

# Propagation of uncertainties and applications in numerical modeling: tutorial

Dominique Barchiesi, Thomas Grosjes

## ► To cite this version:

Dominique Barchiesi, Thomas Grosjes. Propagation of uncertainties and applications in numerical modeling: tutorial . Journal of the Optical Society of America. A, Optics and image science, Optical Society of America, 2017, 34 (9), pp.1602-1619. <<http://doi.org/10.1364/JOSAA.34.001602>>. <10.1364/JOSAA.34.001602>. <hal-01575512>

HAL Id: hal-01575512

<https://hal.inria.fr/hal-01575512>

Submitted on 2 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Propagation of uncertainties and applications in numerical modeling: tutorial

DOMINIQUE BARCHIESI<sup>1,\*</sup> AND THOMAS GROSGES<sup>1</sup>

<sup>1</sup>Group for Automatic Mesh Generation and Advanced Methods (Gamma3 UTT-INRIA), University of Technology of Troyes - French National Institute for computer science and applied mathematics, 12 rue Marie Curie - CS 42060 - F-10004 Troyes Cedex, France

\*Corresponding author: dominique.barchiesi@utt.fr

Compiled July 11, 2017

Some inputs of computational models are commonly retrieved from external sources (handbooks, articles, dedicated measurements) and therefore are subject to uncertainties. The known experimental dispersion of the inputs can be propagated through the numerical models to produce samples of outputs. The stemming propagation of uncertainties is already significant in metrology but has also applications in optimization and inverse problems resolution of the modeled physical system. Moreover, the information on uncertainties can be used to characterize and compare models, and to deduce behavior laws. This tutorial gives tools and applications of the propagation of experimental uncertainties through models. To illustrate the method and its applications, we propose to investigate the scattering of light by gold nanoparticles that also enables the comparison of the full Mie theory and of the dipole approximation. The position of the localized surface plasmon resonance and the corresponding value of the scattering efficiency are more specifically studied. © 2017 Optical Society of America

**OCIS codes:** (080.1753) Computation methods; (100.3190) Inverse problems; (290.4020) Mie theory; (000.4430) Numerical approximation and analysis; (310.6805) Theory and design; (000.5490) Probability theory, stochastic processes, and statistics; (050.1755) Computational electromagnetic methods.

<http://dx.doi.org/10.1364/josa.XX.XXXXXX>

## 1. INTRODUCTION

The basic idea of the propagation of experimental uncertainties through models comes from the handling of uncertainties in experiments. The Guide to the Expression of Uncertainty in Measurement [1–3] and its companion [4] are the founding references of the scientific exploitation of uncertainties [5]: “It is now widely recognized that, when all of the known or suspected components of error have been evaluated and the appropriate corrections have been applied, there still remains an uncertainty about the correctness of the stated result, that is, a doubt about

how well the result of the measurement represents the value of the quantity being measured.” and therefore, the uncertainty of measurement is defined as a “parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand” [1] (section (B.2.18)). Actually, in the case of the comparison between experimental results and theoretical ones, the overlap of uncertainty intervals obtained from experiments and numerical simulations is the necessary condition to assess their agreement [6]. Indeed, some inputs of numerical simulations come from experimental data (the optical properties for instance) and therefore are subject to uncertainties whose influence deserves to be taken into account to refine the discussion on the theoretical results. As uncertainties of model inputs can be considered as the result of a random phenomenon, propagating a random sample of inputs through the numerical model produces an output sample. The purpose of this tutorial is to show how a sample of output numerical results, produced from statistical distributions of the inputs, can be processed. More specifically, models and numerical simulations are very helpful to explain physical phenomena, to deduce behavior laws, to optimize systems and to solve the inverse problems. In this tutorial, these applications of the propagation of uncertainties are addressed.

This tutorial is based upon the propagation of distributions using a Monte Carlo method described in Ref. [2]. No commercial software is required. All results of this tutorial are obtained with softwares under opensource government: GNU Octave [7] and GNU R. The paper is organized as follows: the second section introduces the basis of the propagation of uncertainties and its applications. The third section is devoted to the numerical and physical discussions of the results. For this, the example of the scattering of light by gold nanoparticles is considered. The results of the propagation of uncertainties are used to compare models, to optimize the modeled physical system, to solve the inverse problems and to deduce behavior laws, before concluding. The algorithms of the method of propagation of uncertainties are given in appendix. Glossary of notations may help the reader.

## 2. THE PROPAGATION OF UNCERTAINTIES THROUGH NUMERICAL MODELS

Methods for handling of uncertainties are used by experimentalists to determine significant digits of a measurement, especially in the case of successive measurements of the same measurand,

carried out under the same conditions of measurement [1, 4]. From statistical processing of the measured data, coverage intervals can be deduced. Undoubtedly, the evaluation of uncertainties is valuable in experiments [5]. Conversely, the numerical model is often evaluated only once by using either a specific value or an estimation of the mean value of each of the inputs. However, the experimental uncertainties of the inputs of a model can also be used in numerical simulations. Indeed, repeated evaluations of the model by using the random generation of the inputs in accordance to their uncertainty, produces samples of output results, that can be processed in the same way as any result of repeated experiments. The following subsections introduce the basis of the propagation of uncertainties through analytical and numerical approaches. The concept of coverage intervals is introduced. The use of an output sample of the numerical model for comparison of models, for applications in optimization and in inverse problems are also described, as well as in the possible deduction of behavior laws. Indeed, a deterministic link between outputs and inputs can emerge from the random results.

### A. Analytical approach of the propagation of uncertainties: sensitivity coefficients

Let us consider an *analytical* model  $\mathcal{M}_a(\mathcal{P}, \mathcal{X})$  which depends on  $N$  continuous inputs  $\mathcal{I}$  that are parameters  $\mathcal{P}$  and variables  $\mathcal{X}$ . The analytical model enables to calculate the partial derivatives with respect to each input  $\mathcal{I}_i$ , and therefore to produce an analytical formula of the uncertainty  $u(\mathbf{o})$  on the output quantity  $\mathbf{o} = \mathcal{M}_a(\mathcal{I})$ . This approach is directly inspired by that of the indirect measurement (Eq. (16), p. 21 of Ref. [1]):

$$u^2(\mathbf{o}) = \sum_{i=1}^N \left( \frac{\partial \mathcal{M}_a}{\partial \mathcal{I}_i} \right)^2 u^2(\mathcal{I}_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[ \rho(\mathcal{I}_i, \mathcal{I}_j) \underbrace{\left( \frac{\partial \mathcal{M}_a}{\partial \mathcal{I}_i} \right) \left( \frac{\partial \mathcal{M}_a}{\partial \mathcal{I}_j} \right)}_{\text{sensitivity coefficients } c_i \text{ and } c_j} u(\mathcal{I}_i) u(\mathcal{I}_j) \right], \quad (1)$$

where  $u(\mathcal{I}_i)$  and  $u(\mathcal{I}_j)$  are the uncertainties of each input  $\mathcal{I}_i$  and  $\mathcal{I}_j$  and  $\rho(\mathcal{I}_i, \mathcal{I}_j)$  is the estimated Pearson's correlation coefficient satisfying  $\rho(\mathcal{I}_i, \mathcal{I}_j) = \rho(\mathcal{I}_j, \mathcal{I}_i)$  and  $-1 \leq \rho \leq 1$ . This coefficient is also called the product-moment correlation coefficient [1]:

$$\rho(\mathcal{I}_i, \mathcal{I}_j) = \frac{\sum (\mathcal{I}_i - \mu(\mathcal{I}_i))(\mathcal{I}_j - \mu(\mathcal{I}_j))}{(\sum (\mathcal{I}_i - \mu(\mathcal{I}_i))^2)^{1/2} (\sum (\mathcal{I}_j - \mu(\mathcal{I}_j))^2)^{1/2}}, \quad (2)$$

where  $\mu(\bullet)$  is the mean value of the data, and the sum is achieved on all the considered data. If the inputs  $\mathcal{I}_i$  and  $\mathcal{I}_j$  are *statistically* independent,  $\rho(\mathcal{I}_i, \mathcal{I}_j) = 0$ , but the converse is not necessary true. Evaluations of standard uncertainty components are founded either on frequency distributions of the repeated measurements (Type A), while Type B evaluations are founded on *a priori* distributions of probability using information from references, manuals, technical data sheets, or independent characterization of the measurement apparatus.

The numerical approach of the propagation of uncertainties is an alternative method if no analytical model is available, or if the partial derivative cannot be easily calculated [2]. It is based on a Monte Carlo method to determine the size of sample that ensures a control of the accuracy of the output sample characteristics (mean value, standard deviation and coverage interval).

### B. Numerical approach of the propagation of uncertainties

Let us consider a *numerical* model  $\mathcal{M}_n(\mathcal{P}, \mathcal{X})$  which depends on the input parameters  $\mathcal{P}$  and on the variables  $\mathcal{X}$ . We suppose that inputs  $\mathcal{I} = (\mathcal{P}, \mathcal{X})$  are subject to uncertainties that are supplied from experimental sources. For example,  $\mathcal{P}_i$  is supposed to follow a gaussian (or normal) distribution of probability, with mean value  $\mu(\mathcal{P}_i)$  and standard deviation  $\sigma(\mathcal{P}_i)$ . Therefore, the random generation of  $M$  values of  $\mathcal{P}_i$  allows to compute a sample of output values for the same variable  $\mathcal{X}$ . Therefore, the mean value  $\mu(\mathbf{o})$  and the standard deviation  $\sigma(\mathbf{o})$  of the output sample are similar to those obtained from repeated measurements [1, 2]. The size  $M$  of the sample is determined as follows.

#### B.1. Quantities of interest and sample size

The propagation of the inputs uncertainties through the model produces a sample of output results. Therefore, the output sample  $\mathbf{o}$  is characterized by its estimated mean value  $\mu(\mathbf{o})$  and its estimated standard deviation  $\sigma(\mathbf{o})$ . The standard deviation measures the dispersion of possible outputs around the mean value. Moreover coverage intervals can be deduced from the sample [2]. The coverage interval  $CI_{p\%}^c = [\mathbf{o}_{low}; \mathbf{o}_{high}]$  contains a specified proportion of the outputs, given an arbitrary chosen coverage probability  $p$  (see algorithm 3 in appendix A). Reference [2] (section (7.7)) states that the boundaries of the coverage interval are the minimum and the maximum of the  $M \times p$  values of the sorted sample ( $M \times p$  is the product of the size  $M$  of the output sample  $\mathbf{o}$  by the coverage probability  $p$ , algorithm 3, appendix A). The lower and the upper bounds of the interval  $CI_{p\%}^c$  ( $\mathbf{o}_{low}$  and  $\mathbf{o}_{high}$  respectively) as well as the estimated mean value  $\mu(\mathbf{o})$  and standard deviation  $\sigma(\mathbf{o})$  depend on the realization of the random drawing of the inputs. In Ref. [2], the minimum size of the sample  $M$  is the maximum integer number in  $\{10^4; 100/(1-p)\}$ . For example,  $p = 0.95$  leads to a recommended sample size of  $10^4$ . Nevertheless, the delivery of such quantity of numerical results may require time-consuming computation. Therefore, we prefer to determine the size  $M$  of the sample from both the number of significant digits of the input quantities, retrieved from experimental a priori knowledge, and from the length of their intervals of variation [6]. The sample size  $M$  must be greater than the maximum of the ratios of the lengths of intervals of variation, to the smallest intervals that can be measured (discrimination threshold or finite resolution), for any parameter or variables.

The dispersion of the mean value of outputs ( $\mu(\mathbf{o})$ ), of the standard deviation ( $\sigma(\mathbf{o})$ ) and of the boundaries of the coverage interval ( $\mathbf{o}_{low}$  and  $\mathbf{o}_{high}$ ) must be characterized to deduce the retained significant digits for each of these quantities. Therefore, the statistical study of a sufficient number of output samples must be carried out. In fact, each of these quantities which are deduced from each sample, itself becomes a random number since it varies from one sample to another. Reference [2] proposes an adaptive procedure to achieve a given target accuracy of these quantities. We draw inspiration from that Monte Carlo method: algorithms 1-3 describe the procedure (appendix A). The method is based on the repeated generation of  $h$  samples of outputs (each of them having size  $M$ ), until a given precision on the mean value, the standard deviation of the mean value and on the boundaries of the coverage interval is reached. The aim is to reach a given number of significant digits in these quantities. Actually the repeated achievements of the propagation of uncertainties supplies sets of  $h$  mean values, standard deviations and boundaries of the coverage intervals. The mean value and the

standard deviation of the  $h$  mean values, the standard deviation and the boundaries of the coverage interval of each sample are evaluated. The standard deviations of each of these estimated values (from the  $h$  samples), relatively to the mean values are:  $\mathbf{s}_\mu$ ,  $\mathbf{s}_{y_{low}}$  and  $\mathbf{s}_{y_{high}}$ . The standard deviation for the merged outputs (a single sample of size  $M \times h$ ) is  $\mathbf{s}_{u(O)}$ .

At each step  $h$  of the Monte Carlo loop, an additional sample of size  $M$  is computed. The loop stops when the maximum  $\mathbf{s}_y$  of the relative standard deviations is lower than the chosen target accuracy  $\mathbf{u}_r(\mathbf{o})$  (i.e. 1%). The stopping criterion is [2]:

$$\mathbf{s}_y = \max(\{\mathbf{s}_\mu, \mathbf{s}_{u(O)}, \mathbf{s}_{y_{low}}, \mathbf{s}_{y_{high}}\}) \leq \mathbf{u}_r(\mathbf{o}), \quad (3)$$

At the end of the loop, the number of computed samples is  $\max(h)$ . The merged output data  $\mathcal{O}$  of size  $\max(h) \times M$  can therefore be processed as well as each sample of size  $M$ .

### B.2. Normality test and coverage intervals

Intervals about the result of computations can be defined. They may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the output value. The coverage interval  $CI_{p\%}^c$  is a characteristic of a given sample. Indeed, it contains a specified proportion of the outputs, given an arbitrary chosen coverage probability  $p$ . But in the case of sample of small size, its reliability may be questionable. Therefore, a complementary analysis may be of interest. It consists in testing the null-hypothesis  $H_0$  that the sample comes from a normal distribution of probability (for example) by using non parametric tests of equality of probability distributions (or "goodness-of-fit") [8]. Therefore the question is: "is it reasonable to make the assumption that  $\mathbf{o}$  is randomly distributed following a normal distribution?" If the answer is yes, an interval of confidence at level  $p$  can be deduced.

In a normality test of data (or statistical hypothesis testing), data  $o_i$  are tested against the null hypothesis that it is normally distributed. The test gives a probability value  $p_{test}$ . For examples, the Lilliefors test [9] which is a variant of the Kolmogorov test [10] with estimated mean and variance from the sample, considers the maximum distance between the cumulative distribution functions (CDF). The CDF is also used by the Anderson-Darling test [11] whereas the Shapiro-Wilk test [12] uses the estimate of variance. The Shapiro-Wilk is classically considered as the most powerful especially in the case of small sample size. The Pearson test ( $\chi^2$ ) gives a probability value  $p_{test}$  of deviations between observed and theoretical frequencies. The normalized sum of squared deviations follows a  $\chi^2$  distribution, under the null hypothesis. [8, 13].

The probability value  $p_{test}$  is compared to a significance level  $\alpha$  (arbitrary chosen). The null hypothesis is either:

- rejected if  $p_{test} < \alpha$ , and there is evidence that the output sample  $\mathbf{o}$  was not from a normally distributed population;
- or cannot be rejected if  $p_{test} \geq \alpha$ .

The results of tests are just one piece of evidence that can be helpful to accept (or reject) the hypothesis of normal distribution [14]. However, the quantile-to-quantile plots exhibit the possible asymmetry of the distribution and its agreement with any theoretical distribution. It has to be noticed that the size of the merged output sample  $\mathcal{O}$  may be too large to be processed with conventional statistical tests.

The acceptance or rejection of the supposed normality of the output sample being decided, two cases arise.

**Case #1, accepting the null hypothesis:** coverage intervals having a level of confidence  $p$  can be defined for the sample and for the mean value of the sample.

- The coverage interval  $CI_{p\%}$  for the *sample*, having a level of confidence  $p\%$  is the following:

$$CI_{p\%} = [\mu - z_s \sigma; \mu + z_s \sigma], \quad (4)$$

with  $\mu$  the estimated mean value of the output sample and  $\sigma$  its estimated standard deviation.  $z_s$  is the coverage factor. We are  $p\%$  confident that a *new* value of computation of the model is in our interval. The coverage factor  $z_s$  is given by the inverse of the cumulative distribution function ( $CDF_{\mathcal{N}(0,1)}^{-1}$ ) of the normal distribution with mean 0 and standard deviation 1, as a function of the considered level of confidence  $p$  (see Tab. (G.1) in Ref. [1]):

$$z_s = -CDF_{\mathcal{N}(0,1)}^{-1} \left( \frac{1-p}{2} \right). \quad (5)$$

- The coverage interval  $CI_{p\%}^\mu$  for the *mean value of the sample* is written as:

$$CI_{p\%}^\mu = [\mu - z_m \frac{\sigma}{\sqrt{M}}; \mu + z_m \frac{\sigma}{\sqrt{M}}]. \quad (6)$$

We are  $p\%$  confident that the mean of the output values is in  $CI_{p\%}^\mu$ . Indeed, the mean value may differ from its estimation (the estimated mean value in experiments is the best estimator of the measurand). This interval is a range of values that acts as good estimates of the unknown mean value. The coverage factor  $z_m$  is given by the inverse of the cumulative distribution function ( $CDF_t^{-1}$ ) of the Student t-distribution with mean 0 and standard deviation 1, as a function of the considered level of confidence  $p$  [1]:

$$z_m = -CDF_t^{-1} \left( \frac{1-p}{2} \right). \quad (7)$$

When the size  $M$  of the sample grows, the t-distribution approaches the normal distribution with mean 0 and variance 1 ( $\mathcal{N}(0,1)$ ). The degrees of freedom of the t-distribution is equal to  $\nu_{eff}$ . Reference [1, 2] suggests using the Welch-Satterthwaite formula to adjust the effective degrees of freedom of the output:

$$\nu_{eff} = \frac{u^4(\mathbf{o})}{\sum \left( \frac{c_i^4 u_i^4(\mathcal{I}_i)}{\nu_i} \right)} \quad (8)$$

Both the number of random inputs ( $\nu_i + 1$ ) and the uncertainties of the input parameters and variables  $u_i$  are used to evaluate the effective degrees of freedom.  $c_i$  are the sensitivity coefficients (Eq. 1). If  $c_i$  are unknown, then the degrees of freedom is set to  $M - 1$ .

The mean value is the best estimator of the output of the model. The associated coverage interval  $CI^\mu$  characterizes the statistical realization (i.e. if the computation is repeated on numerous occasions and interval estimates are made on each occasion, the resulting intervals would bracket the true population input in approximately 95 % of the cases [4]).

**Case #2, rejecting the null hypothesis :** the result of the non parametric tests of equality of probability distributions is such as  $p_{test} < \alpha$ . Therefore, the rejection of the hypothesis of normal distribution of the sample is legitimate and forces us to choose an arbitrary value of the coverage factors:  $z_s = z_m = 2$  [1]. In this case, no level of confidence can be associated to the intervals.

A first application of the propagation of uncertainties is the comparison of the results of two models. Indeed, the numerical evaluation of errors between models falls in an interval due to the uncertainties of the model inputs. This approach is close to that of the comparison between experiments and theory.

### C. Application #1: uncertainties for the comparison of models

Two models of the same experiment are considered  $\mathcal{M}_1(\mathcal{P}, \mathcal{X})$  and  $\mathcal{M}_2(\mathcal{P}, \mathcal{X})$ . For example, one of them can be an approximation of the other one, leading to simpler analytical expression. We focus on the agreement of the output data of both models, by processing the output samples. Each model gives output samples  $\mathbf{o}_1$  and  $\mathbf{o}_2$  that can be compared using the coverage intervals described in Sec. 2.B. Indeed, the overlap between coverage intervals of the mean of the output indicates that both models are compatible with each other. On the contrary, in case of no overlap, the approximation fails.

These inputs produce:  $\mathbf{o}_1 = \mathcal{M}_1(\mathcal{P}, \mathcal{X})$  and  $\mathbf{o}_2 = \mathcal{M}_2(\mathcal{P}, \mathcal{X})$ . The relative error  $\mathbf{e}_r$  between the two models can be computed for all inputs that are randomly generated in their respective interval of uncertainty:

$$\mathbf{e}_r(\mathbf{o}_1, \mathbf{o}_2) = \left(1 - \frac{\mathbf{o}_2}{\mathbf{o}_1}\right). \quad (9)$$

The minimum and the maximum of the set of relative error delineate the validity of the approximation, reflecting the propagation of uncertainties of inputs. Indeed, the uncertainties of inputs could screen the discrepancy between both models. In this case, both models could be considered as equivalent, according to the considered level of uncertainties, if the interval of relative errors contains 0.

Even if the random inputs produce apparently random outputs, the models may induce deterministic links between them. The analysis of correlations reveals such a deterministic relation. Indeed, if significant correlations are found, some behavior laws can be deduced [3]. The behavior laws help to explain the underlying physics and to predict or discuss experimental results [6, 15].

### D. Application #2: uncertainties for behavior laws

Basically, the value of the Pearson's correlation coefficient  $\rho$  (Eq. (2)) is a measure of the linear dependence between two quantities in a statistical sample. Therefore,  $\rho$  is an indicator of the behavior law (or causal dependence) that could be deduced from a sample. In case of linear dependence between two quantities,  $|\rho|$  is close to 1. If the quantities have no causal relationship,  $|\rho|$  is close to 0. However the zero correlation should be considered with care: it does not imply no dependence and the visual inspection of the scatter plot of the sorted data is necessary before concluding. The correlation matrix reveal the link between all quantities involved in the problem: the outputs and the inputs.

A correlation coefficient close to  $-1$  or  $+1$  indicates that a deterministic relationship can be found by sorting the data and fitting the results by adequate functions. The fitting uses either models or polynomials. In some cases, the resulting function can be inverted, offering an alternative to solve inverse problems.

The sample of numerical model outputs is actually the result of a numerical experiments that are subject to uncertainties. Therefore, finding inputs that produce outputs that are greater than a given value, is possible. This search belongs to the optimization processes.

### E. Application #3: uncertainties for optimization process

The statistical data obtained from the propagation of uncertainties through the model can be used in an optimization process that corresponds to a change of point of view in the analysis of the results. Solving an optimization problem is seeking values of the variables that lead to an optimal value of the function that is to be optimized [16–18]. Various methods have been proposed for global optimization but some improvements must be introduced for the optimization of plasmonic nano-structures [19–22]. In the present case, we consider the simplest method of optimization based on the propagation of uncertainties. The data obtained from the propagation of uncertainties can be used for an optimization process [23, 24]. The choice of a tolerance  $t_O$  ( $0 < t_O < 1$ ) makes it possible to separate the outputs in two categories. The outputs that fill the corresponding constraint are kept and the corresponding inputs are those of the optimized system.

The optimization method involves two steps:

1. a constraint is chosen. We select the outputs  $\mathbf{o}_O$  of the model  $\mathcal{M}_n$  as :

$$\mathbf{o}_O > t_O \times \max(\mathbf{o}), \quad (10)$$

$\max(\mathbf{o})$  being the maximum of the output sample.

2. Then the intervals for each input parameter  $\mathcal{P}_O$  and input variable  $\mathcal{X}_O$  are deduced from the selected family of solutions which meet the criterion (10) .

These intervals help to improve the experiments modeled by  $\mathcal{M}_n$ , by indicating the tolerated uncertainty of each input. The global maximum  $\max(\mathbf{o})$  can also be found from a specific method of optimization (simulated annealing, particle swarm optimization method (PSO) [25, 26], evolutionary method [27, 28]...). The method based on the propagation of uncertainties provides a family of acceptable sets of inputs and therefore is closer to the experimental design. If a local optimum is found to satisfy the tolerance constraint, it is kept in the family of solutions.

Finding parameters sets or input variables that produce a given value of outputs of the numerical model is also of interest. In this case, the target is not the optimum. This search belongs to the domain of the inverse problems resolution.

### F. Application #4: uncertainties for inverse problems resolution

Solving inverse problems consists in retrieving a set of variables used in numerical model that adjusts itself to given experimental data [29]. If the model is robust and describes accurately the experiment, the recovered variables are expected to be consistent to physical reality. Some optimization methods have been proposed to solve inverse problems [26, 27, 30, 31].

Among all the solutions obtained from the propagation of uncertainties, the target is to retrieve the parameters and the variables  $(\mathcal{P}_I, \mathcal{X}_I)$  giving a specific value  $\mathbf{o}_I = \mathcal{M}_n(\mathcal{P}_I, \mathcal{X}_I)$ . Generally, the problem consists in finding the inverse function:  $(\mathcal{P}_I, \mathcal{X}_I) = \mathcal{M}_n^{-1}(\mathbf{o}_I)$ . This inverse or its approximation is rarely determined, even locally, if we have no local information on the

model. Moreover, in many cases, the solution is not unique, but a family of solutions is expected. Therefore, the output sample of the model, obtained from the propagation of uncertainties, can be used to solve the inverse problems.

The method consists in selecting the variable sets that corresponds to the target  $T$ , with possible introduction of a tolerance  $t_I$  ( $0 < t_I < 1$ ), like in the case of the optimization process:

1. a constraint on the outputs is chosen. For example:

$$T(1 - t_I) \leq \mathbf{o}_I \leq T(1 + t_I). \quad (11)$$

2. Then the possible values for each input variable  $\mathcal{X}_I$  can be found.

### 3. IMPLEMENTATION FOR THE SCATTERING OF LIGHT BY GOLD NANOPARTICLES

The propagation of uncertainties described in Sec. 2 is applied to the scattering of light by spherical gold nanoparticles. We use two numerical models in the examples of applications: the full Mie theory and the dipole approximation. Following is an outline of this section: subsection 3.A deals with the basics of the full Mie theory and of the dipole approximation. Some details on the dipole approximation are given to be able to discuss the following results. The uncertainties of the inputs (parameters and variable) are evaluated in Sec. 3.B from references. Sections 3.C-3.D are devoted to the analytical and statistical approaches of the propagation of uncertainties, respectively. The applications of the propagation of uncertainties to correlation analysis and behavior laws determination, to optimization, to inverse problems can be found in Secs. 3.E, 3.F and 3.G.

#### A. The full Mie theory and the dipole approximation

The Mie theory [32] describes the interaction of light with material particles through an analytical approach. Fully analytical solution can be found for spherical nanoparticles. Notations are summarized in Tab. 7. The scattering efficiency of an homogeneous, isotropic spherical particle of refractive index  $m$  (the surrounding medium is air), and size parameter  $x = 2\pi R/\lambda_0$  ( $R$  is the particle radius and  $\lambda_0$  is the wavelength of the illuminating light) can be written as follows:

$$Q_{sca} = \frac{2}{x^2} \sum_{n=1}^{+\infty} (2n+1) (|a_n|^2 + |b_n|^2). \quad (12)$$

The coefficients  $a_n$  and  $b_n$  of the Mie series, are deduced from the formulation of the scalar wave equation in spherical coordinates and from the boundary conditions on the surface of the sphere:

$$a_n = \frac{m^2 j_n(mx) [x j_n(x)]' - j_n(x) [mx j_n(mx)]'}{m^2 j_n(mx) [x h_n^{(1)}(x)]' - h_n^{(1)}(x) [mx j_n(mx)]'} \quad (13)$$

$$b_n = \frac{j_n(mx) [x j_n(x)]' - j_n(x) [mx j_n(mx)]'}{j_n(mx) [x h_n^{(1)}(x)]' - h_n^{(1)}(x) [mx j_n(mx)]'} \quad (14)$$

where  $j_n$  and  $y_n$  are the spherical Bessel functions and  $h_n^{(1)}$  are the spherical Bessel functions of first kind sometimes called spherical Hankel functions (equations (4.9), (4.10), (4.13) and (4.14) of reference [33]). Free codes for computing the scattering efficiency can easily be found on internet.

For subwavelength particles ( $x \ll 1$ ), power series expansion about  $x = 0$  of the scattering efficiency (Eq. (12)) can be

deduced after some algebra [33, 34]. The result is a function of the relative permittivity  $\epsilon_r = m^2$  (this parameter is a complex number). Limiting the order of the Mie scattering coefficients to the lowest powers of  $x$  seems legitimate for the calculation of the scattering efficiency  $Q_{sca}$ :

$$Q_{sca} = Q_{sca}^d + \frac{16}{5} \left[ \frac{|\epsilon_r - 1|^2 (|\epsilon_r|^2 - 4)}{|\epsilon_r + 2|^4} \right] x^6 - \frac{32}{3} \left[ \Im(\epsilon_r) \frac{|\epsilon_r - 1|^2}{|\epsilon_r + 2|^4} \right] x^7, \quad (15)$$

with  $|\bullet|$  the modulus of a complex number, and  $Q_{sca}^d$  the dipole approximation:

$$Q_{sca}^d = \frac{8}{3} \frac{|\epsilon_r - 1|^2}{|\epsilon_r + 2|^2} x^4, \quad (16)$$

and therefore:  $Q_{sca} = Q_{sca}^d + O(x^6)$ . Complementary theoretical and numerical considerations on the validity of the dipole approximation can be found in Refs. [35, 36].

The maximum of the scattering efficiency  $\max_{\lambda_0}(Q_{sca})$  over the visible spectrum and its position  $\lambda_r$  in the spectrum are the fingerprints of the localized surface plasmon resonance of the metallic nanoparticle. This resonance is a complex phenomenon which depends both on the radius  $R$  of the nanoparticle and on its relative permittivity  $\epsilon_r$ . Therefore the methods presented in Sec. 2 are applied to the problem of the scattering of light by a spherical gold nanoparticles:

- the models are  $\mathcal{M}_1 = \max_{\lambda_0}(Q_{sca})$  and  $\mathcal{M}_2 = \max_{\lambda_0}(Q_{sca}^d)$ ;
- the input of the models is  $\mathcal{I} = (\Re(\epsilon_r), \Im(\epsilon_r), R)$ , where  $\Re(\epsilon_r)$  and  $\Im(\epsilon_r)$  are respectively the real and imaginary parts of the relative permittivity of gold and  $R$  is the radius of particles;
- the output of the models is  $\mathbf{o} = (\max_{\lambda_0}(Q_{sca}), \lambda_r)$  where  $\max_{\lambda_0}(Q_{sca})$  is the maximum of the scattering efficiency within the visible wavelength spectrum, that occurs at  $\lambda_0 = \lambda_r$ . Actually, the outputs are the abscissa and the ordinate of the maximum of the function  $Q_{sca}(\lambda_0)$ .

The direct application of the propagation of uncertainties described in [2] is impossible as the uncertainties of the relative permittivity and of the radius do not play the same role in the numerical model. The radius is a classical input data, while the relative permittivity acts at an intermediate level, when the maximum is searched over the spectrum. The relative permittivity of gold may differ at the same wavelength, according to the handbooks or papers. Therefore, using this dispersion of inputs can help to improve the discussion of validity, from a metrological point of view. The basic idea is to propagate the uncertainties through the model and then to process the output results, in the same way as repeated experimental data could be. The next subsection is therefore devoted to the numerical evaluation of uncertainties for the inputs of the full Mie theory and of the dipole approximation. This first stage makes use of experimental results, bibliographical references about the measurement of the inputs of models: mean values, dispersion and eventually probability distribution.

## B. Uncertainties of the inputs of the model

The uncertainties comes from imperfect measurement, random variations of the observation nor incomplete knowledge of certain physical phenomena [1]: “Uncertainty of measurement is thus an expression of the fact that, for a given measurand and a given result of measurement of it, there is not one value but an infinite number of values dispersed about the result that are consistent with all of the observations and data and one’s knowledge of the physical world, and that with varying degrees of credibility can be attributed to the measurand.” Consequently we evaluate the uncertainties of the radius  $R$  of nanoparticles and that of the relative permittivity of gold ( $\epsilon_r$ ) from bibliographical references.

### B.1. Uncertainty of the radius of the particle

Experimentally, the uncertainty of the radius  $R$  of nanoparticles is related to the preparation method [34]. The size distribution of “gold standard” nanoparticles was carefully addressed from NIST reference materials RM 8011, RM 8012, and RM 8013, with nominal radii of 5, 15, and 30 nm [37]. The reference values from NIST were  $4.55 \pm 0.90$  nm (*i.e.* 20%),  $12.45 \pm 0.60$  nm (*i.e.* 5%) and  $26.6 \pm 2.6$  nm (*i.e.* 10%). The uncertainty can reach 50% for nanometric size particles [38] even if a few papers report multimodal distributions  $6.68 \pm 0.10$  nm (*i.e.* 1.5%) and  $6.82 \pm 0.06$  nm (*i.e.* 0.9%) [39]. However, typical values of relative uncertainties are around 10% [37, 40–42]. According to the dispersion of the available experimental data, for the following numerical applications using  $R \approx 25$  nm, we consider the standard uncertainty for the NIST reference materials RM 8013 (10%):

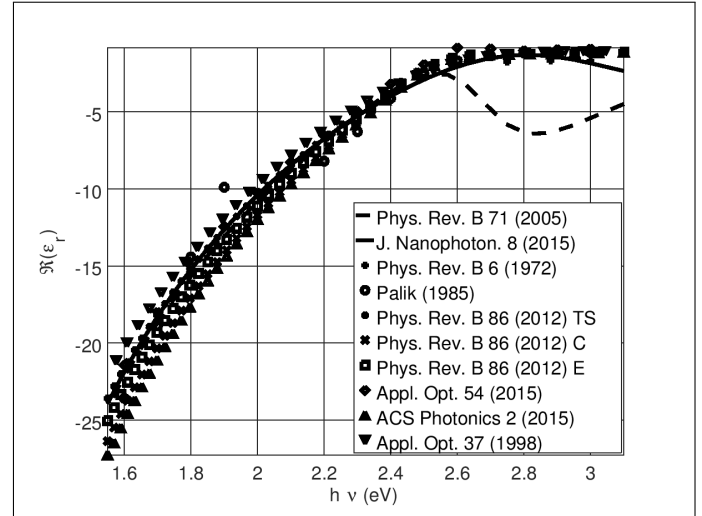
$$u(R) = 0.1R \text{ or } u_r(R) = 10\%, \quad (17)$$

where  $u_r$  is the relative uncertainty. Histograms in most of those references show a gaussian shape (or normal) distribution of size, but tests of the hypothesis of normal distribution have been sparingly performed on the above mentioned experimental data. Actually, these statistical tests of the null hypothesis that the sample comes from a normal distribution, could increase the reliability of the uncertainties and of the coverage intervals which can be deduced from. In the present case, it could also induce the choice of the normal distribution for generating the radii used for the propagation of uncertainties through the model. Let us now investigate the uncertainty of the optical parameters of the gold nanoparticles.

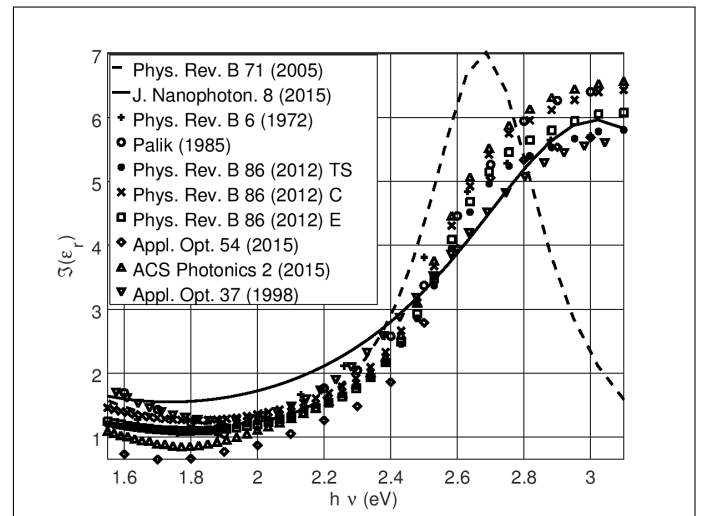
### B.2. Uncertainty of the relative permittivity

The measurement of complex dielectric constant of a single gold nanoparticle revealed high variability [43]. The relative permittivity of gold was shown to depend on the preparation method [34]. A collection of measured relative permittivities of gold can be found in [44]. They come from published articles, private communications and handbooks [45]. Figures 1 and 2 show respectively the real and the imaginary part of the relative permittivity of gold as a function of the photon energy, from references [46] (Phys. Rev. B 6 (1972)), [47] (Palik (1985)), [48] (Phys. Rev. B 86), [49] (Appl. Opt. 54 (2015)), [50] (ACS Photonics 2 (2015)) and [51] (Appl. Opt. 37 (1998)).

We consider the domain of variations of the relative permittivity of gold, as the min-max value of the real part  $\Re(\epsilon_r)$  and imaginary part  $\Im(\epsilon_r)$ , respectively. These min-max values are computed from all the experimental data shown in Figs. 1-2, at each wavelength  $\lambda_0$ . Figure 3 shows the minimum, the mean value and the maximum of both the real and the imaginary parts of the relative permittivity as functions of the wavelength.



**Fig. 1.** Real part of the relative permittivity of gold from various references as a function of the photon energy  $h\nu = \hbar\omega$ . Drude-Lorentz fitting of [46] data (dashed line [52] and solid line [53]), experimental data from [46] (Phys. Rev. B 6 (1972)), [47] (Palik (1985)), [48] (Phys. Rev. B 86), [49] (Appl. Opt. 54 (2015)), [50] (ACS Photonics 2 (2015)) and [51] (Appl. Opt. 37 (1998)).

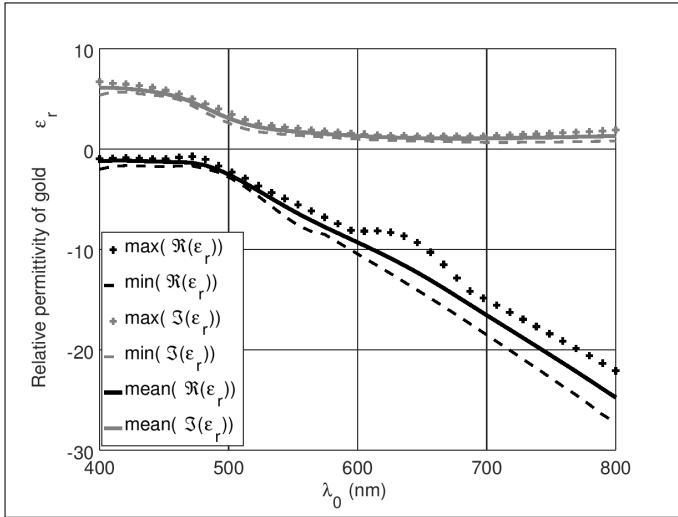


**Fig. 2.** Imaginary part of the relative permittivity of gold from various references as a function of the photon energy  $h\nu = \hbar\omega$ . Drude-Lorentz fitting of [46] data (dashed line [52] and solid line [53]), experimental data from [46] (Phys. Rev. B 6 (1972)), [47] (Palik (1985)), [48] (Phys. Rev. B 86), [49] (Appl. Opt. 54 (2015)), [50] (ACS Photonics 2 (2015)) and [51] (Appl. Opt. 37 (1998)).

According to the experimental setup, the optical properties are generally measured at different wavelengths. Therefore, for computational purpose, we have to evaluate all of them at the same wavelengths. To reach this goal, two approaches are commonly used:

- Special functions are used for fitting: they come from physical considerations on dispersion and are used in computational codes like finite difference time domain (FDTD) [52, 53]. For illustration, the fitting of data proposed in [52] and [53] are plotted (solid and dashed lines) in Figs. 1-2.
- Interpolation or spline fitting of the reference data.

Even if the results from Ref. [53] are better than those from Ref. [52] in the 400 – 800 nm range of wavelengths, the discrepancy between the fitting by special functions and the experimental data leads us to prefer the spline fitting. For numerical applications, both the real and the imaginary parts of the relative permittivity of gold are calculated for  $\lambda_0 \in [510; 530]$  nm using a step of 0.1 nm. In this interval of wavelengths, the ratio of the min-max values to the mean values varies from 25% to 26% and from 16% to 30% for the real and imaginary parts of the relative permittivity of gold, respectively.



**Fig. 3.** Minimum, maximum and mean value of the real (black) and imaginary (gray) part of the relative permittivity used for computing the uncertainty as a function of the wavelength  $\lambda_0$ .

The structure of gold in the particle is supposed unknown and therefore all experimental values found in the literature mentioned above are considered to be equally probable. The uncertainty is therefore deduced from the min-max domains, assuming uniform distribution of probability, and this at each wavelength  $\lambda_0$  [1]:

$$u(\Re(\epsilon_r(\lambda_0))) = \frac{\max(\Re(\epsilon_r(\lambda_0))) - \min(\Re(\epsilon_r(\lambda_0)))}{2\sqrt{3}}, \quad (18)$$

$$u(\Im(\epsilon_r(\lambda_0))) = \frac{\max(\Im(\epsilon_r(\lambda_0))) - \min(\Im(\epsilon_r(\lambda_0)))}{2\sqrt{3}}. \quad (19)$$

The uncertainties of inputs of models being evaluated, analytical calculations of the uncertainty of the scattering efficiency can be deduced from the dipole approximation, and the analytical approach of the propagation of uncertainties can be illustrated.

### C. Propagation of uncertainties: the analytical approach (Illustration of application #1)

Let us consider the dipole approximation of the scattering efficiency (Eq. (16)). The simple formulation of  $Q_{sca}^d$  facilitates the application of the analytical method for the propagation of uncertainties proposed in Sec. 2.A. Indeed, the calculation of the required partial derivatives is straightforward. The partial derivatives (sensitivity coefficients) are calculated from Eq. (16), using  $x = kR$  with  $k = 2\pi/\lambda_0$ . Assuming non negligible uncertainties of both the radius  $R$  of the particle and on the real and imaginary part of the relative permittivity  $\epsilon_r$ , the square of the combined standard uncertainty  $u^2(Q_{sca}^d)$  is deduced from the three partial derivatives of  $Q_{sca}^d(R, \Re(\epsilon_r), \Im(\epsilon_r))$  (see Eq. (20)).

In the present study, the dispersion of radii as well as the evaluation of the uncertainty of the relative permittivity are obtained from independent references. Therefore the correlation between the radius and the relative permittivity is set at 0 [1]. Only the correlation between the real and imaginary parts of the relative permittivity is included in Eq. (20). Considering correlated measurements of the real part and the imaginary part of the relative permittivity, the square of the relative standard uncertainty  $u_r^2(Q_{sca}^d)$  is the sum of four terms:

$$u_r^2(Q_{sca}^d) = \frac{u^2(Q_{sca}^d)}{(Q_{sca}^d)^2} = T_R^2 + T_{Re}^2 + T_{Im}^2 + T_\rho, \quad (20)$$

with:

$$\begin{aligned} Q_{sca}^d T_R &= \left( \frac{\partial Q_{sca}^d}{\partial R} \right) u(R) = \left( \frac{32}{3} k \frac{|\epsilon_r - 1|^2}{|\epsilon_r + 2|^2} \right) x^3 u(R) \\ &= 4 \frac{u(R)}{R}, \end{aligned} \quad (21)$$

$$\begin{aligned} Q_{sca}^d T_{Re} &= \left( \frac{\partial Q_{sca}^d}{\partial \Re(\epsilon_r)} \right) u(\Re(\epsilon_r)) = 16x^4 u(\Re(\epsilon_r)) \\ &\left( \frac{(\Re(\epsilon_r) - 1)(\Re(\epsilon_r) + 2) - (\Im(\epsilon_r))^2}{|\epsilon_r + 2|^4} \right), \end{aligned} \quad (22)$$

$$\begin{aligned} Q_{sca}^d T_{Im} &= \left( \frac{\partial Q_{sca}^d}{\partial \Im(\epsilon_r)} \right) u(\Im(\epsilon_r)) \\ &= \left( 16 \frac{\Im(\epsilon_r)(2\Re(\epsilon_r) + 1)}{|\epsilon_r + 2|^4} \right) x^4 u(\Im(\epsilon_r)), \end{aligned} \quad (23)$$

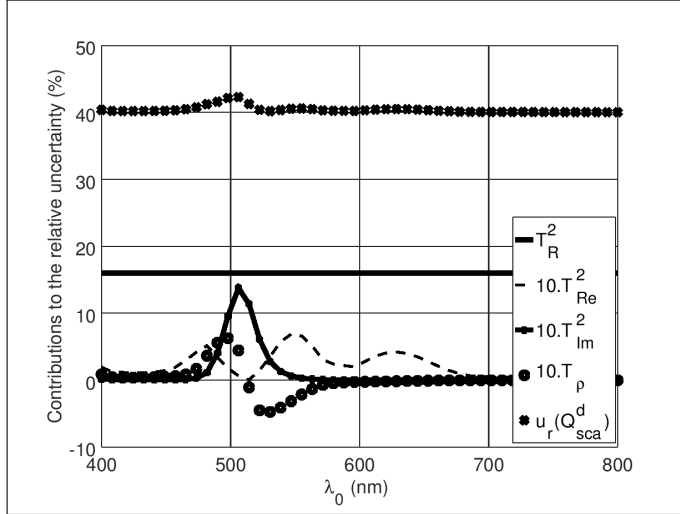
$$T_\rho = 2T_{Re}T_{Im}\rho(\Re(\epsilon_r), \Im(\epsilon_r)). \quad (24)$$

$T_\rho$  can be negative according to the sign of the partial derivative. The estimated correlation coefficient  $\rho$  between the real part  $\Re(\epsilon_r)$  and the imaginary part  $\Im(\epsilon_r)$  of the relative permittivity can be evaluated from retrieved data (see Figs. 1-2): it varies from 0.746 and 0.830. Considering the mean value of both the real  $\mu(\Re(\epsilon_r))$  and the imaginary parts  $\mu(\Im(\epsilon_r))$ , the correlation coefficient is  $\rho = 0.754$ .

This analytical approach is able to reveal the dominant terms, if exist, in the combined standard uncertainty.  $T_R$  is proportional to  $x^3$  whereas the other terms are proportional to  $x^4$ , therefore, it can be dominant for nanoparticles of subwavelength size ( $x < 1$ ). Figure 4 shows the contribution of each additive term in the square of the combined standard uncertainty (Eqs. (20)-(24)), for



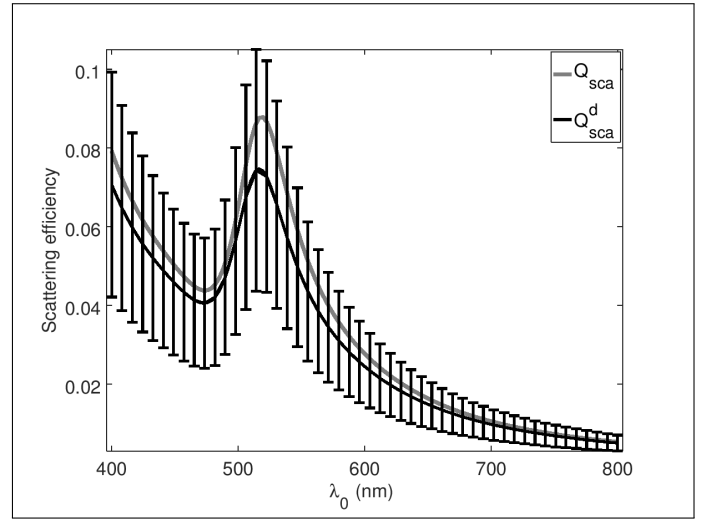
$R = 25$  nm. Clearly the uncertainty of the radius of particle is prevalent unlike that on the relative permittivity. The maximum of the relative uncertainty is 42% and occurs at  $\lambda_0 = 506$  nm. In this zone, the contribution of the real part is vanishing, whereas that of the imaginary part exhibits a maximum. The scattering efficiency is therefore more sensitive to the imaginary part of the relative permittivity. Further explanations and illustrations can be found in Refs. [54, 55] and fall out of the scope of this tutorial.



**Fig. 4.** Contribution of each term in the sum in Eq. (20) to the combined uncertainty calculated from the dipole approximation,  $T_R^2$  (Eq. (21)),  $T_{Re}^2$  (Eq. (22)),  $T_{Im}^2$  (Eq. (23)) and  $T_\rho$  (Eq. (24)) are respectively the contributions of the uncertainties of the radius  $R = 25$  nm, of the real and the imaginary parts of the relative permittivity of gold, and of the Pearson's correlation coefficient  $\rho(\Re(\epsilon_r), \Im(\epsilon_r))$ .

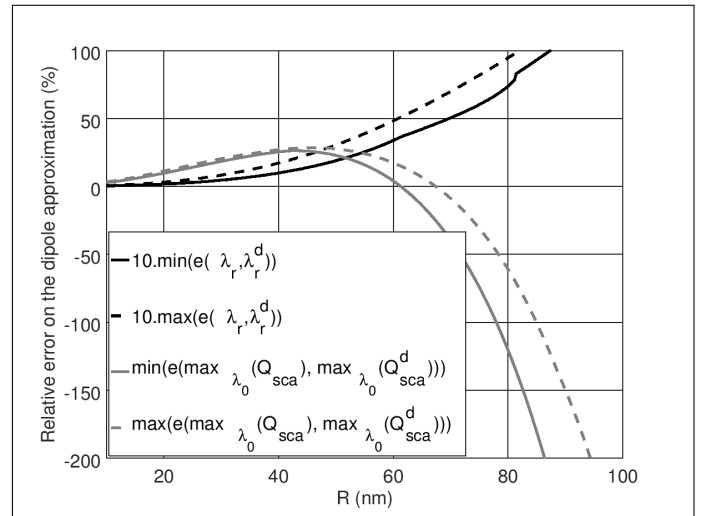
Figure 5 illustrates the propagation of uncertainties of the relative permittivity through the dipole approximation, by plotting uncertainty bars (the uncertainty intervals), as a function of the wavelength, for  $R = 25$  nm. The dipole approximation underestimates (-15%) and shifts the resonance (blue shift of about 2 nm). The scattering efficiency computed for the mean values of the data used for the relative permittivity of gold is shown (gray solid line: full Mie theory, and black solid line: dipole approximation). The curve of the scattering efficiency computed from the full Mie theory pass through the uncertainty intervals and therefore, both models are compatible for  $R = 25$  nm, reflecting the high value of the relative uncertainty of the radius. Considering the combined standard uncertainty of the calculation of the dipole approximation of the scattering efficiency, the discrepancy between the dipole approximation and the full Mie theory appears to be negligible. Nevertheless, both models should be carefully compared before concluding, by calculating the relative error of the approximation, taking into account the uncertainties of their inputs.

The sample generated from the propagation of uncertainties can be used to calculate intervals of error for each radius of particle. Indeed, from either side of the mean value, the dispersion of error induced by uncertainties provides numerical information which maybe useful to justify the choice or the rejection of the dipole approximation. The relative errors  $e_r$  (in percent) of the approximation of the full Mie theory by the dipole term can be evaluated for each output of both models (Eq. 9):  $e_r(\lambda_r, \lambda_r^d)$  and



**Fig. 5.** Scattering efficiency calculated with the dipole approximation (Eq. (16), black solid line) and with the full Mie theory (Eq. (12), gray solid line) for a particle radius  $R = 25$  nm. Uncertainty bars are shown:  $[Q_{sca}^d(\mu(\epsilon_r)) - u(Q_{sca}^d); Q_{sca}^d(\mu(\epsilon_r)) + u(Q_{sca}^d)]$  (Eqs. (16) and (20)).

$e_r(\max_{\lambda_0}(Q_{sca}), \max_{\lambda_0}(Q_{sca}^d))$ . Figure 6 shows the boundaries of relative errors as functions of the radius  $R$  of the nanoparticle. Only the uncertainties of the relative permittivity defined in Sec. 3.B.2 (see Fig. 3 and Eqs. 18-19) are used for these computations, for each radius  $R$  of nanoparticles. The relative error on the maximum of the scattering efficiency over the visible spectrum is computed from Eqs. (12) and (16) (gray lines). The relative error on the spectral position of this maximum is deduced from the same computation (black lines).



**Fig. 6.** Minimum and maximum of the relative errors (in percent)  $e(\lambda_r, \lambda_r^d)$  (black lines) and  $e(\max_{\lambda_0}(Q_{sca}), \max_{\lambda_0}(Q_{sca}^d))$  (gray lines) as functions of the nanoparticle radius  $R$ . All experimental values of the relative permittivity of gold cited in Sec. 3.B are used.

The investigated domain of radii is [15; 100] nm. The min-max range of the relative error  $e_r(\lambda_r, \lambda_r^d)$  on the position of the resonant wavelength obtained from the full Mie theory and

from the dipole approximation is slowly increasing with the radius. This relative error is an increasing function of the radius  $R$  but remains lower than 6% for  $R < 25$  nm. On the other hand, the shape of the min-max range of the relative error  $e_r(\max_{\lambda_0}(Q_{sca}), \max_{\lambda_0}(Q_{sca}^d))$  exhibits a maximum near  $R = 42$  nm, vanishes near  $R = 61$  nm. Then, it is negative and decreases rapidly with the radius, due to the major role of the multipolar terms in the Mie series (Eq. 15). For  $R$  between 62 and 68 nm, the zero value of the error is included in the min-max interval. Therefore, both models are compatible, for this range of radii. This condition is never ensured for the resonant wavelength. Consequently, both models are almost never compatible, when we take only the uncertainty of the relative permittivity of gold into account. It has to be noticed that the distance between the minimum and the maximum of the relative error increases with the radius of the nanoparticle, therefore the uncertainty of the relative permittivity cannot be neglected.

In this section, the analytical approach of the propagation of uncertainties has been used to discuss the respective weight of each uncertainty in that on scattering efficiency, calculated from the dipole approximation (Fig. 4). These first results have allowed us to discuss on the validity of the dipole approximation for  $R = 25$  nm (Fig. 5). Considering the relative error between both models, a more general study of the influence of the uncertainty of the relative permittivity has revealed two different behaviors if the quantity of interest is either the value of the scattering efficiency or that of the resonant wavelength (Fig. 6).

Despite the usefulness of the analytical approach (dipole approximation), it suffers from a limitation: first it is only applicable to simple analytical formula and second, it does not help to answer directly to the question: *assuming the three sources of uncertainty ( $R$ ,  $\Re(\epsilon_r)$  and  $\Im(\epsilon_r)$ ), what are the uncertainties of both the spectral position  $\lambda_r$  and of the value of the resonance peak maximum?* Further analytical calculation would be necessary to address this problem. Considering a model for the dependence of the relative permittivity on the wavelength, finding the inverse function of the derivative of the scattering efficiency is not obvious. However, answering this question is of interest if the tuning of plasmon-polariton is sought [23], in particular for fluorescence enhancement [56, 57]. The following statistical approach of the propagation of uncertainties through models can help to answer to this question.

#### D. Propagation of uncertainties: an example of realization of the statistical approach (Illustration of application #1)

The propagation of uncertainties through models has its origins in the repetition of observations in experiments. "The uncertainty of the result of a measurement reflects the lack of exact knowledge of the value of the measurand. The result of a measurement after correction for recognized systematic effects is still only an estimate of the value of the measurand because of the uncertainty arising from random effects and from imperfect correction of the result for systematic effects" [1]. The natural counterpart in numerical simulation involves the random generation of the inputs which are subject to uncertainties. Therefore, the repeated executions of the numerical model for each set of inputs gives a sample of output results. The output sample can be considered as the realizations of a random variable which involves both the characteristics of the model and the input uncertainties [2]. To illustrate the propagation of uncertainties and its applications, we use the full Mie theory and the dipole approximation as numerical models (Sec. 3.A). Tables 4-6 in appendix B specify the parameters, variables, outputs, and

uncertainties, in relation to the theoretical names in Sec. 2.

The first step is to choose the initial size of the sample, to ensure the reliability of the results. The accuracy (significant digits) of each output quantity is deduced from the precision of classical measurements in physics [6] and from *a priori* information.

- According to the uncertainty of the average radius  $R = 25$  nm, determined in Sec. 3.B.1, the most probable interval of variation of  $R$  is [22.5;27.5] nm and the minimum number of radii is 20 for a target accuracy of 0.25 nm.
- From the results in Sec. 3.B.2, the resonant wavelength  $\lambda_r$  is located around 520 nm. Therefore, the interval of wavelengths  $\lambda_0$  can be limited to [510;530] nm, and we deduce the minimum number of wavelengths  $\lambda_0$ : 200. This sampling enables to locate  $\lambda_r \pm 0.1$  nm.
- At each wavelength, the minimum number of random permittivities is deduced from the min-max interval about 520 nm and the acceptable tolerance. If we set it to 5%, the minimum number of relative permittivities is 20.

In this case, the production of an output sample of  $M = 400$  data requires 80,000 evaluations of each model. The number of inputs has been chosen to make a trade-off between the computing time and the significance of the results. The greater the number of random inputs, the more reliable the results are.

Nevertheless, this first evaluation does not ensure the accuracy of the mean value, of the standard deviation and of the boundaries of the coverage intervals obtained from random generation of the inputs. Fortunately, the accuracy of the outputs can be controlled by processing the results of repeated realizations of the procedure. For this, the results of the repeated realizations of the 400 outputs computation are processed according the Monte Carlo adaptive procedure described in Ref. [2]. The algorithm of the procedure is given in appendix A. The standard deviations of the averages, of the standard deviations and of the boundaries of the coverage intervals are calculated along all Monte Carlo trials. The loop of realizations continues until the stopping criterion is satisfied:

$$\mathbf{s}_y = \max(\{\mathbf{s}_\mu, \mathbf{s}_{u(\mathcal{O})}, \mathbf{s}_{y_{low}}, \mathbf{s}_{y_{high}}\}) \leq \mathbf{u}_r(\mathbf{o}), \quad (25)$$

with  $\mathbf{o} = (\max_{\lambda_0}(Q_{sca}), \lambda_r)$ . The maximum of the relative standard deviations of each of the estimated values from the samples ( $\mathbf{s}_\mu$  for the mean value,  $\mathbf{s}_{u(\mathcal{O})}$  for the merged output along the realizations and  $(\mathbf{s}_{y_{low}}, \mathbf{s}_{y_{high}})$  for the boundaries of the coverage intervals) must be lower than the relative tolerance on each output  $\mathbf{u}_r(\mathbf{o})$ . The results of the Monte Carlo procedure are following.

##### D.1. Coverage intervals for the whole sample (size 80,400)

The standard deviation over the Monte Carlo trials of the averages, of the standard deviations and of the coverage intervals boundaries provide indicators of the convergence of the method (see algorithms 1-3 in appendix A). The stopping criteria ( $\mathbf{u}_r(\mathbf{o})$  in Eq. (25)) are  $u_r(\max_{\lambda_0}(Q_{sca})) = 1\%$  and  $u_r(\lambda_r) = 0.97\%$  (Tab. 1). These values are chosen according a preliminary test showing the dispersion of each quantity and the required precision. The number of iterations required to reach the stopping criterion is  $\max(h) = 201$ . The whole output data set  $\mathcal{O}$  (size 80,400) is the merging of all samples  $\mathbf{o}$  of size 400. The coverage intervals are retrieved from the sample by finding the boundaries of the smallest interval that contains 95% of the data [2] (algorithm 3 in

appendix A). The two first significant digits of the standard deviation  $\sigma$  limit the significant digits of other results. The maximum  $s_y$  of the uncertainties of  $\lambda_r$  decreases more rapidly than that of  $\max_{\lambda_0}(Q_{sca})$ . Hence, its relative value at the end of the loop is lower than 0.01%. Only 12 realizations are necessary to reach  $u_r(\lambda_r) < 0.1\%$ . The coverage intervals with coverage factor  $p = 95\%$  for the results of the full Mie theory and the dipole approximation are overlapping even if the average of maximum of the scattering efficiency obtained from the dipole approximation is about 20% smaller. The blue shift of the resonance is about 1.6 nm for the dipole approximation. Consequently, the choice of the dipole approximation for modelling a nanoparticle of radius  $R = 25$  nm should be based on an *a priori* acceptable error.

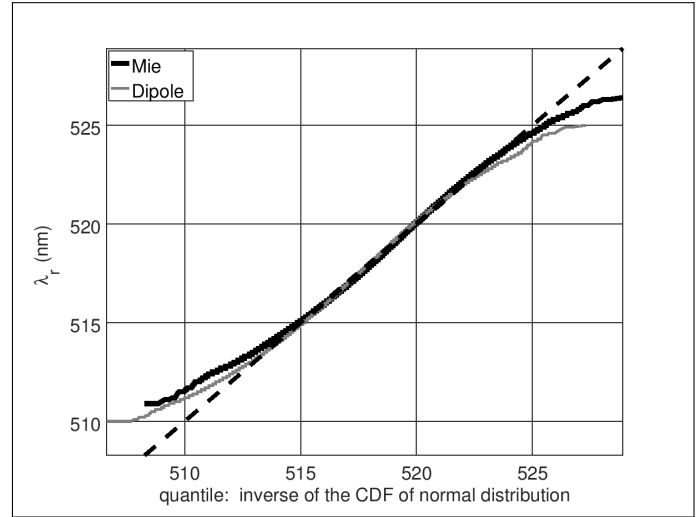
**Table 1.**  $\max_{\lambda_0}(Q_{sca})$  (full Mie theory) and  $\max_{\lambda_0}(Q_{sca}^d)$  (dipole approximation),  $\lambda_r$  (full Mie theory) and  $\lambda_r^d$  (dipole approximation): mean value  $\mu$ , standard deviation  $\sigma$ , coverage intervals  $CI_{95\%}^c$ , for a coverage factor  $p = 95\%$ . The size of the output sample  $\mathcal{O}$  is 80,400. Standard deviations  $\sigma(\cdot)$  of each quantity for the 201 Monte Carlo trials.

	$\max_{\lambda_0}(Q_{sca})$	$\max_{\lambda_0}(Q_{sca}^d)$
$\mu$	0.120	0.100
$s_\mu$ (%)	0.489	0.490
$\sigma$	0.052	0.040
$s_{u(D)}$ (%)	0.499	0.495
$CI_{95\%}^c$	[0.0150; 0.2174]	[0.0138; 0.1746]
$\{s_{y_{low}}; s_{y_{high}}\}$ (%)	{0.496; 0.482}	{0.494; 0.485}
	$\lambda_r$ (nm)	$\lambda_r^d$ (nm)
$\mu$	518.5	516.9
$s_\mu$ (%)	0.0022	0.0053
$\sigma$	2.3	2.3
$s_{u(D)}$ (%)	0.0053	0.0016
$CI_{95\%}^c$	[510.9; 522.6]	[510.0; 520.9]
$\{s_{y_{low}}; s_{y_{high}}\}$ (%)	{0.0045; 0.0029}	{0.0054; 0.0016}

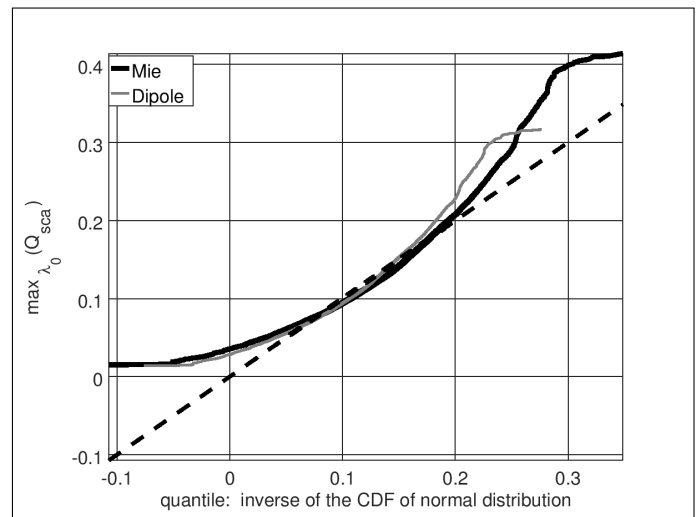
Over the visible spectrum, the analytical relative uncertainty  $u(Q_{sca}^d)$  is between 40% and 42% (Eq. (20), Fig. 4). The same values are obtained when calculating the ratio  $\sigma/\mu$ .

The large size of this merged sample prevents from doing statistical test of the  $H_0$  hypothesis of normal distribution of the outputs  $\mathcal{O} = (\max_{\lambda_0}(Q_{sca}), \lambda_r)$ . However, the visual inspection of the quantile-quantile plot (Q-Q plot) also gives an indication of the normal distribution of the sample. Actually, data from a normal distribution are close to the first bisector of the Q-Q plot (dashed line). Figures 7-8 show that the distribution of the values of the resonant wavelength may be normally distributed unlike the maximum of the scattering efficiency.

The mean value, the standard deviation and the boundaries of the coverage interval can be computed from a specific data set  $\mathbf{o}$ . In this case, the coverage interval associated to the level of confidence  $p$  provided the null hypothesis of normal distribution of the data can be accepted.



**Fig. 7.** Quantile to quantile plot of values of the maximum of the scattering efficiencies  $\max_{\lambda_0}(Q_{sca})$  (black) and  $\max_{\lambda_0}(Q_{sca}^d)$  (gray).



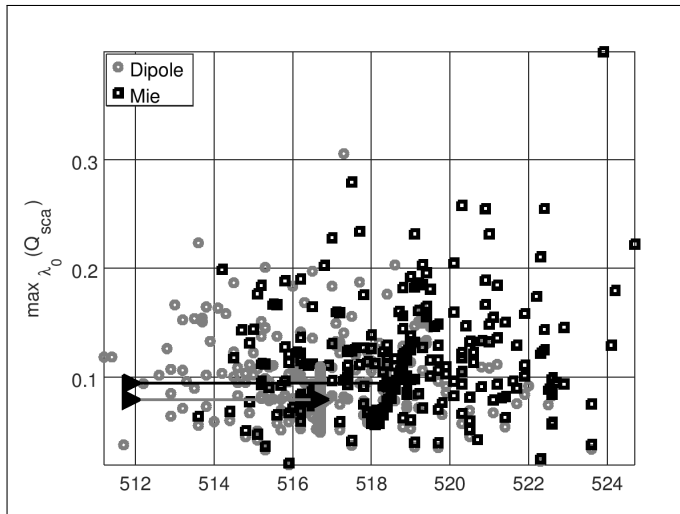
**Fig. 8.** Quantile to quantile plot of the resonant wavelengths  $\lambda_r$  (black) and  $\lambda_r^d$  (gray).

## D.2. Coverage intervals for the small sample (size 400)

A single realization of the method of the propagation of uncertainties is under consideration. This approach should be restricted to numerical models requiring time-consuming computation. The size of the sample is  $M = 400$ , with random generation of:

- 20 values of the radius  $R$  around the mean value  $\mu = 25$  nm following a normal distribution with standard deviation  $\sigma = 2.5$  nm ( $\mathcal{N}(25, 2.5)$ );
- 20 values of the relative permittivity of gold following a uniform distribution as described in Sec. 3.B.2 (Eqs. 18-19). These values are generated for each wavelength  $\lambda_0$  ( $\mathcal{U}(\mu(\epsilon_r(\lambda_0)), \sigma(\epsilon_r(\lambda_0)))$ ).

Consequently, a set of 400 inputs is generated for this realization. It requires 80,000 evaluations of both models. For each wavelength  $\lambda_0$ , the scattering efficiencies are computed from the full Mie theory ( $Q_{sca}$ ) and from the dipole approximation ( $Q_{sca}^d$ ). Then the maximums of the scattering efficiencies ( $\max_{\lambda_0}(Q_{sca})$ ,  $\max_{\lambda_0}(Q_{sca}^d)$ ) are found for each set of inputs. Each maximum corresponds to a scattering efficiency resonance which occurs at  $\lambda_r$ . Figure 9 shows the scatter plot of a realization of the algorithm. The maximum values of the scattering efficiencies



**Fig. 9.** Scatter plot of a realization of the statistical method: resonance of the scattering efficiencies  $\max_{\lambda_0}(Q_{sca})$  (black) and  $\max_{\lambda_0}(Q_{sca}^d)$  (gray) as a function of  $\lambda_r$ , the wavelength at which resonance occurs. Deterministic calculation for  $R \in [22.5; 27.5]$  nm and the mean value of  $\epsilon_r$  (thick segments).

( $\max_{\lambda_0}(Q_{sca})(\lambda_r)$  (black) and  $\max_{\lambda_0}(Q_{sca}^d)(\lambda_r)$  (gray)) are plotted as functions of the resonant wavelengths (resp.  $\lambda_r$  and  $\lambda_r^d$ ). The dispersion of the results is due to the propagation of uncertainties of both radius and relative permittivity. The two thick lines (indicated by arrows) come from the deterministic computation for  $R \in [22.5; 27.5]$  nm (dipole approximation (gray) and full Mie theory (black)), by using the mean values of the relative permittivities. The scatter of results around these small segments shows the influence of the input uncertainties of the outputs. The maximum of the scattering efficiency can reach 140% of the maximum of the deterministic evaluation (for  $R = 27.5$  nm). The small value for the dispersion of  $\lambda_r$  and  $\lambda_r^d$  confirms the analytical analysis in Sec. 3.A. On the contrary, high variations

of  $\max_{\lambda_0}(Q_{sca})$  and  $\max_{\lambda_0}(Q_{sca}^d)$  can be observed leading to the conclusion of a high sensitivity of the value of the scattering efficiency to the inputs.

Table 2 shows the statistical results of 400 computations of  $\max_{\lambda_0}(Q_{sca})$  (full Mie theory) and of  $\max_{\lambda_0}(Q_{sca}^d)$  (dipole approximation). Table 3 gives the abscissa  $\lambda_r$  and  $\lambda_r^d$  of these maximums. Each column of Tabs. 2-3 are computed from normal distributions  $\mathcal{N}(25, 2.5)$  of radii and from uniform distributions of the relative permittivity of gold (Eqs. (18)-(19)). Tables 2-3 give the mean value, the standard deviation, the results of normality tests: Lilliefors (SK), Pearson ( $\chi^2$ ), Anderson-Darling (AD), Shapiro-Wilk (SW) tests of normality, coverage factors for the coverage intervals  $z_s$  of the sample and of the mean value  $z_m$ . The effective degrees of freedom  $\nu_{eff}$  is deduced from Eq. 8 with Eqs. (21)-(23):

$$\nu_{eff}(\max_{\lambda_0}(Q_{sca}^d)) = \frac{u^4(\max_{\lambda_0}(Q_{sca}^d))}{\frac{T_R^2(\lambda_r^d)^2}{19} + \frac{T_{Re}^2(\lambda_r^d)}{19} + \frac{T_{Im}^2(\lambda_r^d)}{19}}. \quad (26)$$

In Tab. 2, the tests of normal distribution of  $\max_{\lambda_0}(Q_{sca})$  are all in agreement, considering the 201 Monte Carlo trials: all results cause the rejection of the normal distribution. The values of the standard deviation  $\sigma$  can be compared to the estimated value from the analytical approach (Sec. 3.A). Over the visible spectrum, the analytical relative uncertainty  $u(Q_{sca}^d)$  is between 40% and 42% (Eq. (20), Fig. 4). About the same values are obtained by calculating the ratio  $\sigma/\mu$ .

**Table 2.**  $\max_{\lambda_0}(Q_{sca})$  (full Mie theory) and  $\max_{\lambda_0}(Q_{sca}^d)$  (dipole approximation): mean value  $\mu$ , standard deviation  $\sigma$ , results of tests: Lilliefors ( $p_{SK}$ ), Anderson-Darling ( $p_{AD}$ ), Shapiro-Wilk ( $p_{SW}$ ), Pearson ( $p_{\chi^2}$ ). Coverage factors for the sample  $z_s$  and for the mean value  $z_m$ , coverage intervals of the sample  $CI$  and of the mean value  $CI^\mu$ . The coverage intervals  $CI_{p\%}^c$  have a level of confidence  $p = 0.95$ .

	$\max_{\lambda_0}(Q_{sca})$	$\max_{\lambda_0}(Q_{sca}^d)$
$\mu$	0.117	0.097
$\sigma$	0.051	0.040
$CI_{95\%}^c$	[0.042;0.232]	[0.037;0.185]
$p_{SK}(\mathcal{N}(\mu, \sigma))(\%)$	0	0
$p_{AD}(\%)$	0.0	0.0
$p_{SW}(\%)$	0.0	0.0
$p_{\chi^2}(\%)$	0.0	0.0
$z_s$	2	2
$CI$	[0.015;0.219]	[0.017;0.177]
$z_m$	2	2
$CI^\mu$	[0.111;0.122]	[0.093;0.101]

Table 3 gives the results for the resonant wavelengths  $\lambda_r$  and  $\lambda_r^d$  respectively computed from the full Mie theory and the dipole approximation. These wavelengths are the abscissa of  $\max_{\lambda_0}(Q_{sca})$  and  $\max_{\lambda_0}(Q_{sca}^d)$ . Yet, the retrieved resonant wavelengths seem to follow a normal distribution unlike the maximum of the scattering efficiency. Indeed, all tests of normality are in agreement ( $p_{test} > \alpha$  with  $\alpha = 5\%$  being the threshold

generally considered), the computed probabilities suggest that the hypothesis of normal distribution of  $\lambda_r$  is much more probable than that of  $\max_{\lambda_0}(Q_{sca})$ . The Anderson-Darling, Lilliefors and Pearson tests are in line with the Shapiro-Wilk's test. According to these results, the hypothesis of normal distribution of the values of  $\lambda_r$  and  $\lambda_r^d$  cannot be rejected. Consequently, the coverage intervals having a 95% level of confidence can be calculated from Eqs. (5) and (7). The coverage intervals  $CI_{95\%}^c$  are a little bit wider than the coverage interval  $CI_{95\%}$  of the sample in Tab. 1.

**Table 3.**  $\lambda_r$  (full Mie theory) and  $\lambda_r^d$  (dipole approximation): mean value  $\mu$ , standard deviation  $\sigma$ , results of tests: Lilliefors ( $p_{SK}$ ), Anderson-Darling ( $p_{AD}$ ), Shapiro-Wilk ( $p_{SW}$ ), Pearson ( $p_{\chi^2}$ ). Coverage factors for the sample  $z_s$  and for the mean value  $z_m$ , coverage intervals of the sample  $CI$  and of the mean value  $CI^\mu$ . The coverage intervals  $CI_{p\%}^c$  have a level of confidence  $p = 0.95$

	$\lambda_r$	$\lambda_r^d$
$\mu$	518.6	516.8
$\sigma$	2.3	2.3
$CI_{95\%}^c$	[512.7;522.5]	[510.4;521.0]
$p_{SK}(\mathcal{N}(\mu, \sigma))(\%)$	50.1	31.28
$p_{AD}(\%)$	55	20
$p_{SW}(\%)$	26	7
$p_{\chi^2}(\%)$	52	35
$z_s$	1.9600	1.9600
$CI$	[514.0;523.1]	[512.2;521.3]
$z_m$	1.9742	2.0003
$CI^\mu$	[518.3;518.8]	[516.5;517.0]

Even if the mean values and the standard deviations are of the same order of magnitude in Tabs. 1-3, this is not the case for the coverage intervals in Tabs. 1-3. The boundaries of the coverage intervals differ strongly for each model. These quantities are more much more sensitive to the sample size than the mean value and the standard deviation. Therefore, to deduce accurate coverage intervals from samples, the Monte-Carlo method must be used. Moreover, it gives the dispersion of all quantities. Nevertheless, the overlap of the coverage intervals of the sample  $CI$  and  $CI^c$  in each column of Tabs. 1-3 show that the Mie theory and the dipole approximation are compatible for  $R = 25$  nm. Actually, both methods can give values in the same order of magnitude.

On the other hand, the coverage intervals of the mean  $CI^\mu$  do not overlap in both columns of Tabs. 2-3. This discrepancy between the results obtained from the full Mie theory and from the dipole approximation, are the evidence of the fact that the dipole approximation is not sufficiently accurate to evaluate the maximum of the scattering efficiency over the visible spectrum, for  $R \approx 25$  nm, nor the resonant wavelength. Both model are very different in terms of metrology. This result can be related to the non vanishing error in Fig. 6, even if the uncertainty of the radius was not taken into account.

### D.3. Standard expression of the mean value

All the following numerical results are written by keeping two significant digits for the standard uncertainties of the mean value, as recommended in [1]. The number of digits kept in the estimated values of the scattering efficiency or of the resonant wavelength is limited accordingly. The last digit should be only considered as supplementary information on the order of magnitude of the estimated values. Let us note that the propagation of experimental uncertainties of inputs through a numerical model allows a discussion on the overlap between coverage intervals obtained from experiments and models. This comparison can support the often found conclusion on the "excellent agreement between experiment and theory". The results of computations can be written as a result of measurement as follows:

$$\max_{\lambda_0} Q_{sca} = (0.1174 \pm 0.0051), \quad (27)$$

$$\max_{\lambda_0} Q_{sca}^d = (0.0977 \pm 0.0040), \quad (28)$$

where the number following the symbol  $\pm$  is the numerical value of an expanded uncertainty,  $U = z_m \cdot u$ , determined from a standard uncertainty  $u$  computed from 20 values of radii following a normal distribution and from 20 random relative permittivity following an uniform distribution. The coverage factors  $z_m = 2$  is arbitrary, since the distribution of values is probably not normal. On the other hand, the resonant wavelengths seem to follow a normal distribution. The result for the Mie theory can be written as follows:

$$\lambda_r = (518.55 \pm 0.23)nm, \quad (29)$$

where the number following the symbol  $\pm$  is the numerical value of an expanded uncertainty,  $U = z_m \cdot u$ , determined from a standard uncertainty  $u$  computed from 20 values of radii following a normal distribution and from 20 random relative permittivity following an uniform distribution. The coverage factor  $z_m = 1.9742$  is based on the  $t$ -distribution for 168 degrees of freedom, and define an interval estimated to have a level of confidence of 95 percent. The resonant wavelength for the dipole approximation is:

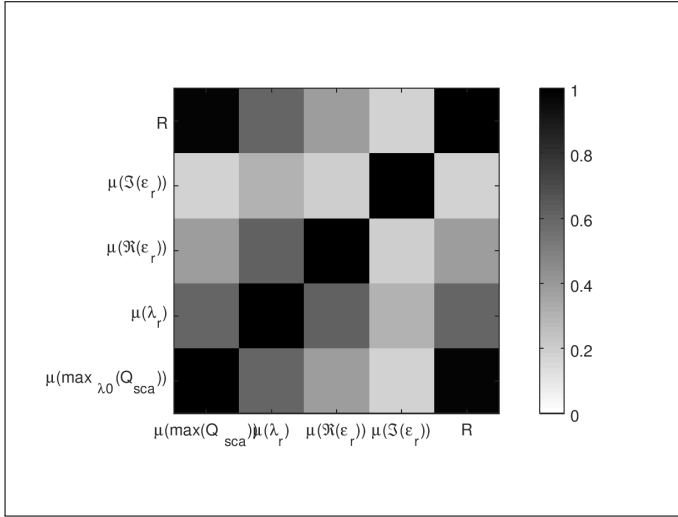
$$\lambda_r^d = (516.82 \pm 0.23)nm \quad (30)$$

where the number following the symbol  $\pm$  is the numerical value of an expanded uncertainty,  $U = z_m \cdot u$ , determined from a standard uncertainty  $u$  computed from 20 values of radii following a normal distribution and from 20 random relative permittivity following an uniform distribution. The coverage factor  $z_m = 2.0003$  is based on the  $t$ -distribution for 60 degrees of freedom, and define an interval estimated to have a level of confidence of 95 percent.

These results show that the full Mie theory and the dipole approximation are not compatible in terms of metrology, the estimated average and its dispersion being used to draw this conclusion. The behavior laws which can be deduced from the input and output samples may confirm this result.

### E. Illustration of application #2: correlation and behavior laws

For this application of the propagation of uncertainties, the Pearson's correlation coefficients  $\rho$  (Eq. 2) between the inputs ( $(\Re(\epsilon_r), \Im(\epsilon_r)), R$ ) and the output of the full Mie theory ( $\lambda_r, \max_{\lambda_0}(Q_{sca})$ ) are calculated. Figure 10 shows the correlation matrix computed from the sample calculated from the full Mie theory. The elements on the diagonal of the matrix equal to one. They correspond to the autocorrelations.



**Fig. 10.** Correlation matrix of the input parameters, variable and outputs: the mean values of the real and imaginary parts of the relative permittivity of gold ( $\epsilon_r$ ), the radius  $R$ , the values of the scattering efficiency maximum  $\max_{\lambda_0}(Q_{sca})$  and the resonant wavelength  $\lambda_r$ , calculated from the full Mie theory. The whole output data  $\mathcal{O}$  of size 80,400 is used.

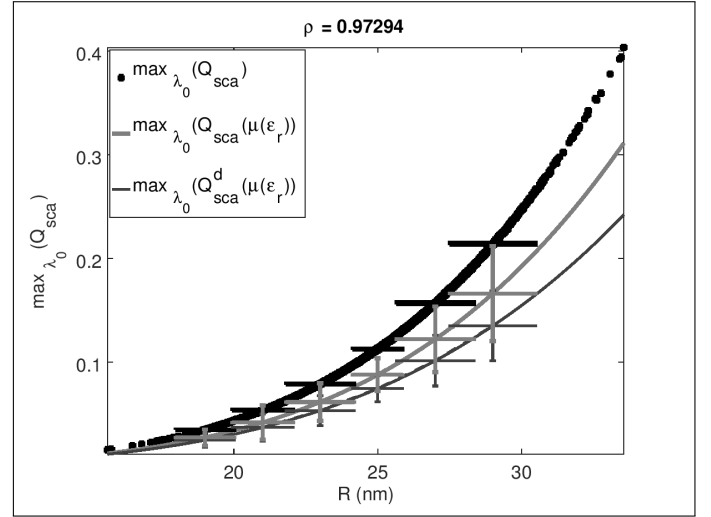
### E.1. Correlation study

The most significant correlation coefficient is  $|\rho(R, \max_{\lambda_0}(Q_{sca}))| = 0.9744$ . The next ones are  $|\rho(\Re(\epsilon_r), \lambda_r)| = 0.5566$  and  $|\rho(R, \lambda_r)| = 0.5185$ . These two last values are not close enough to 1 to be significant. These results can be confirmed by the visual inspection of all the scatter plots (not shown here). The dispersion of the correlation coefficients can be computed along the 201 realizations of the Monte Carlo procedure. The standard deviation of  $|\rho(R, \max_{\lambda_0}(Q_{sca}))|$  is 0.0013, that of  $|\rho(\Re(\epsilon_r), \lambda_r)|$  is 0.020. The dispersion decreases with the absolute value of the correlation coefficient. These uncertainties govern the number of significant digits.

### E.2. Visualization of the behavior law

Let us focus on the relationship between the radius  $R$  and the mean value of  $\max_{\lambda_0}(Q_{sca})$  computed over the 20 values of relative permittivities of gold. Sorting the random values of  $R$  and the values of the scattering efficiency maximum accordingly, reveals a deterministic law that links  $R$  and  $\mu(\max_{\lambda_0}(Q_{sca}))$ . Figure 11 shows the corresponding points. The plots of the scattering efficiencies calculated from the mean values  $\mu(\epsilon_r)$  of the relative permittivities for each wavelength (see Sec. 3.B.2). These deterministic approaches use the full Mie theory (gray line) and from the dipole approximation (black line). The corresponding uncertainties also are plotted as crosses.

The sample which comes from the propagation of uncertainties falls in the interval of uncertainties of the full Mie theory, but not in those of the dipole approximation. The maximum of the scattering efficiencies found by propagating the uncertainties is much greater than those calculated from the mean value of the 20 relative permittivities following a uniform distribution for each wavelength  $\lambda_0$ . Therefore, calculating a behavior law is possible, and the effective relative permittivities can be deduced.



**Fig. 11.** Scatter plot of the sorted results of the propagation of uncertainties. Scattering efficiency calculated from the full Mie theory (gray lines) and from the dipole approximation (black lines), using the mean value of the relative permittivities of gold for each wavelength. Uncertainties are calculated for  $\mu(R) = 19, 21, 23, 25, 27$  and  $29$  nm, with the method of propagation of uncertainties.

### E.3. Behavior laws from models

Two approximations have been proposed in Sec. 3.A (Eqs. (15) and (16) (the dipole approximation)). The dipole approximation depends on  $x^4$  and the second term of the series in Eq. (15) depends on  $x^6$ . Therefore, the fitting of the scatter plot in Fig. 11 is possible with these polynomials. Using the dipole approximation (Eq. (16):  $Q_{sca} = \alpha_1 x^4$ ) for fitting the sorted statistical data, leads to a new coefficient obtained from Levenberg-Marquardt nonlinear regression [58, 59]:  $\alpha_1 = 13.630$  (uncertainty 0.016) instead of  $\alpha_1 = 8.830$  (calculated using the mean value of relative permittivities of gold  $\epsilon_r = -3.789 + 2.310i$  at  $\lambda_0 = \lambda_r = 518$  nm, where  $i$  is the imaginary unit ( $i^2 = -1$ )). The correlation coefficient of the best-fitting curve using the  $\alpha_1 x^4$  model is  $\rho = 0.9997$  against  $\rho = 0.996$ .

The fitting of output data by the power series expansion for  $Q_{sca}(x)$  for about  $x = 0$  to order  $x^6$  (Eq. (15)) gives  $Q_{sca} = C_1 x^4 + C_2 x^6$  with  $C_1 = 11.602$  and  $C_2 = 18.90$  (uncertainties 0.013 and 0.14 respectively). The correlation coefficient is improved to  $\rho = 0.99995$ . Therefore the model used for fitting is more accurate for  $x \in [0.2149; 0.3699]$  (the dipole approximation is not sufficient to fit the data). The great advantage of using two terms in series is the possibility to deduce the effective relative permittivity of particles:  $\epsilon_r^{eff} = -2.28 + 1.76i$ :

$$\Re(\epsilon_r^{eff}) = \frac{2(5 - 3C_1)C_1 + 5C_2}{(4 + 3C_1)C_1 - 10C_2} \quad (31)$$

$$\Im(\epsilon_r^{eff}) = 3\sqrt{\frac{(-4C_1^2 + 24C_1^3 - 20C_1C_2 - 25C_2^2)}{(4C_1 + 3C_1^2 - 10C_2)^2(-16 + 5C_2)}} \quad (32)$$

### E.4. Behavior laws from polynomials

The “blind” polynomial fitting is used if no available model enables the choice of the polynomials. It gives the following

results, with correlation coefficient 0.99996:

$$\begin{aligned}\max_{\lambda_0}(Q_{sca}) &= a_0 + a_1R + a_2R^2 + a_3R^3 \\ &= 0.1128 + 0.0487R + \\ &\quad 0.00839R^2 + 0.000657R^3.\end{aligned}\quad (33)$$

The uncertainties of the polynomial coefficients are respectively 0.0015, 0.0012 nm<sup>-1</sup>, 0.00038 nm<sup>-2</sup> and 4.9 × 10<sup>-5</sup> nm<sup>-3</sup>. In this case, increasing the degree of polynomials does not improve significantly the correlation coefficient and the polynomials has no physical sense, but can also be used to predict the scattering efficiency for all radii within the considered investigated interval. Moreover, the inverse function can be calculated, giving  $R$  (nm) as a function of the desired value of the maximum of the scattering efficiency:

$$\begin{aligned}R &= \frac{1}{3a_3} \left[ \left( -a_2 + 2^{1/3} \frac{D_1}{-D_2 + \sqrt{D_2^2 - 4D_1^3}} \right)^{1/3} + \right. \\ &\quad \left. \frac{1}{2^{1/3}} \left( -D_2 + \sqrt{D_2^2 - 4D_1^3} \right)^{1/3} \right]\end{aligned}\quad (34)$$

with:

$$D_1 = a_2^2 - 3a_1a_3, \quad (35)$$

and:

$$D_2 = 2a_2^3 - 9a_1a_2a_3 + 27a_3^2(a_0 - \max_{\lambda_0}(Q_{sca})). \quad (36)$$

The sample obtained from the propagation of uncertainties has been used to deduce behavior laws and effective optical properties. The study of correlation is a first indicator of the deterministic link that can exist between inputs and outputs. Then, the non linear fitting of data, according to known models or polynomials is possible. The results of the fittings and curves in Fig. 11 confirm that the dipole approximation cannot be considered as sufficiently accurate. Consequently, only the sample  $\mathcal{O}$  obtained from the Mie theory will be used in the following applications: optimization and resolution of inverse problem.

### F. Illustration of application #3: optimization

The optimization process only corresponds to a change of point of view in the analysis of the results of the propagation of uncertainties (Sec.2.E). In the present case, optimization consists in selecting the sets of model inputs (the radii) that guarantee a sufficient level of the scattering efficiency. According to the above results, the lack of accuracy of the dipole approximation leads us to use the samples obtained from the full Mie theory.

#### F.1. Processing the whole sample of size 80,400

A tolerance  $t_O = 0.90$  gives  $\max(\max_{\lambda_0}(Q_{sca})) = 0.414$ . The number of selected data is 78. The evaluated mean value of the selected radii is  $\mu = 24.7$  nm. The standard deviation is  $\sigma = 2.4$  nm and therefore, the relative uncertainty should be reduced to 9.7%. The coverage interval ( $p = 95\%$ ) is  $CI_{95\%}^c = [20.3; 28.9]$  nm and  $s_y = 0.17\%$ . Consequently, reducing the uncertainty of the radius of the produced particles is unnecessary, if that on the relative permittivity remains unchanged.

#### F.2. Comparison with classical optimization

We can compare these results of optimization to those obtained from evolutionary optimization, by using the method described in [27]. The evolutionary optimization gives the best result over the domain of search. For each wavelength  $\lambda_0$ , 30 random relative permittivities and 30 radii are generated (uniform distributions). These populations are recombined, are mutated to produce an input set (size 100). The 30 best input sets are selected within the evolutionary loop until their relative dispersion is lower than a target (10<sup>-6</sup>). For each wavelength  $\lambda_0$ , the best radius is 27.5 nm (the upper boundary of the domain of search). Indeed, the function  $Q_{sca}(R)$  is an increasing function of the radius. The best result is  $\max(\max_{\lambda_0}(Q_{sca})) = 0.1764$  at  $\lambda_r = 519.9$  nm. Moreover, the real part and imaginary part of the relative permittivity of gold are found respectively at the maximum and the minimum of their interval of search at  $\lambda_0 = \lambda_r$ :  $\Re(\epsilon_r) = -3.4242$  and  $\Im(\epsilon_r) = 1.7702$ . The imaginary part of the relative permittivity contributes to the dimming and widening of the resonance peak of the scattering efficiency and therefore should be as small as possible. The real part tends toward the pole of the scattering efficiency  $\epsilon_r = -2$ . This result is also supported by the optimization using a wider domain of search for the relative permittivity:  $[-4; 0]$  for the real part and  $[0; 2]$  for the imaginary part. Indeed, the optimum is found for  $\Re(\epsilon_r) = -2.1887$ ,  $\Im(\epsilon_r) < 10^{-6}$  and  $R = 22.5$ . In this case, the smaller the radius is, the better the dipole approximation is. The first terms of series (15) are therefore dominant.

Even if the classical optimization is of interest to explain the physical behavior of the light scattering by gold nanoparticles, finding the global optimum does not allow us to get intervals of suitable inputs, useful for experimental tuning. In contrast to this global optimization strategy, the proposed method gives a set of acceptable inputs, by relaxing the constraint on the searched optimum. Indeed, intervals can be deduced and therefore, the improvement of the experiments is possible by indicating the critical parameters and their tolerated uncertainty.

Saving computing time is of interest and therefore, the comparison of the results obtained from the whole sample and from a single realization are of interest.

#### F.3. Processing a sample of size 400

As an example, we consider the first sample (400 values), for which  $\max(\max_{\lambda_0}(Q_{sca})) = 0.188$ . The number of selected sets of inputs is 25. The mean value for selected radii is  $\mu = 24.7$  nm with a standard deviation  $\sigma = 1.9$  nm. The relative dispersion of radii is therefore 8%. The coverage interval ( $p = 95\%$ ) is  $CI_{95\%}^c = [21.3; 28.0]$  nm. A single realization gives therefore the same order of magnitude except for the maximum of the scattering efficiency and for the boundaries of the coverage interval. Moreover, the number of significant digits of these results cannot be determined if the Monte Carlo scheme is not used.

Processing each of the 201 samples of 400 values, the minimums of the mean values, the standard deviations and the boundaries of the coverage interval for  $R$  are 26.4 nm,  $1.4 \times 10^{-5}$  nm, 25.9 nm, 26.5 nm, respectively. The corresponding maximums are 33.6 nm, 0.314 nm, 33.6 nm, 33.6 nm. Therefore, the results of the processing of single sample of size 400 should be considered as indicative and not accurate.

The random data obtained from the propagation of uncertainties have been used for optimization. Therefore, in this application of the method of propagation of uncertainties, the challenge was to recover the sets of inputs leading to a sufficient scattering efficiency. However, the target also could be the

recovery of inputs that lead to a short interval around a given target value (output). This application of the propagation of uncertainties through a model belongs actually to the resolution methods of the inverse problems.

### G. Illustration of application #4: inverse problem resolution

For decades, nanoparticles are used to enhance the signal of either Raman spectroscopy [34, 60–65] and fluorescence [56, 66–72]. For these applications, even if the excitation wavelength is matched to an electronic transition of the investigated molecule or crystal, the detected signal is weak. Therefore, the resonance of the interaction of light with nanoparticles can be used to enhance the detected signal. In this case, the challenge consists in finding the radii  $R$  of nanoparticles that could produce a resonance at a specific wavelength  $\lambda_r$ . Actually, the variable should be deduced from a constrained output of the model  $\lambda_r$ , but the inverse function of  $\max_{\lambda_0}(Q_{sca})$  cannot be calculated analytically. Fortunately the data obtained in Sec. 3.D are pairs of values  $(R, \lambda_r)$ . Therefore, the selection of nanoparticles whose resonance is located within a narrow range of wavelengths around the target is possible. [73].

As an illustration, let us search the variable sets that verify  $\lambda_r = 518 \pm 0.5$  nm. The tolerance is therefore  $t_I = 0.096\% = 0.5/518$  (Eq. (11)). The results are the followings:

- **Sample of size 80,400:** The size of the selected sample  $\mathcal{O}_I$  is 13,722. The mean value is  $\mu = 24.9$  nm and the standard deviation is  $\sigma = 2.5$  nm. The coverage interval is  $CI_{95\%}^c = [15.6; 29.1]$  nm. The uncertainty over the 201 realizations is 0.18 nm.
- **Sample of size 400:** (Tabs. 2-3), the number of selected variables is 72. The mean radius is  $\mu = 24.9$  nm and the standard deviation is  $\sigma = 2.4$  nm. The coverage interval is  $CI_{95\%}^c = [20.0; 29.8]$  nm. Consequently, the average and dispersion of the solutions retrieved from this atom sample are significant unlike the coverage interval. The small size of the second sample explains the discrepancy between both coverage intervals.
- **All samples of size 400:** the minimums of the mean values, the standard deviations and the boundaries of the coverage interval for  $R$  are 23.4 nm, 3.81 nm, 15.6 nm, 26.5 nm, respectively. The corresponding maximums are 26.8 nm, 3.81 nm, 23.8 nm, 33.6 nm. Again, the repeated realizations through the Monte Carlo scheme requires more computing time but improves the results of the inverse problem.

## 4. CONCLUSION

The analytical and statistical methods of the propagation of uncertainties were presented. We illustrated the methods by the problem of the scattering of light by a gold nanoparticle. We have discussed the validity of the dipole approximation for a gold nanosphere, 25 nm in radius. The comparison of models, the applications in optimization and in inverse problems, the determination of behavior laws from the propagation of uncertainties through a model have been presented. Cited references and suggestions for improvement could help to go further and to see other illustrations. The purpose was to give tools to make the most of the repeated computations from random generated inputs of numerical models. The proposed applications can be applied to any physical model or repeated measurements. The opportunity to address the metrological problem of

the propagation of uncertainties could enlighten the method of comparison of experimental and theoretical results with similar approaches [1].

## APPENDIX

### A. ALGORITHM USED TO DETERMINE THE SIZE OF THE SAMPLE

To make easily the connection with the Monte Carlo adaptive procedure in Ref. [2], we use the same names as much as possible. The coverage interval  $CI_{p\%}$  for the output quantity is defined in Ref. [2](7.7). This is the shortest interval which contains  $p\%$  of the output values,  $p\%$  being the coverage factor in percent. The basic idea of this algorithm is to generate  $h$  Monte Carlo trials of the  $M$  outputs  $\mathbf{o}$  until the dispersion of the mean value, the standard deviation and the boundaries of the coverage interval falls within a stop criterion. This stop criterion is defined as the relative tolerance on the output data.

**Algorithm 1.** Propagation of uncertainties algorithm. The number of inputs of the model is  $N$  ( $N = 3$  in our case:  $(R, \Re(\epsilon_r), \Im(\epsilon_r))$ ). All vectors are of dimension  $M$  ( $M = 400$ ).  $\mathcal{M}_n$  is the numerical model,  $p$  is the coverage probability,  $\mu$  is the average (or mean value) and  $\sigma$  is the standard deviation.  $\mathbf{u}_r(\mathbf{o})$  is the target accuracy for the output data.

---

```

1: procedure SAMPLE( $u_r(\mathcal{P}), I_s(\mathbf{o}), p, \mathbf{u}_r(\mathbf{o}), \mathcal{M}_n$ ) ▷ The sizes
   of  $\mathcal{P}$  and  $\mathbf{o}$  are  $M$ .
2:    $M \leftarrow \prod_{i=1}^N [u_r(\mathcal{P}_i) \cdot I_s(\mathcal{P}_i)]$  ▷ Initialization
3:    $\mathcal{O} \leftarrow \emptyset$ 
4:    $h \leftarrow 1$  ▷ The following condition must be fulfilled.
5:   while  $2s_y > \mathbf{u}_r(\mathbf{o})$  do ▷ Stopping criterion
6:      $h \leftarrow h + 1$ 
7:      $[\mathbf{o}^{(h)}, \mathcal{O}^{(h)}, \mu^{(h)}, \sigma^{(h)}, CI_{p\%}^{(h)}] \leftarrow \text{POU}(M, \mathcal{O}, \mathcal{M}_n)$ 
8:     if  $h > 1$  then
9:        $s_\mu \leftarrow \sigma(\mu^{(1..h)}) / \mu(\mu^{(1..h)}) / \sqrt{h}$ 
10:       $s_{u(\mathcal{O})} \leftarrow 2\sigma(\mathcal{O}) / \mu(\mathcal{O}) / \sqrt{h \times M}$ 
11:       $s_{y_{low}} \leftarrow \sigma(y_{low}^{(1..h)}) / \mu(y_{low}^{(1..h)}) / \sqrt{h}$ 
12:       $s_{y_{high}} \leftarrow \sigma(y_{high}^{(1..h)}) / \mu(y_{high}^{(1..h)}) / \sqrt{h}$ 
13:       $s_y \leftarrow \max(\{s_\mu, s_{u(\mathcal{O})}, s_{y_{low}}, s_{y_{high}}\})$ 
14:   return  $[\mathcal{O}]$  ▷ The sample that verifies the convergence
   criterion.
```

---

The output  $\mathcal{O}$  of the procedure SAMPLE is the sample of size  $M \times h$  that fulfills the convergence criterion: the maximum of the relative standard deviations of the average, the standard deviation and both boundaries of the coverage interval, is lower than the relative tolerated uncertainty of the outputs  $(\mathbf{u}_r(\mathbf{o}))$ . The mean values  $\mu(\mathcal{O})$ , the standard deviation  $\sigma(\mathcal{O})$  and the upper and lower boundaries of the coverage interval  $CI_{p\%}^c \leftarrow [y_{low}(\mathcal{O}); y_{high}(\mathcal{O})]$  are deduced. The relative uncertainty of each output limits the significant digits on these data. The procedure SAMPLE (Alg. 1) uses the functions POU (Alg. 2) to compute the mean value, the standard deviation and the coverage interval (function  $CI_{p\%}^c$  (Alg. 3), considering a confidence level  $p$ .

### B. GLOSSARY

The notations and quantities used in this tutorial are summarized in Tabs. 4-7.



**Algorithm 2.** Propagation of uncertainties function.  $\mathcal{M}_n$  is the numerical model,  $p$  is the coverage probability,  $\mu$  is the average (or mean value) and  $\sigma$  is the standard deviation.  $\mathbf{u}_r(\mathbf{o})$  is the target accuracy for the output data.

```

1: function POU( $M, \mathcal{O}, \mathcal{M}_n, p\%$ )
2:    $\mathcal{P} \leftarrow \mathcal{N}(\mu(\mathcal{P}), \sigma(\mathcal{P}))$   $\triangleright$  For normal law distribution.
3:   for  $j = 1..M$  do  $\mathbf{o}_j \leftarrow \mathcal{M}_n(\mathcal{P}_j, \mathcal{X})$   $\triangleright \mathcal{P}_j$  is the  $j$ -th vector
4:    $\mu \leftarrow \mu(\mathbf{o}_j)$   $\triangleright$  Vector of estimate averages.
5:    $\sigma \leftarrow \sigma(\mathbf{o}_j)$   $\triangleright$  Vector of estimate standard deviations.
6:    $[y_{low}; y_{high}] \leftarrow CI_{p\%}^c(\mathbf{o}_j, p\%, M)$   $\triangleright$  Coverage interval.
7:    $\mathcal{O} \leftarrow \mathcal{O} \cup \mathbf{o}$   $\triangleright$  Merged outputs
8:   return  $[\mathbf{o}, \mathcal{O}, \mu, \sigma, CI_{p\%}]$ 

```

**Algorithm 3.** Coverage interval.  $\lfloor \cdot \rfloor$  is the integer part [2](7.7)

```

1: function  $CI_{p\%}^c(\mathbf{y}, p\%, M)$ 
2:    $y \leftarrow \text{sort}(\mathbf{y}), \delta \leftarrow 10^{-20}$ 
3:   if  $y_{i+1} = y_i$  then
4:      $y_{i+1} = y_i + \delta$ 
      $\triangleright \mathbf{y}$  is a vector of strictly increasing data, size  $M$ 
5:    $q \leftarrow \lfloor M \times p\% / 100 \rfloor$   $\triangleright$  Integer part of  $M \times p$ 
6:    $y_{min} \leftarrow \mathbf{y}(1), y_{max} \leftarrow \mathbf{y}(M)$ 
7:   for  $r = \lfloor (M - q) / 2 \rfloor .. \lfloor (M + q) / 2 \rfloor$  do
8:     if  $\mathbf{y}(r + q) - \mathbf{y}(r) < y_{max} - y_{min}$  then
9:        $y_{max} \leftarrow \mathbf{y}(r + q), y_{min} \leftarrow \mathbf{y}(r)$ 
10:  return  $[y_{min}, y_{max}]$ 

```

**Table 4.** Summary of the notations, quantities, and uncertainties for models (Sec. 2 and Sec. 3).

Names	Quantities
$\mathcal{M}_a$	analytical model
$\mathcal{M}_n$	numerical model
$\mathcal{M}_1(\mathcal{P}, \mathcal{X})$	model
$\mathcal{M}_2(\mathcal{P}, \mathcal{X})$	model approximation
$e_r(\cdot, \cdot)$	relative error on the approximation
$\mathcal{P}$	$\epsilon_r(\lambda_0)$ input parameter
$\mathcal{X}$	$R$ input variable
$u_r(R) = 10\%$	relative uncertainty
$N = 3$	$(R, \Re(\epsilon_r), \Im(\epsilon_r))$ , inputs of the models
$\mathbf{o}$	outputs: $\max_{\lambda_0}(Q_{sca}), \lambda_r$
$u_r(\max_{\lambda_0}(Q_{sca})) = 1\%$	relative uncertainty
$u_r(\lambda_r) = 0.1\%$	relative uncertainty
$M = 400$	size of $\mathbf{o}$
$\mathcal{O}$	merged output results
$M \times \max(h) = 80,400$	size of $\mathcal{O}$ ( $201 \times 400$ )

**Table 5.** Summary of the general notations for statistics, propagation of uncertainties, optimization and inverse problems (Sec. 2.F-D and Sec. 3).

Names	Quantities
	Statistics
$\mathcal{N}(m, \sigma)$	Normal distribution, mean $m$ , standard deviation $\sigma$
$\mathcal{U}(m, \sigma)$	Uniform distribution, mean $m$ , standard deviation $\sigma$
$\mu(\cdot)$	estimated mean value (or average)
$\sigma(\cdot)$	estimated standard deviation
$s_\mu$	estimated relative standard deviation of the mean
$\rho(\cdot, \cdot)$	Pearson's correlation coefficient
$p_{test}$	Computed probability of $H_0$
$\alpha$	level of significance of the test
$\nu_{eff}$	effective degrees of freedom
$t_O = 0.90$	tolerance for optimization
$\mathcal{X}_O, \mathcal{P}_O$	solutions for optimization
$t_I = 0.096\%$	tolerance for inverse problem
$\mathcal{X}_I, \mathcal{P}_I$	solutions for inverse problem

**Table 6.** Summary of the notations, quantities for coverage intervals (Sec. 2 and Sec. 3).

Names	Quantities
	Coverage intervals
$p$ or $p\%$	level of confidence
$CI_{p\%}^c$	coverage interval
$\mathbf{y}_{low}$	lower boundary
$\mathbf{s}_{y_{low}}$	relative standard deviation
$\mathbf{y}_{high}$	higher boundary
$\mathbf{s}_{y_{high}}$	relative standard deviation
$CI_{p\%}^\mu$	coverage interval for the sample
$z_s$	coverage factor (sample)
$CI_{p\%}^\mu$	coverage interval for the estimated mean
$z_m$	coverage factor (mean)

**Table 7. Summary of the notations and quantities for the scattering of light by nanoparticles.**

Names	Quantities
$\lambda_0$	wavelength of light in vacuum
$h\nu = \hbar\omega$	photon energy
$\epsilon_r(\lambda_0)$	relative permittivity of gold (complex number)
$R$	radius of a spherical particle
$x = 2\pi \frac{R}{\lambda_0}$	size parameter of the particle
$Q_{sca}$	scattering efficiency (full Mie model)
$Q_{sca}^d$	scattering efficiency (dipole approximation)
$\max_{\lambda_0}(\cdot)$	maximum of $(\cdot)$ on an interval of wavelengths
$\max(\max_{\lambda_0}(\cdot))$	maximum on the realizations
$\lambda_r$	resonant wavelength

## REFERENCES

- Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG 1), *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*, Paris, 1st ed. (2008). Corrected version 2010 - GUM 1995 with minor corrections.
- Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG 1), *Evaluation of measurement data - Supplement 1 to the "Guide to the expression of uncertainty in measurement" - Propagation of distributions using a Monte Carlo method*, Paris, 1st ed. (2008).
- Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG 1), *Evaluation of measurement data - Supplement 2 to the "Guide to the expression of uncertainty in measurement" - Extension to any number of output quantities*, Paris, 1st ed. (2011).
- B. N. Taylor and C. E. Kuyatt, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, National Institute of Standards and Technology, Washington (1994). Supersedes NIST Technical Note 1297, January 1993.
- L. Kirkup, "A guide to gum," *European Journal of Physics* **23**, 483 (2002).
- D. Barchiesi and T. Grosgees, "Inverse problem method: A complementary way for the design and the characterization of nanostructures," *AASCI Communication* **2**, 296–300 (2015).
- J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring, *GNU Octave version 4.2.0 manual: a high-level interactive language for numerical computations* (2016).
- A. A. Borovkov, *Mathematical Statistics* (Overseas Publishers Association N. V. - Gordon and Breach Science Publishers, Amsterdam, 1998).
- H. W. Lilliefors, "On the kolmogorov-smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association* **62**, 399–402 (1967).
- A. Kolmogoroff, "Sulla determinazione empirica di una legge di distribuzione." *G. Ist. Ital. Attuari* **4**, 83–91 (1933).
- T. W. Anderson and D. A. Darling, "A test of goodness of fit," *Journal of the American Statistical Association* **49**, 765–769 (1954).
- S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)." *Biometrika* **52**, 591 (1965).
- S. Kotz and N. L. Johnson, eds., *Breakthroughs in Statistics: Methodology and Distribution* (Springer New York, New York, NY, 1992).
- G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters* (John Wiley & Sons Inc., New York, 1978).
- S. Kessentini and D. Barchiesi, *Nanostructured Biosensors: Influence of Adhesion Layer, Roughness and Size on the LSPR: A Parametric Study* (INTECH Open Access, 2013), chap. 12, pp. 311–330. ISBN 978-953-51-1004-0.
- H. Nakamura, T. Sato, H. Kambe, K. Sawada, and T. Saiki, "Design and optimization of tapered structure of near-field probe based on finite-difference time-domain simulation," *J. Microscopy* **202**, 50–52 (2001).
- N. Harris, M. J. Ford, and M. B. Cortie, "Optimization of plasmonic heating by gold nanospheres and nanoshells," *The Journal of Physical Chemistry B* **110**, 10701–10707 (2006).
- D. Barchiesi, *Numerical optimization of plasmonic biosensors* (INTECH Open Access, 2011), chap. 5, pp. 105–126. ISBN 978-953-307-448-1.
- D. H. Wolpert and W. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation* **1**, 67–82 (1997).
- S. Kessentini, D. Barchiesi, T. Grosgees, and M. Lamy de la Chapelle, "Selective and collaborative optimization methods for plasmonics: A comparison," *PIERS Online* **7**, 291–295 (2011).
- S. Kessentini and D. Barchiesi, "Particle swarm optimization with adaptive inertia weight," *International Journal of Machine Learning and Computing* **5**, 368–373 (2015).
- D. Barchiesi, "Adaptive non-uniform, hyper-elliptic evolutionary method for the optimization of plasmonic biosensors," in "Proc. Int. Conf. Computers & Industrial Engineering CIE 2009," (2009), pp. 542–547.
- T. Grosgees, D. Barchiesi, T. Toury, and G. Gréhan, "Design of nanostructures for imaging and biomedical applications by plasmonic optimization," *Opt. Lett.* **33**, 2812–2814 (2008).
- D. Barchiesi, *Surface Plasmon Resonance Biosensors: Model and Optimization* (Pan Stanford, 2013), chap. 11, pp. 333–358, High-Tech of Biotechnology.
- D. Barchiesi, "Numerical retrieval of thin aluminium layer properties from SPR experimental data," *Opt. Express* **20**, 9064–9078 (2012).
- D. Barchiesi, "The Lycurgus cup: inverse problem using photographs for characterization of matter," *J. Opt. Soc. Am. A* **32**, 1544–1555 (2015).
- D. Macías, A. Vial, and D. Barchiesi, "Application of evolution strategies for the solution of an inverse problem in Near-Field Optics," *J. Opt. Soc. Am. A* **21**, 1465–1471 (2004).
- D. Macías and D. Barchiesi, "Identification of unknown experimental parameters from noisy apertureless scanning near-field optical microscope data with an evolutionary procedure," *Opt. Lett.* **30**, 2557–2559 (2005).
- A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2005).
- J. Salvi and D. Barchiesi, "Measurement of thicknesses and optical properties of thin films from surface plasmon resonance (SPR)," *Applied Physics A* **115**, 245–255 (2014).
- D. Barchiesi, S. Kessentini, N. Guillot, M. Lamy de la Chapelle, and T. Grosgees, "Localized surface plasmon resonance in arrays of nanogold cylinders: inverse problem and propagation of uncertainties," *Opt. Express* **21**, 2245–2262 (2013).
- G. Mie, "Beiträge zur Optik trüber Medien speziell kolloidaler Metallösungen (contributions to the optics of turbid media, especially colloidal metal solutions)," *Ann. Phys.* **25**, 377–445 (1908).
- C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (John Wiley & Sons, Inc., New York, 1998).
- S. Link and M. A. El-Sayed, "Shape and size dependence of radiative, non-radiative and photothermal properties of gold nanocrystals," *International Reviews in Physical Chemistry* **19**, 409–453 (2000).
- D. Barchiesi and T. Grosgees, "Control of the applicability of the dipole approximation for gold nanoparticles," *Advanced Studies in Biology* **7**, 403–412 (2015).
- D. Barchiesi and T. Grosgees, "Short note on the dipole approximation for electric field enhancement by small metallic nanoparticles," *J. Opt.* **17**, 114003 (4 pp.) (2015).
- R. Bienert, F. Emmerling, and A. F. Thünemann, "The size distribution of 'gold standard' nanoparticles," *Analytical and Bioanalytical Chemistry* **395**, 1651 (2009).
- O. P. na, L. Rodríguez-Fernández, V. Rodríguez-Iglesias, G. Kellermann, A. Crespo-Sosa, J. C. Cheang-Wong, H. G. Silva-Pereyra, J. Arenas-Alatorre, and A. Oliver, "Determination of the size distribution of metallic nanoparticles by optical extinction spectroscopy," *Appl. Opt.* **48**, 566–572 (2009).
- R. P. Carney, J. Y. Kim, H. Qian, R. Jin, H. Mehenni, F. Stellacci, and O. M. Bakr, "Determination of nanoparticle size distribution together with density or molecular weight by 2d analytical ultracentrifugation," *Nature Communications* **2**, 1–8 (2011).
- W. Haiss, N. T. K. Thanh, J. Aveyard, and D. G. Fernig, "Determination of size and concentration of gold nanoparticles from uv-vis spectra," *Anal. Chem.* **79**, 4215–4221 (2007).
- Y. Mori, M. Furukawa, T. Hayashi, and K. Nakamura, "Size distribution of gold nanoparticles used by small angle x-ray scattering," *Particulate Science and Technology* **24**, 97–103 (2006).
- H. N. Verma, P. Singh, and R. M. Chavan, "Gold nanoparticle: synthesis and characterization," *Veterinary World* **7**, 72–77 (2014).
- P. Stoller, V. Jacobsen, and V. Sandoghdar, "Measurement of the complex dielectric constant of a single gold nanoparticle," *Opt. Lett.* **31**, 2474–2476 (2006).
- (2016). <http://refractiveindex.info>.
- L. Gao, F. Lemarchand, and M. Lequime, "Comparison of different dispersion models for single layer optical thin film index determination," *Thin Solid Films* **520**, 501 – 509 (2011).
- P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," *Phys. Rev. B* **6**, 4370–4379 (1972).
- E. D. Palik, *Handbook of Optical Constants* (Academic Press Inc., San Diego USA, 1985).
- R. L. Olmon, B. Slovick, T. W. Johnson, D. Shelton, S.-H. Oh, G. D. Boreman, and M. B. Raschke, "Optical dielectric function of gold," *Phys.*

- Rev. B **86**, 235147 (2012).
49. S. Babar and J. H. Weaver, "Optical constants of Cu, Ag, and Au revisited," *Appl. Opt.* **54**, 477–481 (2015).
  50. K. M. McPeak, S. V. Jayanti, S. J. P. Kress, S. Meyer, S. Iotti, A. Rossinelli, and D. J. Norris, "Plasmonic films can easily be better: Rules and recipes," *ACS Photonics* **2**, 326–333 (2015). PMID: 25950012.
  51. A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski, "Optical properties of metallic films for vertical-cavity optoelectronic devices," *Appl. Opt.* **37**, 5271–5283 (1998).
  52. A. Vial, A.-S. Grimault, D. Macias, D. Barchiesi, and M. Lamy de la Chapelle, "Improved analytical fit of gold dispersion: Application to the modeling of extinction spectra with a finite-difference time-domain method," *Phys. Rev. B* **71**, 085416–085423 (2005).
  53. D. Barchiesi and T. Grosge, "Errata: Fitting the optical constants of gold, silver, chromium, titanium, and aluminum in the visible bandwidth," *J. Nanophoton.* **8**, 089996 (2015).
  54. D. Barchiesi, E. Kremer, V. P. Mai, and T. Grosge, "A Poincaré's approach for plasmonics: The plasmon localization," *J. Microscopy* **229**, 525–532 (2008).
  55. D. Barchiesi and A. Otto, "Excitations of surface plasmon polaritons by attenuated total reflection, revisited," *La Rivista del Nuovo Cimento* **36**, 173–209 (2013).
  56. T. Pagnot, D. Barchiesi, D. van Labeke, and C. Pieralli, "Use of SNOM architecture to study fluorescence and energy transfer near a metal," *Opt. Lett.* **22**, 120–122 (1997).
  57. G. Parent, D. Van Labeke, and D. Barchiesi, "Fluorescence lifetime of a molecule near a corrugated interface. application to near-field microscopy," *J. Opt. Soc. Am. A* **16**, 896–908 (1999).
  58. K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.* **2**, 164–168 (1944).
  59. D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics* **11**, 431–441 (1963).
  60. M. Futamata, "Surface-plasmon-polariton-enhanced Raman scattering from self-assembled monolayers of p-nitrothiophenol and p-aminothiophenol on silver," *J. Phys. Chem.* **99**, 11901–11908 (1995).
  61. N. Féridj, J. Aubard, G. Lévi, J. R. Krenn, M. Salerno, G. Schider, B. Lamprecht, A. Leitner, and F. R. Aussenegg, "Controlling the optical response of regular arrays of gold particles for surface-enhanced Raman scattering," *Phys. Rev. B* **65**, 075419–075427 (2002).
  62. N. Féridj, J. Aubard, G. Lévi, J. R. Krenn, A. Hohenau, G. Schider, A. Leitner, and F. R. Aussenegg, "Optimized surface-enhanced Raman scattering on gold nanoparticles arrays," *Appl. Phys. Lett.* **82**, 3095–3097 (2003).
  63. J. Grand, M. Lamy de la Chapelle, J.-L. Bijeon, P.-M. Adam, A. Vial, and P. Royer, "Role of localized surface plasmons in surface-enhanced Raman scattering of shape-controlled metallic particles in regular arrays," *Phys. Rev. B* **72**, 033407 (2005).
  64. L. Billot, M. Lamy de la Chapelle, A. S. Grimault, A. Vial, D. Barchiesi, J.-L. Bijeon, P.-M. Adam, and P. Royer, "Surface enhanced Raman scattering on gold nanowire arrays: Evidence of strong multipolar surface plasmon resonance enhancement," *Chem. Phys. Lett.* **422**, 303–307 (2006).
  65. G. Sun, J. B. Khurgin, and D. P. Tsai, "Comparative analysis of photoluminescence and Raman enhancement by metal nanoparticles," *Opt. Lett.* **37**, 1583–1585 (2012).
  66. H. Chew and M. M. P. J. and Kerker, "Model for Raman and fluorescent scattering by molecules embedded in small particles," *Phys. Rev. A* **13**, 396 (1976).
  67. P. J. McNumy, S. D. Druger, M. Kerker, and H. Chew, "Fluorescent scattering by anisotropic molecules embedded in small particles," *Appl. Opt.* **18**, 1484–1488 (1979).
  68. T. Hayakawa, S. T. Selvan, and M. Nogami, "Field enhancement effect of small Ag particles on the fluorescence from  $\text{Eu}_3^+$ -doped  $\text{SiO}_2$  glass," *Appl. Phys. Lett.* **74**, 1513–1515 (1999).
  69. T. Pagnot, D. Barchiesi, and G. Tribillon, "Energy transfer from fluorescent thin films to metals in near-field optical microscopy: Comparison between time-resolved and intensity measurements," *Appl. Phys. Lett.* **75**, 4207–4209 (1999).
  70. G. Parent, D. Van Labeke, and D. Barchiesi, "Surface imaging in near-field optical microscopy by using the fluorescence decay rate: A theoretical study," *J. Microscopy* **194**, 281–290 (1999).
  71. C. Baffou, M. Kreuzer, F. Kulzer, and R. Quidant, "Temperature mapping near plasmonic nanostructures using fluorescence polarization anisotropy," *Opt. Express* **17**, 3291–3298 (2009).
  72. V. V. Datsyuk and O. M. Tovkach, "Optical properties of a metal nanosphere with spatially dispersive permittivity," *J. Opt. Soc. Am. B* **28**, 1224–1230 (2011).
  73. D. Barchiesi, "Handling uncertainties of models inputs in inverse problem: the u-discrete pso approach," in "Control, Decision and Information Technologies (CoDIT), 2014 International Conference on," I. Kacem, P. Laroche, and Z. Róka, eds. (IEEE, 2014), pp. 747–752.