

# UQ

## Uncertainty Quantification in Monte Carlo simulation

Maria Grazia Pia

*INFN Sezione di Genova*

[Matej Batic], [Gabriela Hoff], Paolo Saracco

Collaborators:

Fondazione Bruno Kessler, MPI HLL, XFEL,  
UC Berkeley, State Univ. Rio de Janeiro, Hanyang Univ. (Korea)

INFN CCR

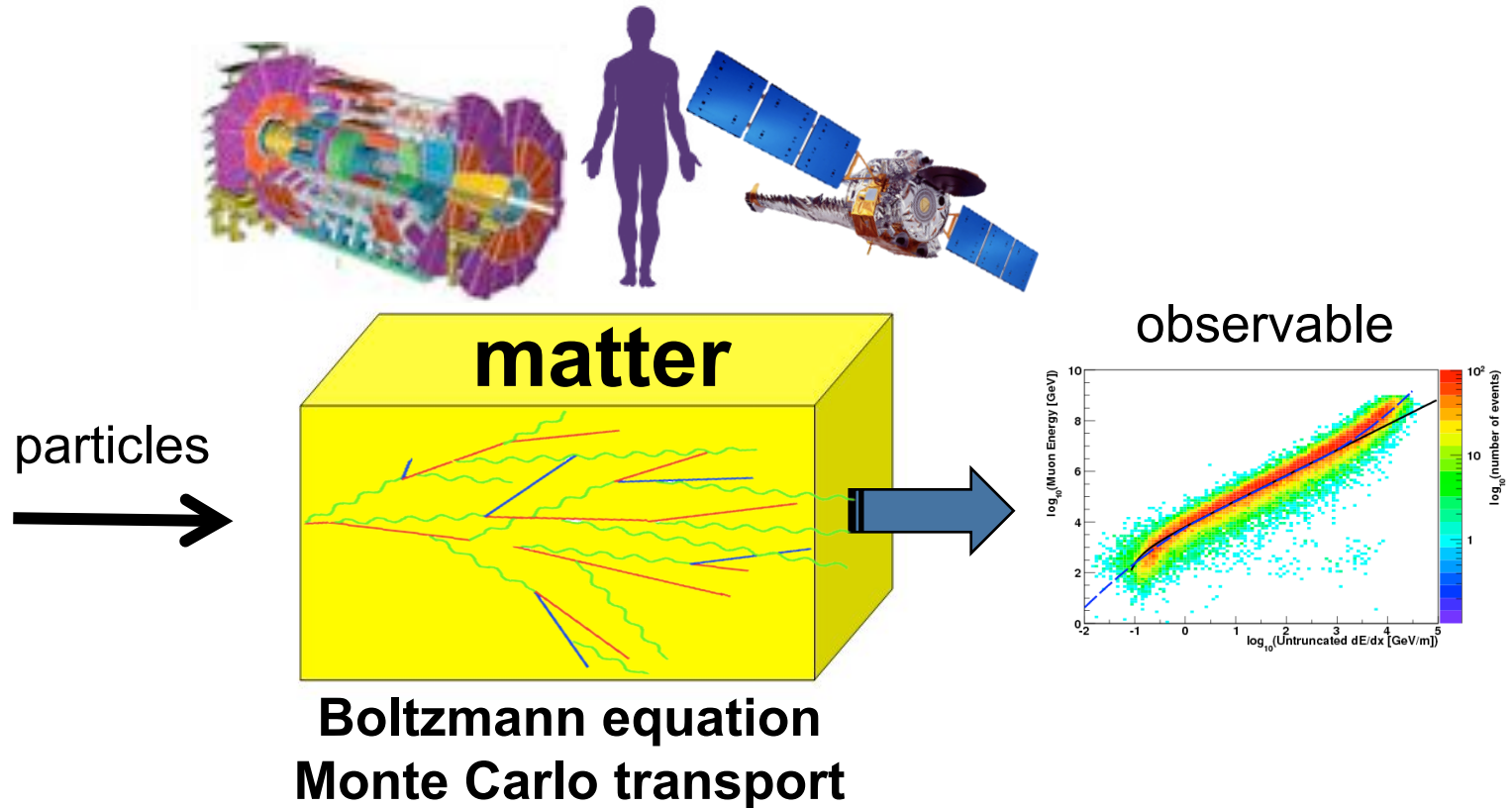
24-26 October 2012



# Simulation validation

- Geant4 was first released on 15 December 1998
- Geant4 reference paper (2003) is the **most cited INFN publication**, and INFN is the institute contributing the largest number of citations
  - Large user community worldwide, multi-disciplinary applications
- What fraction of Geant4 has been validated to date?
- By whom?
  - Geant4 kernel
  - Use cases
- Uncertainty quantification in simulation (not only Monte Carlo) is a major research topic these days
  - Essential to critical applications

# A predictive instrument



How do the uncertainties in the simulation model affect the observables?  
Can we quantify the accuracy of the simulation output,  
given the uncertainties in its “ingredients”?

# UQ

## Uncertainty quantification in Monte Carlo simulation

**Theoretical foundation**

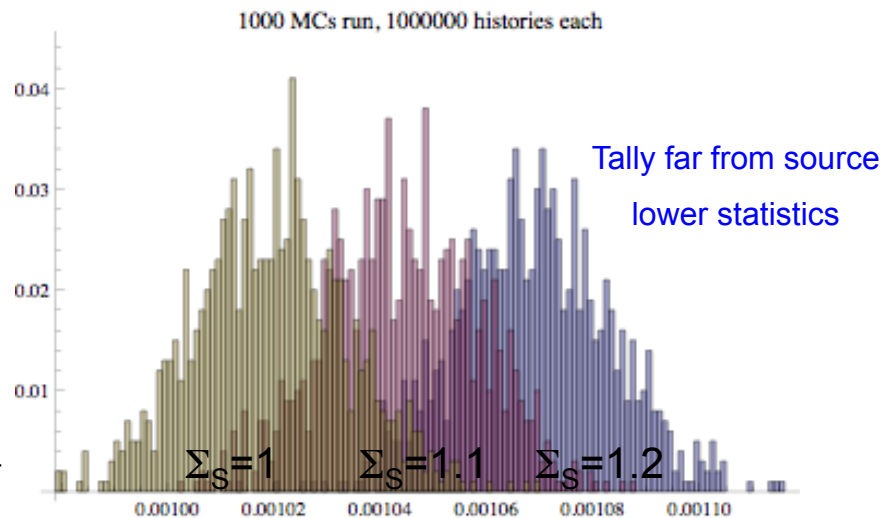
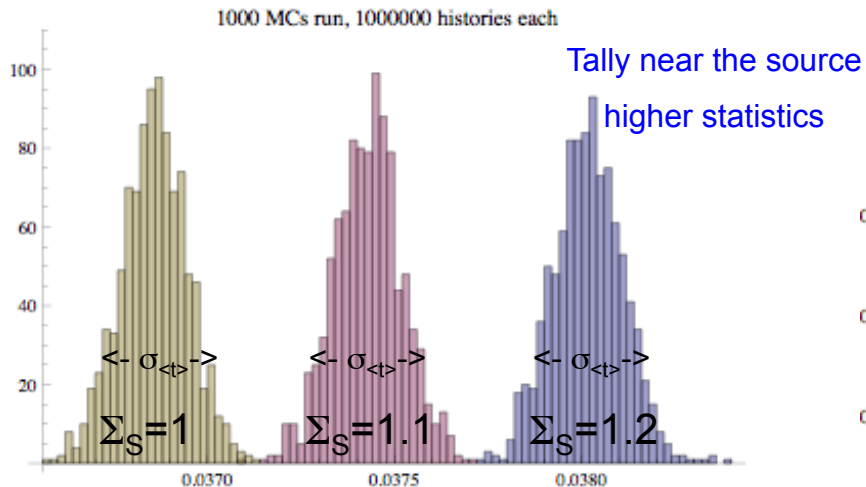
**Geant4 validation**

Identification of **epistemic uncertainties**  
in the simulation

**Physics models** for particle transport  
**Software design** features enabling UQ  
**Statistical and mathematical methods** supporting UQ

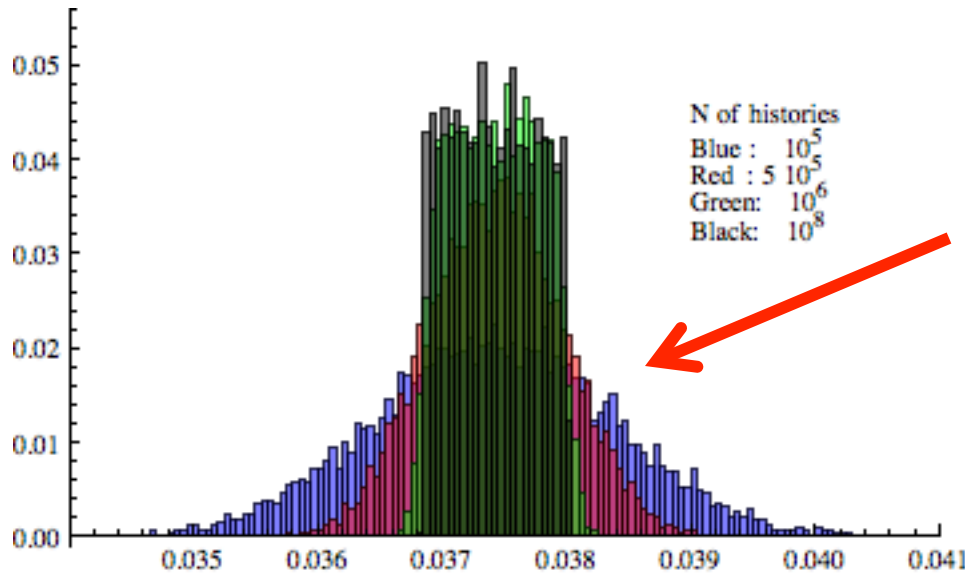
# Theoretical foundation

- Studied in a simple set-up
  - A sphere
  - Probability (i.e. macroscopical cross section  $\Sigma_S$ ) assigned to each physics process considered (scattering, absorption etc.)
  - Tallies (i.e. observables) scored in given positions
- What happens to the tallies, when  $\Sigma_S$  varies?



The Central Limit Theorem assures that we can reduce the widths, by increasing the number of histories in each MC:  $\sigma_{\langle t \rangle} \cong \sigma_t / \sqrt{N}$

# Uncertainty transfer



Uncertainty in  $\Sigma_S$

$1.0 < \Sigma_S < 1.2$  (flat distribution)

**transfers into the tally**  
(+ some noise)

If the number of generated events is “not sufficient”, one may mistake the tally distribution as a gaussian

This derives from the identity

$$\lim_{\sigma \rightarrow 0} \sqrt{\frac{1}{2\pi\sigma^2}} \int_{-\infty}^{\infty} p(x_0) \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] dx_0 = p(x)$$

The noise can be reduced by increasing the number of generated events



# Impact

Accurate knowledge of uncertainties in the “ingredients” of Monte Carlo simulation allows the quantification of the uncertainties of the observables produced by the simulation

## **Simulation with predictive power**

### **Validation of the “ingredients”**

Heuristic proof

Higher statistics study foreseen

Still a long way to go... multiple “ingredients”, correlations, multiple tallies etc.

# Validation of Geant4 “ingredients”



Validation of Geant4  
photon interactions

Photoelectric effect:  
total cross section

Preliminary results at NSS 2012

IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 59, NO. 4, AUGUST 2012

Epistemologically correct validation

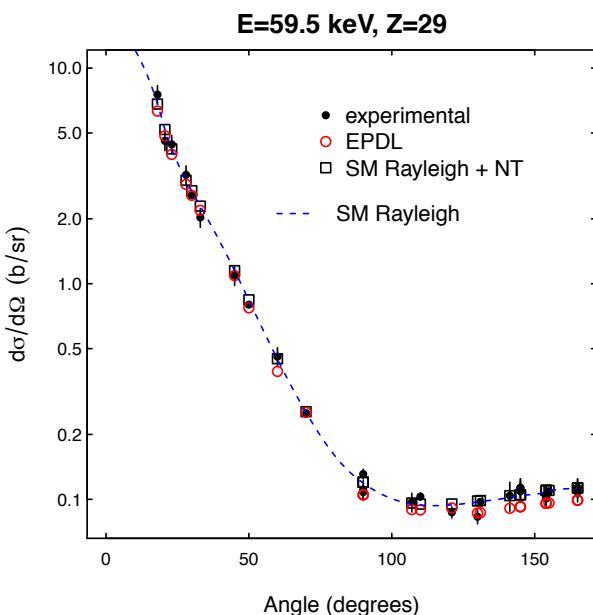
## Photon Elastic Scattering Simulation: Validation and Improvements to Geant4

Matej Batič, Gabriela Hoff, Maria Grazia Pia, and Paolo Saracco

TABLE VIII

$\chi^2$  TEST OUTCOME: TEST CASES COMPATIBLE WITH EXPERIMENT AT 0.01 SIGNIFICANCE LEVEL

Test		Penelope 2001	Penelope 2008-2011	EPDL	EPDL ASF	SM	RF	NF	MF	MF ASF	RF ASF
all	Test cases	71	71	71	71	71	71	71	71	71	71
	Pass	19	27	27	18	55	18	25	35	37	34
	Fail	52	44	44	53	16	53	46	36	34	37
	Efficiency	0.27	0.38	0.38	0.25	0.77	0.25	0.35	0.49	0.52	0.48
	Error	$\pm 0.05$	$\pm 0.06$	$\pm 0.06$	$\pm 0.05$	$\pm 0.06$	$\pm 0.05$	$\pm 0.06$	$\pm 0.06$	$\pm 0.06$	$\pm 0.06$
$\theta \leq 90^\circ$	Test cases	67	67	67	67	67	67	67	67	67	67
	Pass	19	27	27	18	55	18	25	35	36	32
	Fail	48	40	40	49	12	49	42	32	31	35
	Efficiency	0.28	0.40	0.40	0.27	0.82	0.27	0.37	0.52	0.54	0.48
	Error	$\pm 0.05$	$\pm 0.06$	$\pm 0.06$	$\pm 0.05$	$\pm 0.05$	$\pm 0.05$	$\pm 0.06$	$\pm 0.06$	$\pm 0.06$	$\pm 0.06$
$\theta > 90^\circ$	Test cases	17	17	17	17	17	17	17	17	17	17
	Pass	1	1	1	1	10	1	1	0	2	4
	Fail	16	16	16	16	7	16	16	17	15	13
	Efficiency	0.06	0.06	0.06	0.06	0.59	0.06	0.06	< 0.06	0.12	0.24
	Error	$\pm 0.06$	$\pm 0.06$	$\pm 0.06$	$\pm 0.06$	$\pm 0.12$	$\pm 0.06$	$\pm 0.06$		$\pm 0.08$	$\pm 0.10$



Geant4 radioactive decay: collaboration con  
Darmstadt Tech. Univ. + XFEL

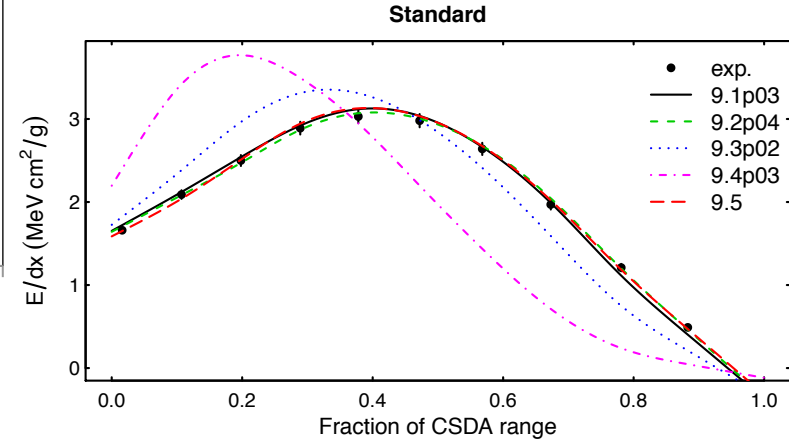
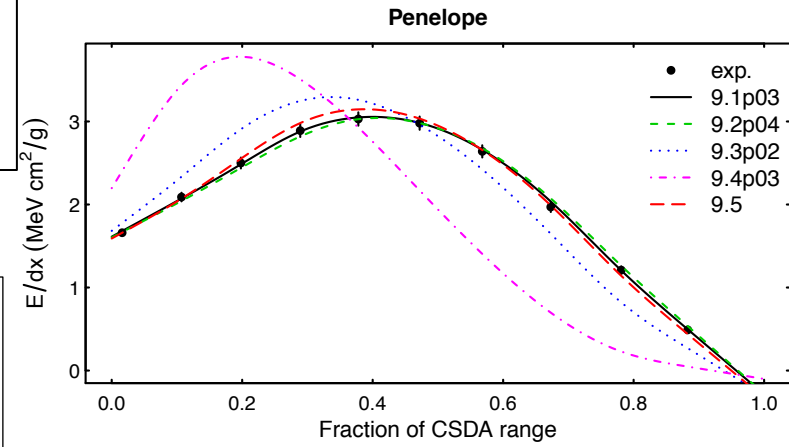
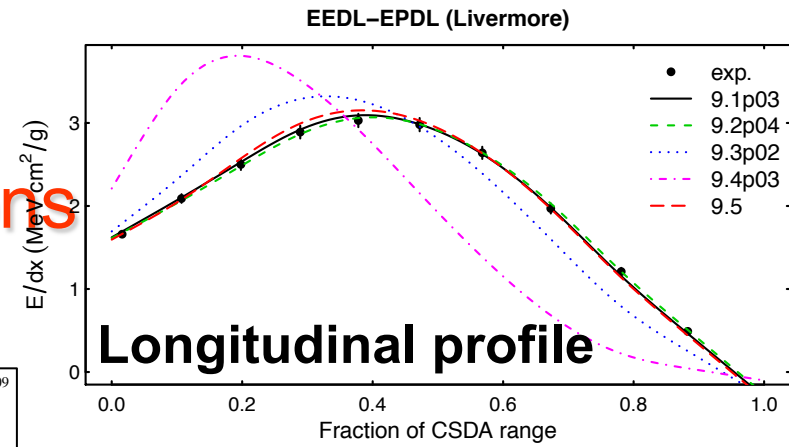
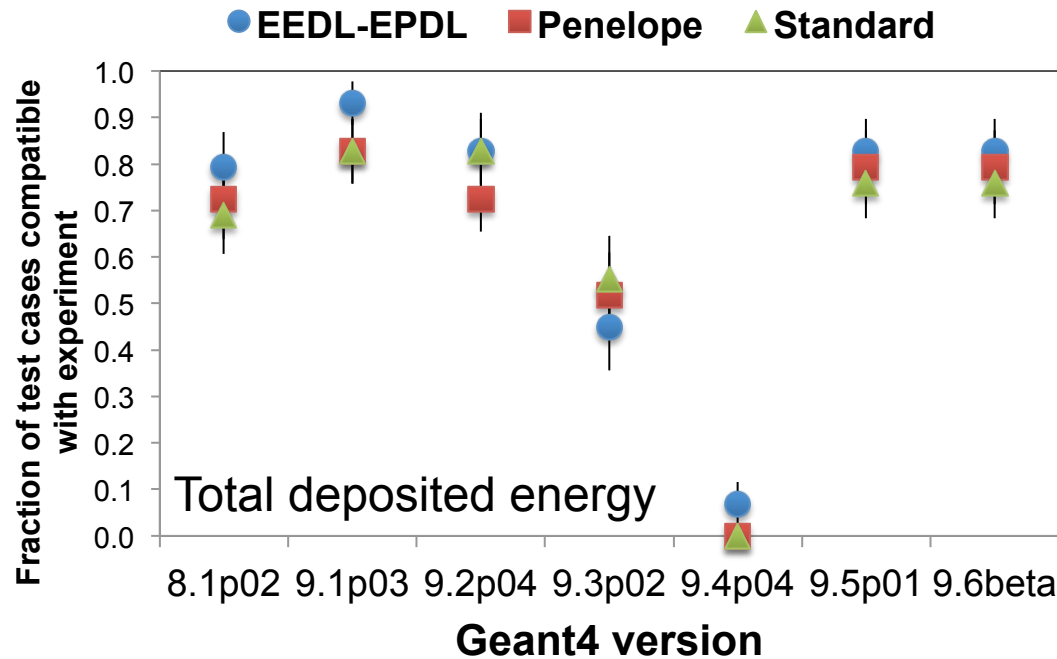


# Validation of Geant4 energy deposition by low energy electrons

398 IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 56, NO. 2, APRIL 2009

Validation of Geant4 Low Energy Electromagnetic Processes Against Precision Measurements of Geant4 Electron Energy Deposition

8.1p02, 9.1 Anton Lechner, Maria Grazia Pia, and Manju Sudhakar



# Publications

- M. Batic, G. Hoff, M. G. Pia, P. Saracco  
**Photon elastic scattering simulation: validation and improvements to Geant4**  
*IEEE Trans. Nucl. Sci.*, vol. 59, no. 4, pp. 1636-1664, Aug. 2012
- 2 papers on Geant4 radioactive decay to be submitted in the next days

# CHEP 2012

- M. G. Pia et al.,  
**Refactoring, reengineering and evolution: paths to Geant4 uncertainty quantification and performance improvement**
- M. G. Pia, T. Basaglia, Z. W. Bell, P. V. Dressendorfer  
**Publication patterns in HEP computing**
- M. Batic, G. Hoff, M. G. Pia, P. Saracco,  
**Algorithms and parameters for improved accuracy in physics data libraries**
- M. Batic, G. Hoff, M. G. Pia,  
**Precision analysis of Geant4 condensed transport effects on energy deposition in detectors**
- M. Batic et al.  
**A new development cycle of the Statistical Toolkit  
(Winner of Best Poster selection)**

*Proc. CHEP (Computing in High Energy Physics ) 2012,  
J. Phys. Conf. Series, 2012*

# NSS 2012

- Paths to Geant4 Evolution: Refactoring, Reengineering and Physics
- State-of-the-Art Simulation of Photon Interactions with Matter
- Precision Analysis of Electron Energy Deposition in Detectors Simulated by Geant4
- Code and Papers: Computing Publication Patterns in the LHC Era
- *Physics Data Libraries: Content and Algorithms for Improved Monte Carlo Simulation*
- *New Developments of the Statistical Toolkit*

# Training

- Lecture on “Refactoring – Dealing with legacy code” at Advanced Programming Concepts School, DESY, Oct. 2012
- Geant4 Refresher Course, IEEE NSS 2012
- Invitation to organize a Geant4 Short Course, IEEE NSS 2013, Seoul, Korea

# Uncertainty Quantification in generic Monte Carlo simulations

---

*P. Saracco<sup>(1)</sup>, M. Batic<sup>(2)</sup>, G. Hoff<sup>(3)</sup>, M.G. Pia<sup>(1)</sup>*

*<sup>(1)</sup>I.N.F.N. sez. Genova - Italy*

*<sup>(2)</sup>I.N.F.N. sez. Genova (Italy) & Jožef Stefan Institute (Slovenija)*

*<sup>(3)</sup>I.N.F.N. sez. Genova (Italy) & P.U.C.R.S. (Brasil)*

What UQ means?

Which UQ can be done in MC simulations?

Best practices for UQ in MC simulation

Conclusions, suggestions & outlook

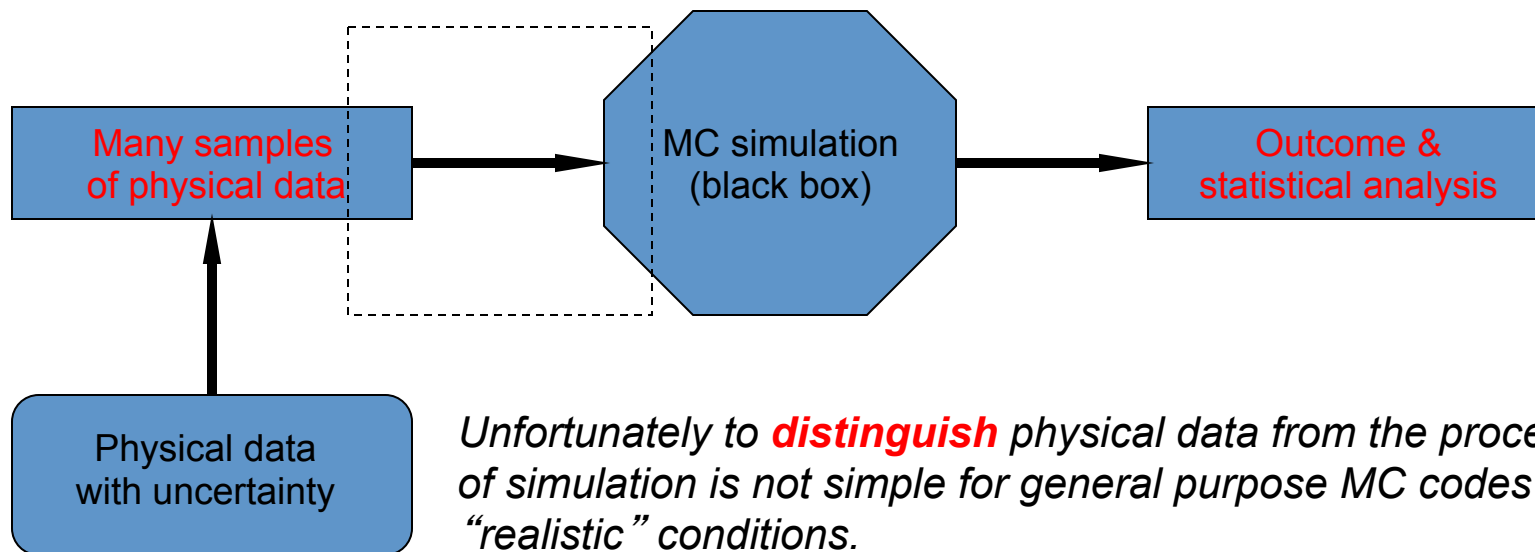
---

Nuclear Science Symposium, Medical Imaging Conference 2012 - Anaheim, CA, USA

*“the lack of complete certainty, that is the existence of more possibilities. The true outcome/state/result cannot be known”*

Usually in such a situation one asks to know the probabilities of the possible outcomes:  
is this the final goal of UQ in a (Monte Carlo) simulation environment?

In a “true” (or “realistic”) MC simulation many different aspects of the problem interact, but we roughly imagine something like



*Unfortunately to **distinguish** physical data from the process of simulation is not simple for general purpose MC codes in “realistic” conditions.*

Moreover “many samples” has its counterpart in the time duration of the simulation in realistic conditions (it depends strongly on the complexity of the specific problem at hand). **So our first purpose is to simplify the simulation context to extract infos useful as a guidance to analyze more complex situations. (this is a STARTUP project...)**

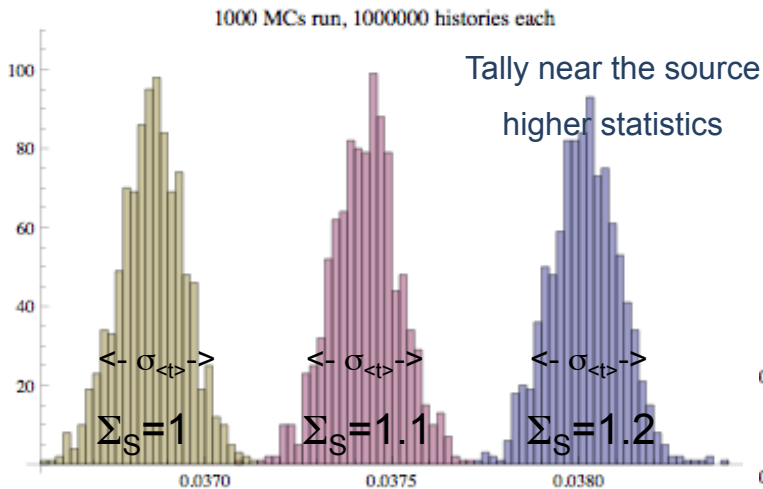
## A (very) simple Monte Carlo Game

So we imagine a very simplified situation: a probability, i.e. a macroscopical cross section, is directly assigned to each physical process considered (scattering, absorption, ...) in a single geometry (spherical): simulation essentially corresponds to build a random path from these assigned probabilities.

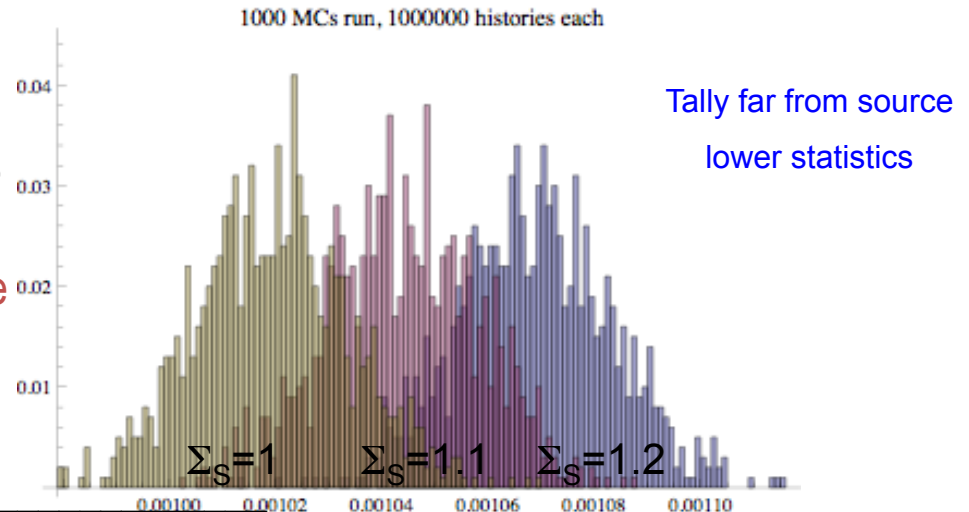
Tallies are then chosen in some given positions.

We can easily run this simulation varying the physical parameters:

by varying, for example,  $\Sigma_S$  we obtain, for a flux tally in different (radial) positions, different outcomes (mean time for history  $\approx 250 \mu\text{sec}$ , depending on the number of tallies recorded).



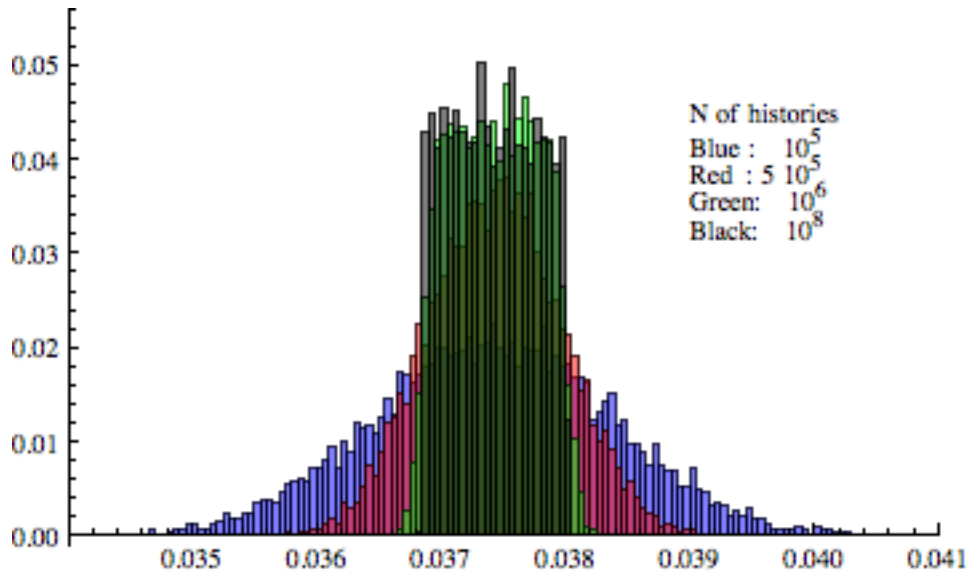
This is expected and almost obvious



The Central Limit Theorem assures that we can reduce the widths, by increasing the number of generated events in each MC:

$$\sigma_{\langle t \rangle} \cong \sigma_t / \sqrt{N}$$





When  $\Sigma_S$  varies continuously in the same interval we can appreciate directly the effect of epistemic uncertainties on a MC simulation: here we ran  $10^5$  MCs with  $1.0 < \Sigma_S < 1.2$  (flat distribution).

If the number of histories  $N$  is “not sufficient” the outcome can be confused with a gaussian distribution with a larger standard deviation than the one resulting from a single MC run.

But as the number of histories in each MC increases the outcome become closer and closer to a flat distribution (green and black histograms) - the same form we assumed for the variability of the physical data: then we conclude naively that (details later on):

**MC simulation “simply” transfers the original (form of the) epistemic uncertainty in the final outcome adding some statistical “noise”.** I anticipate that this derives from the identity

$$\lim_{\sigma \rightarrow 0} \sqrt{\frac{1}{2\pi\sigma^2}} \int_{-\infty}^{\infty} p(x_0) \exp\left[-\frac{(x - x_0)^2}{2\sigma^2}\right] dx_0 = p(x)$$

and then this “statistical noise” can be reduced in principle arbitrarily by increasing the number of histories run in each MC.

Let us see how this works: if we run many N-histories MCs varying the physical parameter(s) - here is simply  $\Sigma_S$  - in some interval with probability  $f(\Sigma_S)$  the final distribution of (any given) tally will be

$$G(x) = \int_{-\infty}^{+\infty} f(\Sigma_S) \exp\left[-\frac{(x - x_0(\Sigma_S))^2}{2\sigma_{x_0}^2 / N}\right] \sqrt{\frac{N}{2\pi\sigma_{x_0}^2}} d\Sigma_S$$

where we finally integrated over the possible values of the physical parameter(s) with their (unknown) probabilities.

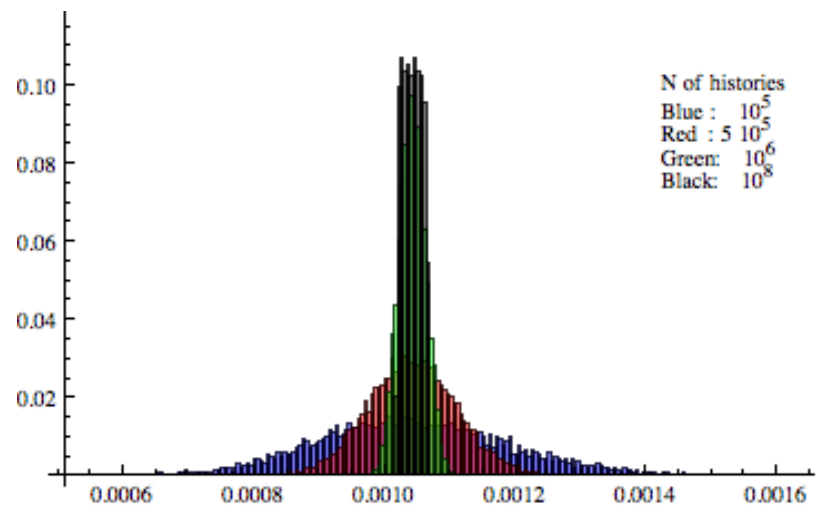
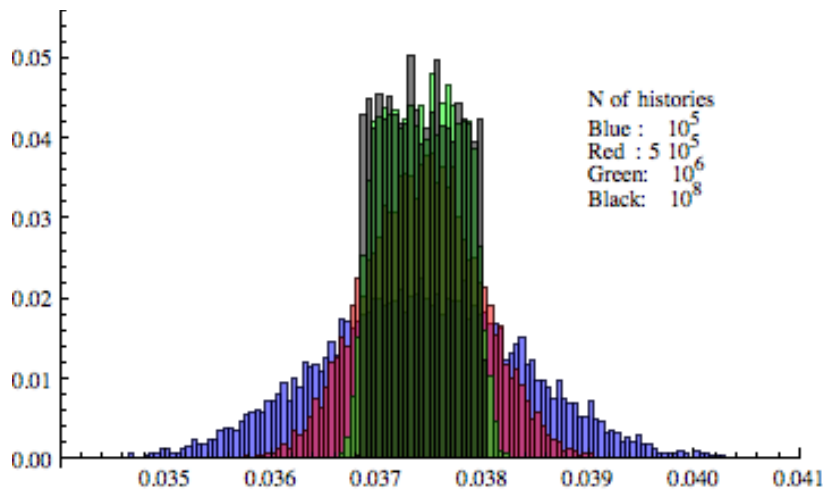
**This hypothesis derives from the CLT if N is sufficiently large.**

We made the seemingly obvious assumption that it exists a (invertible) function  $x_0(\Sigma_S)$  relating the “exact” tally mean value with the original value of the physical parameter that corresponds to this result: this assumption relies on the interpretation we give to the process of simulation as a “surrogate” for the solution of the Boltzman transport equation. The tally mean is a (statistical) estimate for the particle flux/density which (obviously) is a function of the physical parameters: if their range of variability is not too large (and if we are not too much unlucky) this function is also invertible. It is “easy” (with some subtleties) that in the limit of infinite histories we expect something like

$$G(x) = K \left| \frac{d\Sigma_S(x)}{dx} \right| f(\Sigma_S^{-1}(x))$$

If the range of variability of  $\Sigma_S$  is sufficiently small  $x(\Sigma_S)$  can be approximated by a linear relation, i.e. the final distribution of tallies has the same form of the original “epistemic uncertainty” (with some displacement and stretch).

**This is an heuristic proof -----> be careful !!!!**



By varying the position of the tally far from the source the same effect is observed, but we possibly need to enhance the number of histories run in each MC to see it clearly; the relevant parameter to look at is the ratio between the single MC standard deviation and the range of variability of the tally itself as it can be determined by the last equations:

for any given range of the physical parameters  $\Sigma_{\min} \leq \Sigma_S \leq \Sigma_{\max}$  the process of simulation transforms it in an interval of variability for the tally  $x_{\min} \leq x \leq x_{\max}$ .

This interval clearly depends on the choice we made for the tally.

It is required clearly that - for any single MC run -  $\sigma_{\langle t \rangle} \ll x_{\max} - x_{\min}$ : this is the practical translation of the limit for “infinite histories” in previous equations.

**When required this condition can be met by increasing the number of histories run, since for each fixed parameters MC:  $\sigma_{\langle t \rangle} \sim \sigma_t N^{-1/2}$  for sufficiently large N.**

**These observations open the possibility to define some heuristical “good practices” to follow for UQ in MC simulations.**

## Good practices for UQ in realistic problems

It should be clear that in a realistic situation the final uncertainty depends on a wide set of physical parameters: for example in a scattering experiment one should consider the dependence from beam uncertainties, target CS, detector geometry/composition and more.

So it is better to reduce, as far as possible, the complexity of the problem at hand by decomposing it in simpler ones, each depending on a smaller subset of physical parameters. Then composition rule for conditional probabilities can be invoked.

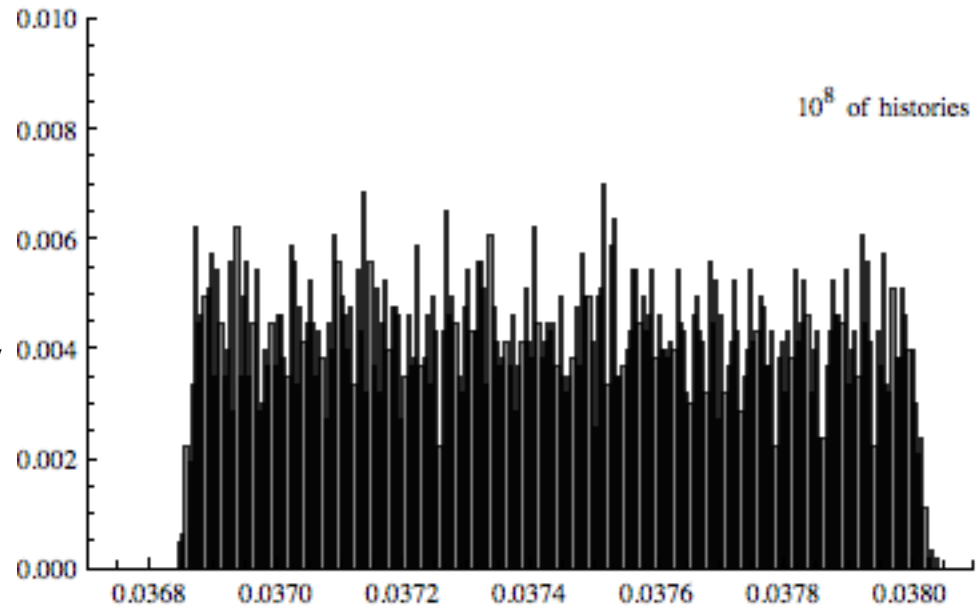
Then one should perform an “**interval analysis**” on each of the (relevant) parameters involved in each of the subprocesses: on the basis of the previous results and if the conditions under which they are valid - a fact that if necessary can be explored - one simply need to determine the interval of uncertainty for the (intermediate) simulation **by examining with an accuracy higher as needed which interval is spanned by the result of the simulation: to do this one need in principle two single MC run for each of the parameters to be investigated.**

I stress that it is not required to analyze in detail the probability distribution of the outcome of each simulation in the cascade - a task with very high CPU time requirements - because we (suppose to) know how the input probability distribution transfers in the output one. **The final (expected) distribution is the (linear) map of the original probability distribution of physical data (thata is an epistemic uncertainty) onto the found interval.**

At the end to obtain the UQ for the realistic problem one must combine all these results by using repeatedly the laws of probability composition.

Having outlined the procedure, let us give some more details: since we are going to examine all these features in a statistical context, it is better to go in more deep insight on some critical aspects.

For the first tally we considered, very near to the particle source (.1 cm far) we plot again the final distribution for the tally means alone: this plot was built by  $\approx 7000$  runs each with  $10^8$  histories: this amounts to some 20000 hours of CPU, in a very simple geometry and with a single homogeneous material. This means one week with 100 CPU farm.



**It is clear from this plot “alone” that we have not a satisfactory reproduction of a flat distribution over an interval: there are tails at the extrema and a wide variability in the “flat” zone.**

- tails at the extrema depend (as we have seen) on the number of histories run
- variability in the “flat” zone depends on the number of MCs we are able to run

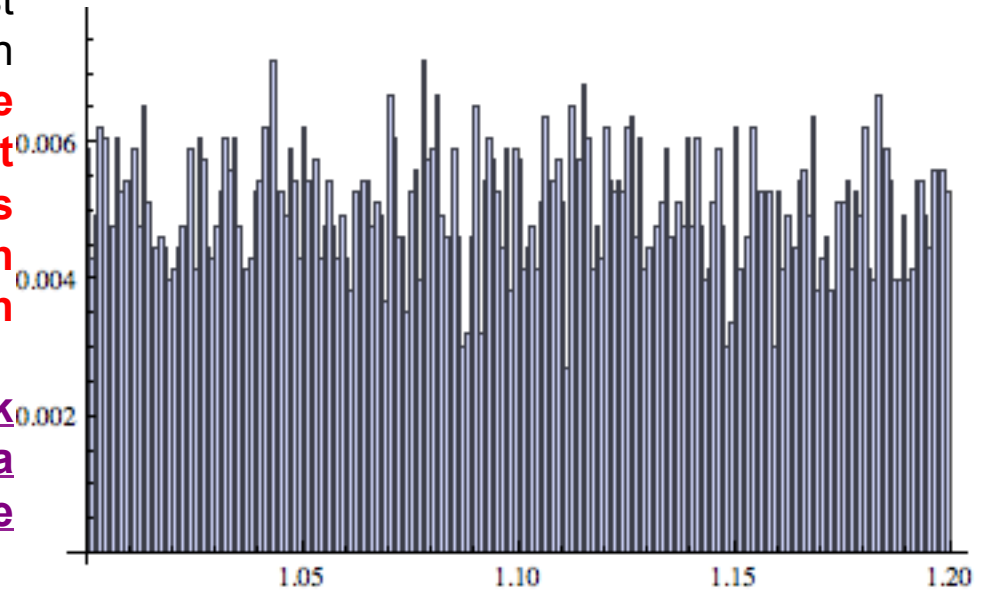
In this case we have  $x_{<}=0.036846$ ,  $x_{>}=0.0380403$ ,  $3\sigma \approx 2.7 \cdot 10^{-4}$  then  $x_{\max}-x_{\min} \approx 1.10^{-3}$ ; then it is clear on the basis of previous argument why we still see some tail at the extrema. We should reduce  $\sigma$  at least by a factor 10, which means  $10^{10}$  histories !!!

On the other hand because we have chosen a 250 bins histogram we have in mean  $\approx 30$  MCs in each bin: even this number is too much low to expect a truly “flat” distribution (this depends also on the fact that the sample of the  $\Sigma$  we extracted from a flat distribution is not a “good sample” ---> we would need more and more MCs).

In conclusion we should multiply the CPU time required to statistically test our result by a factor 1000 or more, in fact **confidence analysis against the hypothesis of flat distribution is not satisfactory:  $p=0.03$ , but this number is rapidly increasing with the number of histories (and with the number of MCs run).**

Then the job of statistically check case by case what happens in a realistic simulation is a no-hope task.

*The sample of the  $\Sigma_S$  from the flat distribution*



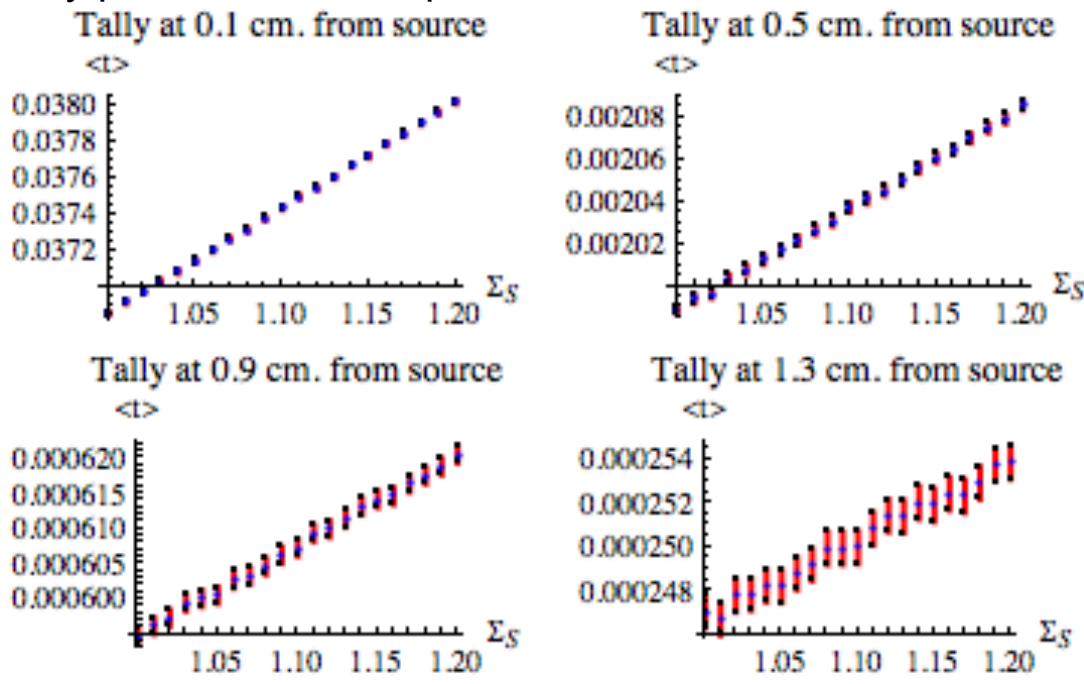
**BUT THIS IS NOT A JOB FOR MC USERS, because theory assures us about the final outcome (even if we, better our CPU farm, are working to give it better statistical evidence).**

**Other simpler checks can, if case, be performed**

In the present case, in particular, we have chosen a rather high cross section variability (20%) and a tally near to the source to enhance tally statistics: both of these conditions help making the resulting effect more evident and clear.

But, looking in detail to the plot, one could question about the linearity of the relation  $x(\Sigma)$ : if this relation is not linear, simulation could transfer input uncertainty into the final result in a more complicated way than expected, for instance distorting the flat (input) distribution of the  $\Sigma_t$  into a non flat output distribution for tally means.

Here we show the (sampled) relation  $x(\Sigma)$  with a confidence interval at 1/1000, for various tally positions with respect to the source. Here we run  $10^9$  histories for various values of  $\Sigma_S$



to check if  $x(\Sigma)$  is really a linear relation, at least in this case.

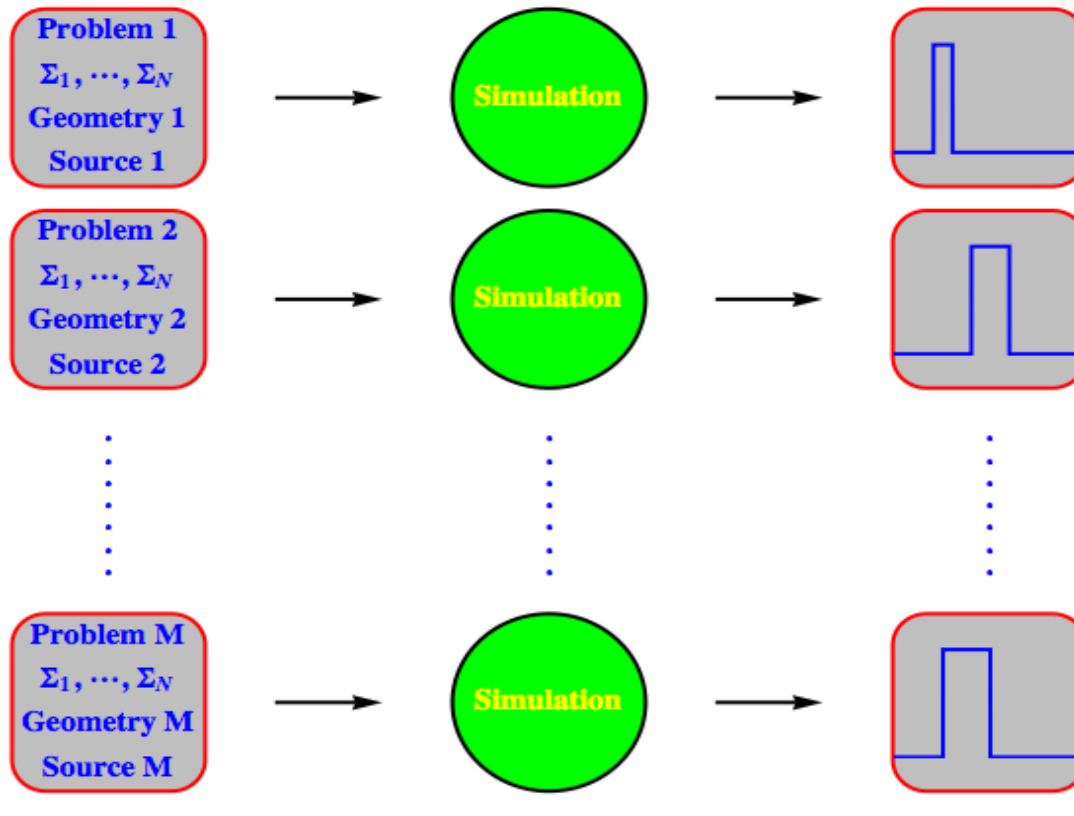
We remark that **position of the tally** with respect to the source can **influence** the quality of the sample: we must always remember that even if “theory is exact”, results we are trying to assess about UQ come from sampling the exact results in a statistical environment.

## Conclusions (I)

As we said before **MC simulation “simply” transfers the original epistemic uncertainty in the final outcome adding some statistical “noise”**: this seems quite obvious, if not even tautological, but “a priori” we did not know how this really happens.

We have shown that under some (wide and verifiable) conditions this is predictable;

If we simulate different problems depending on the same set of physical parameters, for



instance with different geometries/sources, we expect something like it is shown here aside (if original uncertainties are assumed to be flat distributions), without any other distortions on the functional form of the final uncertainty (here we neglected for simplicity the noise added by simulation)



## Conclusions (II)

### What MC projects should do:

Implement in MC codes the possibility for users to handle/modify easily cross-sections and other physical parameters subject to (epistemic) uncertainty.

### What UQ projects should do:

Enhance this modelization of UQ, if possible providing quantitative estimates.  
Give detailed probabilistic description of the variability of physical data

### What MC users should do:

Use as better as possible our suggested “good practices” to quantify the uncertainties in realistic problems.

Be careful when splitting the original problem in simpler ones: optimum cannot practically reached !

Keep care of correlations because not all physical parameters involved in a given realistic problem are really independent.

(employ more time and more people in this job)