models,
physics parameters..

input

MC
engine

observable



primary particles,
deposited energy…

Event generator
(Pythia, Herwig…)
Particle transport
(Geant4, MCNP, MARS…)

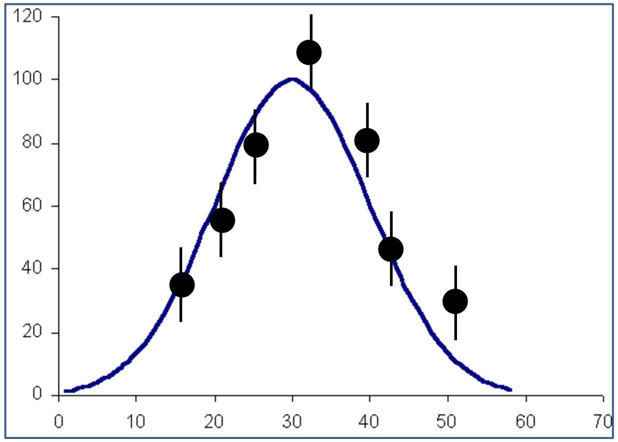
How much can we trust the observables produced by MC?

Measure and compare

Test beam

Courtesy CERN CDS, CERN-EX-0305054

Measured and simulated **observable**



If the accuracy of observable **A** is assessed by comparison with experiment, **what about observable B**?

And what about the simulation of **detector concepts**, which do not exist yet?
And observables which cannot be measured in practice?

Uncertainties

Input

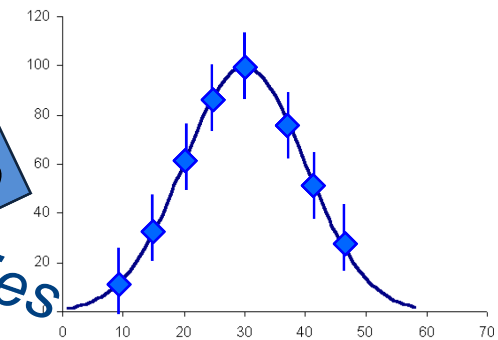
Parameter uncertainties

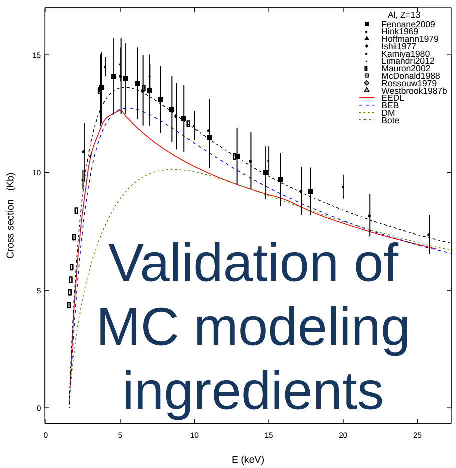
Monte Carlo method

Statistical uncertainty

observable

with uncertainties





Beware: input uncertainties can be hidden in models and algorithms in the code

Uncertainties deriving from
- input uncertainties
- Monte Carlo algorithm
- simulation model

Uncertainty quantification is the ground for **predictive Monte Carlo simulation**

Parameter uncertainties

They are the uncertainties of the **“ingredients”** of the simulation engine *(particle transport system, event generator)*

Can they be disentangled from statistical uncertainties associated with the Monte Carlo method?

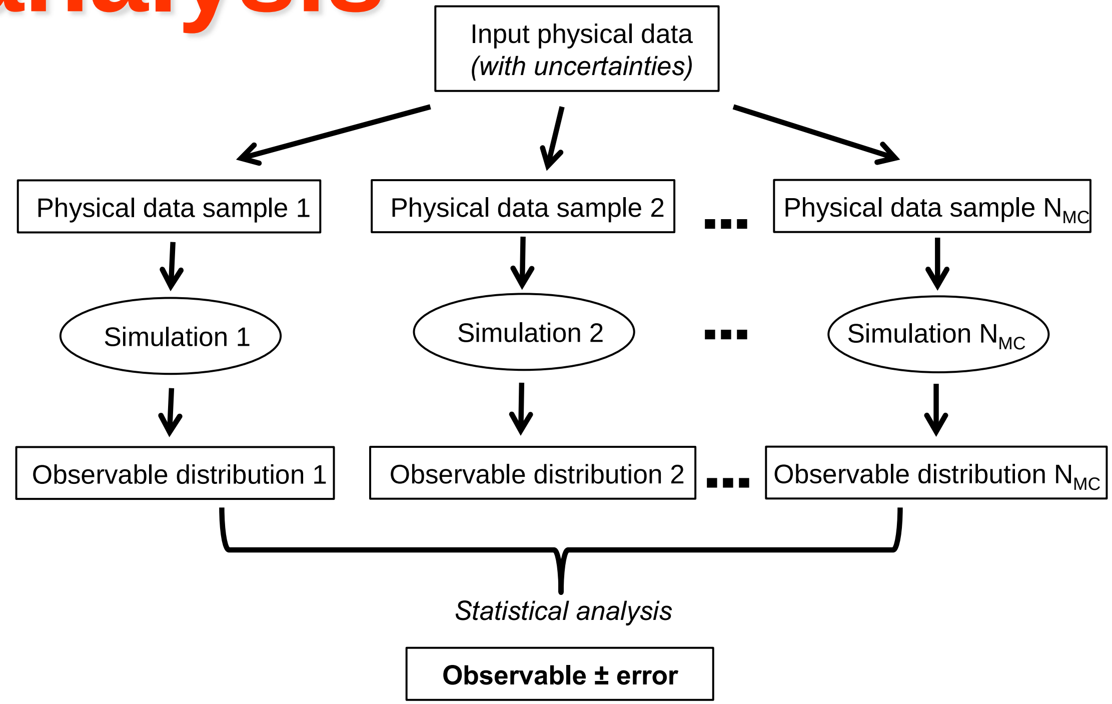
Can we estimate their effect on the observables produced by the simulation?

Sensitivity analysis

Exact calculation

Sensitivity analysis

Computational cost



To reduce computational costs, one can reduce the sensitivity analysis to the search for the most probable output
➔this means giving up a full statistical characterization of the output

- Mathematical methods
- Software toolkits *(DAKOTA, PSUADE..)*

Disentangling uncertainties I

Algorithmic (statistical) and parameter uncertainties

How do statistical and parameter uncertainties mix in a simulation environment?

EXAMPLE: a (very) simplified "transport code", a random path generator ruled by two constant parameters describing the relative probability of absorption (Σ_A) and scattering processes (Σ_S), sampling an observable – track-length in this case.

(by the way, this simulates the propagation of neutral particles in an uniform medium with constant scattering and absorption cross-sections and isotropic scattering)

If Σ_S is affected by some uncertainty, say $\Sigma_{S,\ min} < \Sigma_S < \Sigma_{S,\ max}$ we run many simulations varying its value with some known probability (for instance flat)

Disentangling uncertainties

Algorithmic (statistical) and parameter uncertainties

Results for a track length observable scored in a volume near the source

15000 simulations

(a) 10^5 events

(b) $5\ 10^5$ events

(c) 10^6 events

(d) 10^8 events

As statistical errors decrease, the distribution of the observable **is dominated by parameter uncertainties only**

Uncertainty propagation

EXPLANATION

The PDF for the result of the simulation is given by

$$G_{MC}(x) \cong \int_{-\infty}^{+\infty} d\Sigma_S\, f(\Sigma_S) \exp\left[-\frac{(x - x_0(\Sigma_S))^2}{\sigma^2_{x_0}/N} \right] \sqrt{\frac{N}{2\pi\sigma^2_{x_0}}}$$

that in the limit $N \to \infty$ becomes

$$G(x) = \int_{-\infty}^{+\infty} d\Sigma_S\, f(\Sigma_S)\, \delta(x - x_0(\Sigma_S)) = \left| \frac{d\Sigma_S(x_0)}{dx_0} \right|_{x_0 = x} f(\Sigma_S(x))$$

This last shows how an input uncertainty exactly "propagates" into the probability distribution of the output.

We must know (and invert) the "susceptivity" $x_0(\Sigma_S)$!!!
We can use MC to study this. **Lower computational cost**

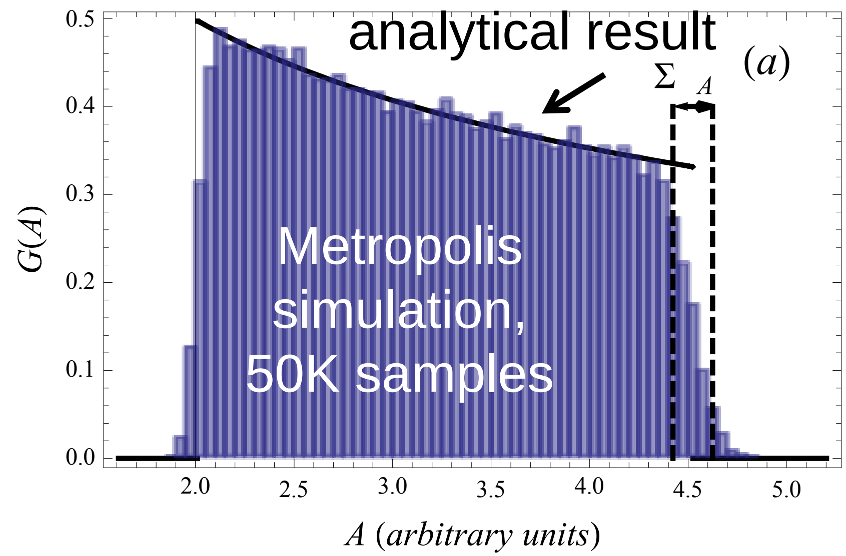
Uncertainty propagation: verification

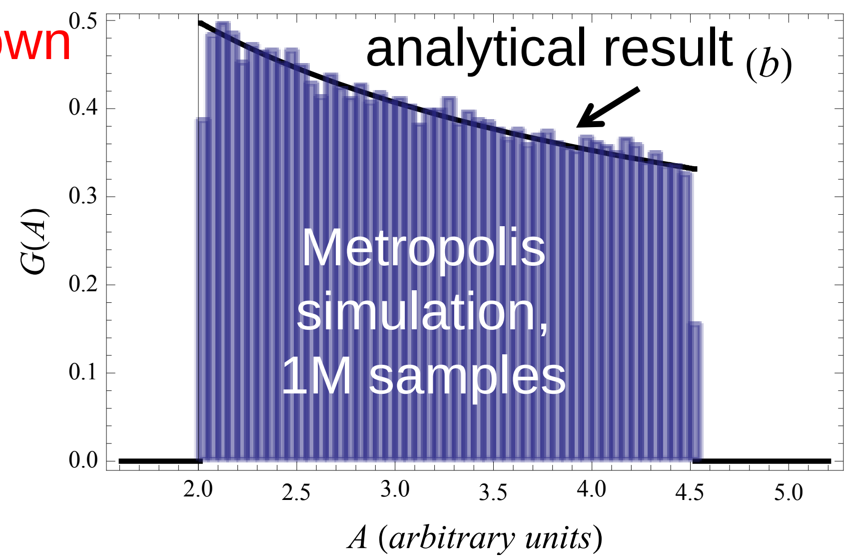
Example: evaluate the **area of the circle** with some **flat uncertainty** on the measure of its **radius**

$f(R)=\theta(R-R_{min})\theta(R_{max}-R)/(R_{max}-R_{min})$

$A(R) = \pi R^2$ the exact solution is known

$$G(A) = \frac{\theta(A - A_{min})\theta(A_{max} - A)}{2\left(\sqrt{A_{max}} - \sqrt{A_{min}}\right) \times \sqrt{A}}$$





The task of UQ

The feasibility of UQ requires:

1) To **know input uncertainties** and their probability distributions
Validation of MC modeling ingredients needed

1) To be able to **solve explicitly for G(x)**:
An exact mathematical context needed

2) To use MC simulations to **determine parameters in G(x)**:
(in the previous example, to find $A_{max/min}$ *from simulation and to determine the proper behavior* $A^{-1/2}$*)*
Possible with few simulations with predetermined accuracy

2) is independent from the features of the specific problem and can be solved within the scope of wide assumptions
this is an exact mathematical frame for UQ

UQ for a 1 parameter problem

Back to the transport example where $x_0(\Sigma_S)$ is **unknown**

We must determine the relationship $x_0(\Sigma)$ between the input parameter Σ and the (exact) result for the required physical observable x_0.

This can be done with few (21 in the case) MC simulations at fixed values of Σ_S within its range of variability.

An example of this procedure in the previous simple transport problem for observables at different distances from the source.

Note the linear relationship $x_0(\Sigma)$

Many parameters problem

In the generic case we have many input parameter unknowns:

$$G(x) = \int_{-\infty}^{+\infty} d\vec{\Sigma} f(\vec{\Sigma}) \delta(x - x_0(\vec{\Sigma}))$$

We make two “reasonable” assumptions:

- the Σ_k are independent
- $x_0(\vec{\Sigma})$ is linear (if necessary subdivide the domain of variability of the unknowns in such a way to fulfill the condition)

Under these hypothesis the evaluation of G(x) reduces to a well known problem in probability theory: the determination of the weighted sum of a certain number of independent stochastic variables.

Unfortunately even this problem is not soluble in general

Can we solve it?

Some remarks

Under these assumptions $\sigma_x^2 = \sum_k \left(\frac{\partial x}{\partial S_k}\right)^2 \sigma_k^2$ with σ_k^2 the variances of the individual input unknowns.

For M input unknowns we need *a priori* M+1 simulations to determine the values $\frac{\partial x}{\partial S_k}$, a task that can be pursued reasonably if the number of input unknowns is not so large.

So detailed physical knowledge of the problem at hand is required to select a proper set of physical parameters on which is meaningful to attempt a full Uncertainty Quantification.

We then emphasize that a full UQ is **PROBLEM SPECIFIC**

BUT we are not sure that σ_x^2 is a proper measure of the output uncertainty, since we do not know the exact form of G(x).

In some **useful selected cases** the form of G(x) is known:

- all the input unknowns are normally distributed: in such case G(x) is normal with the quoted variance
- all the input unknown are uniformly distributed: in such case a generalization of the Irwine-Hall distribution holds
- all the input unknowns have α-stable distributions with the same α value: in such case G(x) is again a stable distribution with the same α value (e.g. the Lorentz distribution)

We recently proved that a general form exists for the weighted sum of generic polynomial distributions over different intervals: this result can be used in principle to find an approximate form of G(x) with arbitrary predetermined accuracy in the general case.

A specific software tool is needed to reliably perform this step

Current scope of applicability

Single parameter uncertainty (see [1]):

- complete analysis of uncertainty propagation available
- simulation is used solely to determine the values of the parameters defining the output probability density function
- confidence intervals for the output are known with a statistical error that can be predetermined

Many parameter uncertainty (see also [2]):

- a complete UQ is possible only for independent input uncertainties.
- calculability issues may exist, in practice, if the number of parameters considered is high and/or if linearity of $x_0(\Sigma_k)$ is questionable
- confidence interval for the output are affected by the statistical errors in the determination of the required parameters AND by errors in the polynomial approximations required
- in principle a predefined accuracy can be obtained
- calculation issues must be studied

[1] - P. Saracco, M. Batic, G. Hoff, M.G. Pia – "Theoretical ground for the propagation ofuncertainties in Monte Carlo particle transport", submitted to IEEE Trans. Nucl. Phys., 2013.

[2] – P. Saracco, M.G. Pia – "Uncertainty Quantification and the problem of determining the distribution of the sum of N independent stochastic variables: an exact solution for arbitrary polynomial distributions on different intervals", submitted *to Journ. Math. Phys.*, 2013.

Validation of physics ingredients

- The validation of the physics "ingredients" of Monte Carlo codes is a *(complex, slow)* still ongoing process
  - at least regarding Monte Carlo particle transport
- Quantitative estimates of input uncertainties *(cross sections, angular distributions, BR etc.)* are necessary for uncertainty propagation
  - A qualitative plot is not enough…
- Epistemic uncertainties are often embedded in the code, without being documented
  - See M. G. Pia, M. Begalli, A. Lechner, L. Quintieri, P. Saracco, **Physics-related epistemic uncertainties of proton depth dose simulation,** *IEEE Trans. Nucl. Sci.*, vol. 57, no. 5, pp. 2805-2830, 2010

Conclusion and outlook

We have established

- A **novel conceptual approach**
- A **mathematical framework**
- **Calculation methods** for single and many parameter uncertainties

to determine the **intrinsic uncertainty** of the results of Monte Carlo simulation
(beyond statistical uncertainty)

These developments are applicable to Monte Carlo simulation in general
Particle transport
Event generators

Outlook

- Verification in a realistic experimental scenarios
- Application software system

In parallel, we pursue extensive **Geant4 physics validation**:
http://www.ge.infn.it/geant4/papers

Collaboration with HEP experiments is welcome!

Collezione di immagini utili

Don’t forget the chef

With the same ingredients we obtain very different tastes of the soup allowing the chef to modify the recipe

In a simulation context the recipe is chosen by MC user, when he defines the experimental configuration: geometry, composition, external conditions, …

→ MC user is the chef ←

For the purpose of UQ different experimental configurations – if not VERY similar – must be analyzed as different problems

The “simulation engine” is the CODE together with the CHOSEN EXPERIMENTAL CONFIGURATION