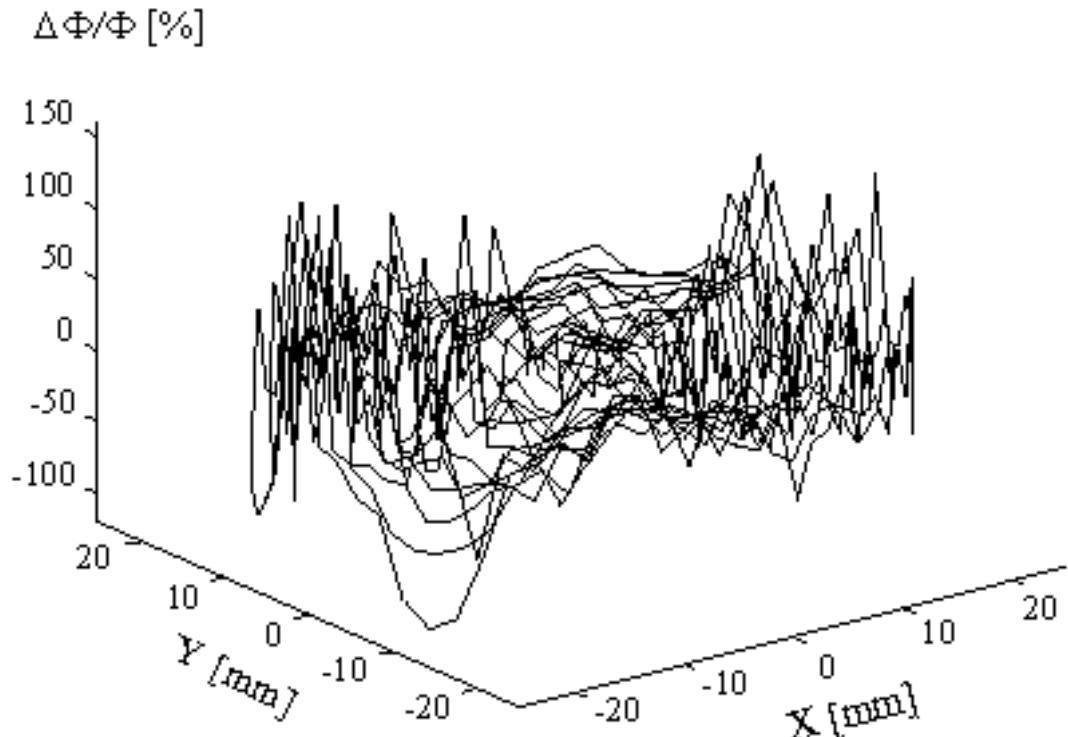


PHR 101

"Principes et outils pour l'analyse et la mesure"

Chapitre 6

Notions de statistique pour l'analyse et la mesure



6. Notions de statistique pour l'analyse et la mesure

6.1. Phénomènes déterministes ou aléatoires

Jusqu'à maintenant on a étudié des élongations, vitesses, accélération de masse dans un oscillateur mécanique. Ces grandeurs pouvaient être représentées pour une courbe représentant leur variation en fonction du temps. Il en était de même dans le cas de la charge de l'intensité du courant et de la tension aux bornes des composants d'un circuit oscillant électrique.

De même, dans l'étude des phénomènes de propagation nous avons supposé que l'on pouvait connaître à tout instant et en tout point l'évolution de l'onde mécanique le long de la corde, dans ce cas on dit que l'on a des **phénomènes déterministes** dont l'évolution peut être prédictive par un modèle mathématique, mais dans la réalité, il existe beaucoup d'autres phénomènes dont le comportement est imprévisible, ce sont les **phénomènes aléatoires** ou **stochastiques** (figure 55)

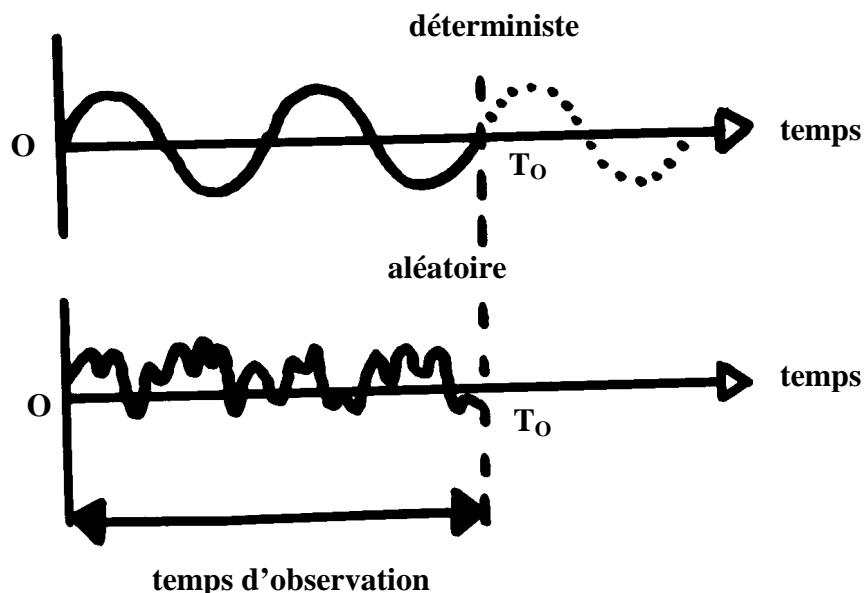


Figure 55 : la courbe de l'évolution sinusoïdale d'un phénomène en fonction du temps est typique de l'allure d'un phénomène déterministe dont on peut prévoir l'évolution au delà du temps d'observation. La courbe du bas de la figure représente un signal aléatoire dont on ne peut rien dire à priori au delà du temps d'observation.

Un phénomène aléatoire est **stationnaire** si toutes les propriétés statistiques sont invariantes dans le temps. Cette propriété de **stationnarité** est intéressante car elle permet une certaine "prédition" sur le comportement "en moyenne" des phénomènes et des signaux de mesure qui peuvent leur être associés.

6.2. Représentation statistique d'un phénomène aléatoire

La figure n°56 représente un "morceau" de signal aléatoire observé ou mesuré pendant un temps T_o . Ce signal étant aléatoire, nous ne pouvons pas écrire l'équation $y(t)$ qui le représente à l'instant t .

On peut par exemple connaître les valeurs y_1 à y_N correspondant à N "échantillons" (ici $N = 15$) pris à des intervalles réguliers de temps ΔT pendant un temps d'observation égal à T_o . En traitement du signal on montre que la période **d'échantillonnage** $\Delta T = T_o/N$ doit satisfaire certaines conditions (théorème de Nyquist) sous peine d'obtenir un résultat de mesure qui ne correspond pas à la réalité.

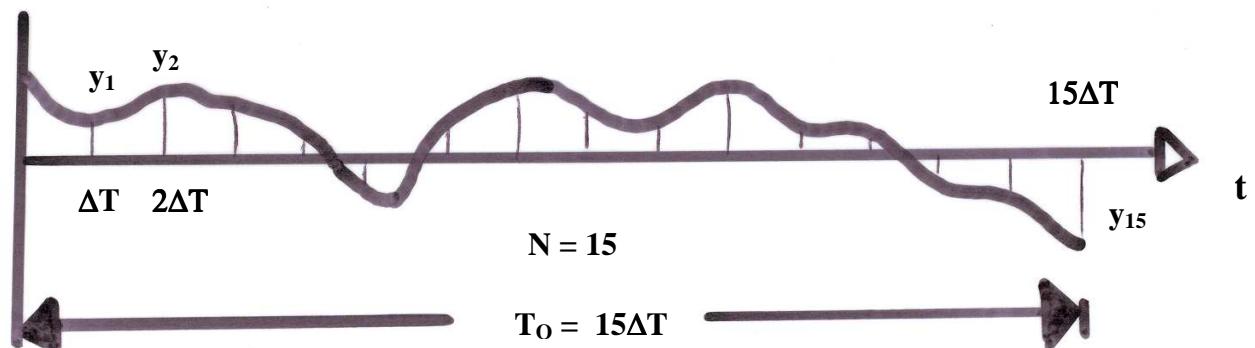


Figure 56 : Echantillonnage d'un signal aléatoire

Ainsi une variable aléatoire (signal électrique, vibration, mouvement de poussière, etc.) qui peut prendre un nombre fini de valeurs est dite **variable aléatoire discrète**, si on a un nombre infini non dénombrable de valeurs elle est dite **variable aléatoire continue**.

6.3. Valeur moyenne, variance et écart type

6.3.1. Valeur moyenne

Une grandeur continue $y(t)$ définie dans un intervalle entre 0 et T_o a pour valeur moyenne

$$\bar{y} = \frac{1}{T_o} \int_0^{T_o} y(t) dt \quad (175)$$

Une grandeur discrète définie par un ensemble de N valeurs y_N a pour valeur moyenne

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{i=N} y_i \quad (176)$$

6.3.2. Variance

Pour évaluer de combien une grandeur ou un signal s'écarte de sa valeur moyenne \bar{y} , on définit la variance σ^2 . Dans le cas d'une grandeur continue, on a :

$$\sigma^2 = \frac{1}{T_0} \int_0^{T_0} [y(t) - \bar{y}]^2 dt \quad (177)$$

Dans le cas de N échantillons d'une grandeur discrète, on obtient :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \quad (178)$$

6.3.3. Ecart type

L'**écart type** est la racine carrée de la variance.

Dans le cas particulier où la valeur moyenne \bar{y} est nulle, l'écart type σ est égal à la racine carrée de la valeur quadratique moyenne y_m .

$$\sigma = \sqrt{\frac{1}{T_0} \int_0^{T_0} y^2 dt} \quad (179)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2} \quad (180)$$

6.4. La fonction densité de probabilité ou loi de distribution ou loi de probabilité

La fonction **densité de probabilité** est fonction de la grandeur y , elle permet de connaître quelle est la **probabilité** d'obtenir, dans une certaine gamme Δy , la valeur y .

La figure n° 57(a) représente un ensemble de valeurs y_i et l'axe vertical y est divisé en m parties (ici $m = 8$) correspondant chacune à une variation Δy . On peut ainsi compter le nombre n_j de fois où y_i apparaît dans l'intervalle numéro j . Finalement, la **probabilité** P_j que le signal (ou la grandeur mesurée) apparaisse dans l'intervalle j est donnée par

$$P_j = \frac{\text{nombre total de fois où la grandeur apparaît dans l'intervalle } j}{\text{nombre total "d'échantillons" du signal}}$$

$$P_j = \frac{n_j}{N} \quad j = 1, \dots, m \quad (181)$$

La figure n° 57(b) représente cette **probabilité** P_j en fonction du numéro j de l'intervalle

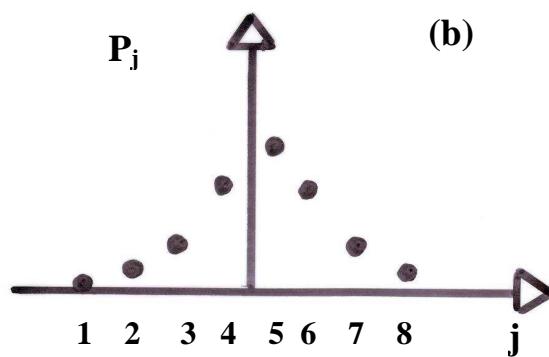
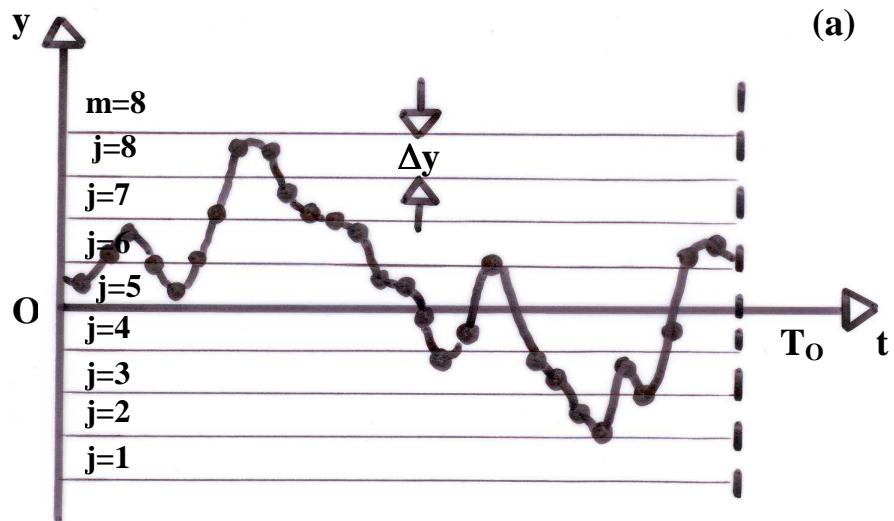


Figure 57 (a) partage de y en 8 intervalles Δy et ceci pendant un temps T_0 .

(b) probabilité P_j que le signal y apparaisse dans l'intervalle j . (j varie de 1 à 8)

On définit maintenant la **fonction de répartition** C_j qui représente la probabilité totale que le signal apparaisse dans les intervalles allant de 1 à j . Cette somme de probabilité est donnée par :

$$C_j = P_1 + P_2 + \dots + P_j \quad (182)$$

$$C_j = \frac{1}{N} (n_1 + n_2 + \dots + n_j) \quad (183)$$

La figure n°58 représente cette **fonction de répartition** C_j dans le cas particulier de la figure 57(a)

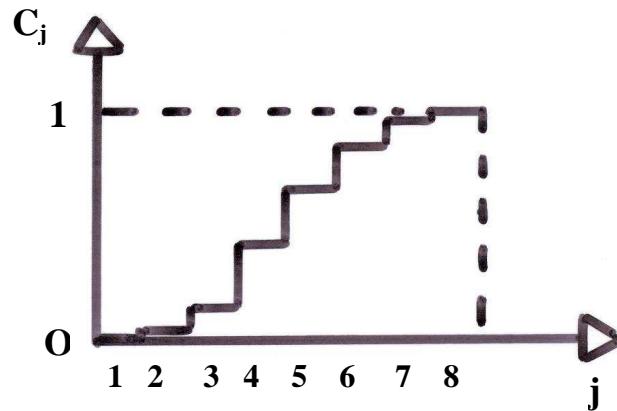


Figure 58 : fonction de répartition C_j correspondant au cas illustré sur les figures 57(a) et (b) quand $y = m$, la **fonction de répartition** devient :

$$\begin{aligned} C_m &= \frac{1}{N} (n_1 + n_2 + \dots + n_m) & (184) \\ &= \frac{1}{N} \cdot N = 1 \end{aligned}$$

La probabilité de trouver la grandeur y dans tous les intervalles est égale à 1 !

On peut également définir la fonction de répartition de $P(y)$ définie par :

$$P(y) = \lim_{\Delta y \rightarrow 0} C_j \quad (185)$$

La figure n° 59 représente l'allure de cette **fonction de répartition** dans notre cas particulier (figure n°57)

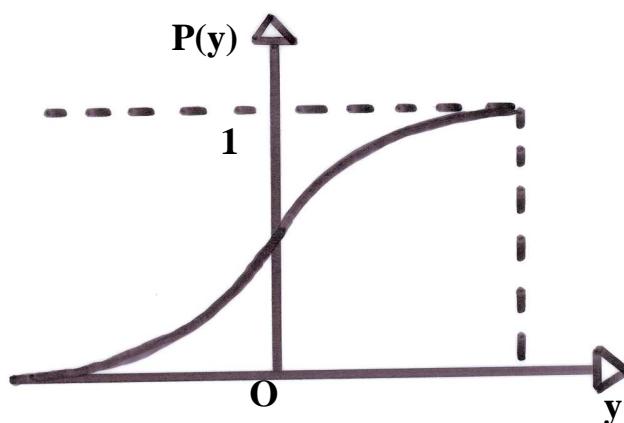


Figure 59 : allure de la fonction de répartition pour la grandeur y dans le cas particulier de la figure 57.

On définit enfin la fonction **densité de probabilité** $p(y)$ (**ou loi de distribution ou encore loi de probabilité**) qui est la dérivée par rapport à y de la fonction de répartition $P(y)$

$$p(y) = \frac{dP}{dy} \quad (186)$$

Un exemple d'allure de variation d'une densité de probabilité $p(y)$ en fonction de y est représenté sur la figure n°60.

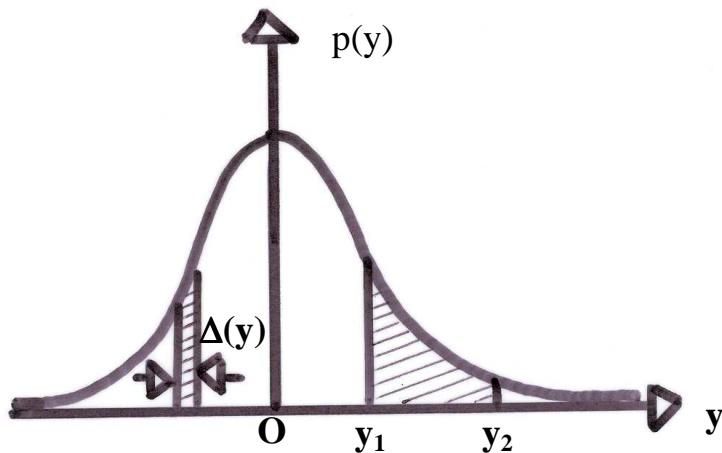


Figure 60 : allure typique de la densité de probabilité pour une grandeur y de valeur moyenne \bar{y} nulle.

La **densité de probabilité (ou loi de distribution ou encore loi de probabilité)** d'une grandeur est très employée.

La probabilité $P_y, y + \Delta y$ que la grandeur se trouve dans l'intervalle y et $y + \Delta y$ est donnée par :

$$P_{y, y + \Delta y} = p(y) \Delta y \quad (187)$$

Ce qui est représenté par la surface hachurée de "largeur" Δy sur la courbe de la figure n°60.

De même, on obtiendrait la probabilité de trouver y_1 et y_2 à partir de la relation :

$$p_{y_1, y_2} = \int_{y_1}^{y_2} p(y) dy \quad (188)$$

qui est représentée sur la figure n°60 par l'aire hachurée comprise entre y_1 et y_2 .

La surface totale sous la courbe donnant la **densité de probabilité** sera égale à l'unité correspondant à la probabilité totale que la grandeur prenne toutes les valeurs y possible.

Dans ce qui suit nous allons illustrer quelques-uns de ces concepts statistiques sur un exemple concret : la mesure de la rugosité d'une surface.

6.5. Description de la rugosité d'une surface

La figure n°61 représente une surface rugueuse. Les coordonnées x, y, z définissent les différents points de la surface.

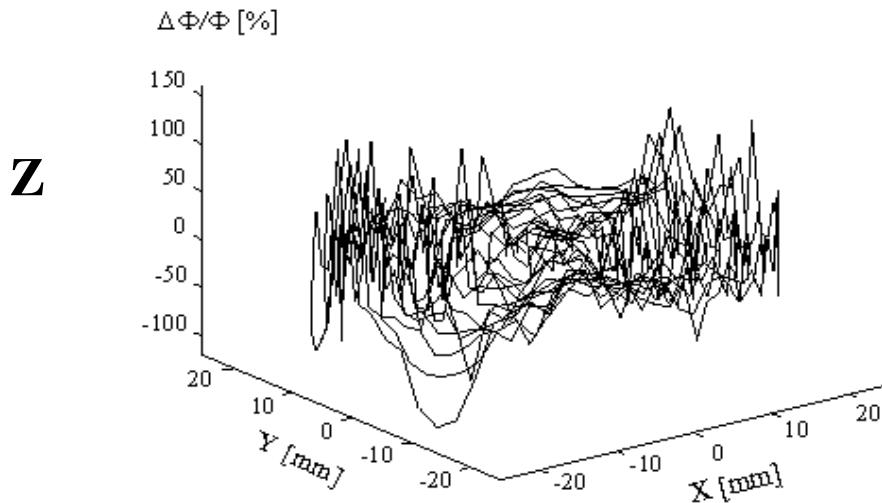


Figure 61 : mesure de la rugosité d'une surface circulaire d'un étalon de masse. Cette mesure utilise la diffusion de la lumière*

Les différents points de la surface rugueuse sont définis par leur cote Z_i par rapport au niveau moyen de la surface (figure n°62). Ce niveau moyen est défini par :

$$\sum_{i=1}^N Z_i = 0 \quad (189)$$

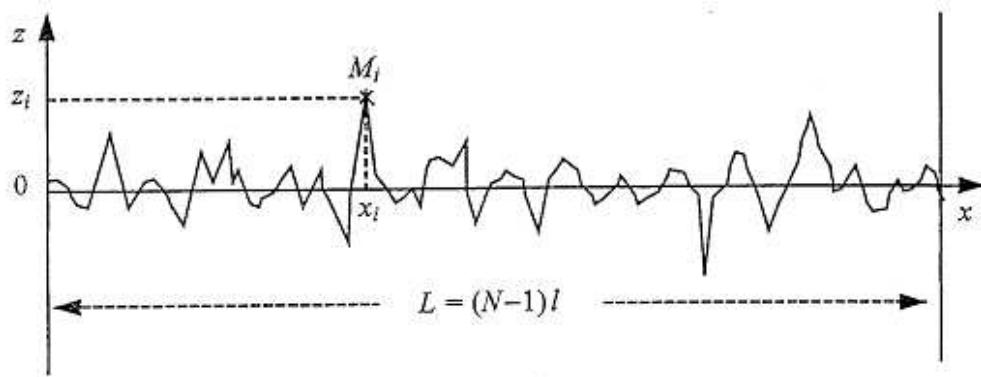


Figure 62 : représentation schématique d'une surface rugueuse (avec $L = \text{longueur du profil}$ $N = \text{nombre de points de mesure et } l = \text{distance d'échantillonnage})^*$

Le paramètre le plus utilisé pour caractériser la rugosité est la hauteur quadratique moyenne.

* C.Zerrouki et al. The nanometric roughness of mass standards and the effect of BIPM cleaning-washing techniques, Metrologia, 1999, 36, 403-414

$$\bar{\delta} = \sqrt{\frac{1}{N} \sum_{i=1}^N Z_i^2} \quad (190)$$

La figure n°63 illustre le résultat de la mesure de la hauteur Z_i en fonction de la distance δ (height) correspondant à une surface réelle dont la hauteur quadratique moyenne est égale à $4,7 \text{ \AA}$ soit $4,7 \cdot 10^{-10} \text{ m}$.

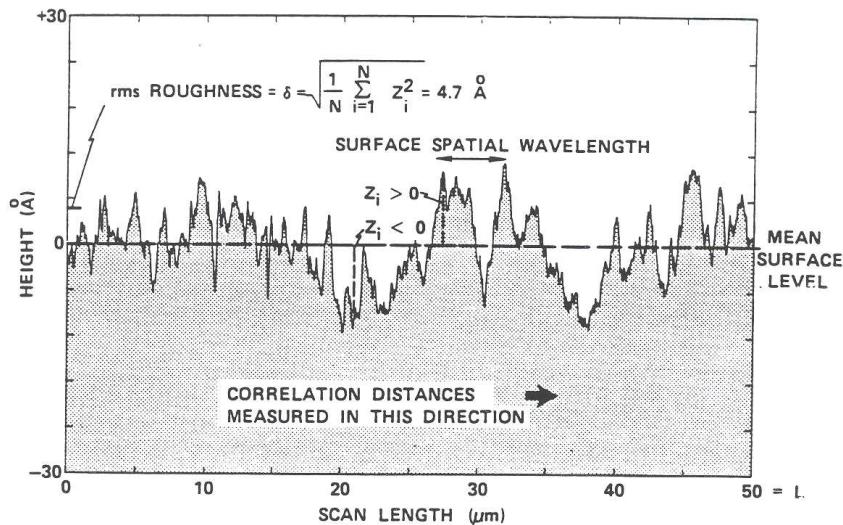
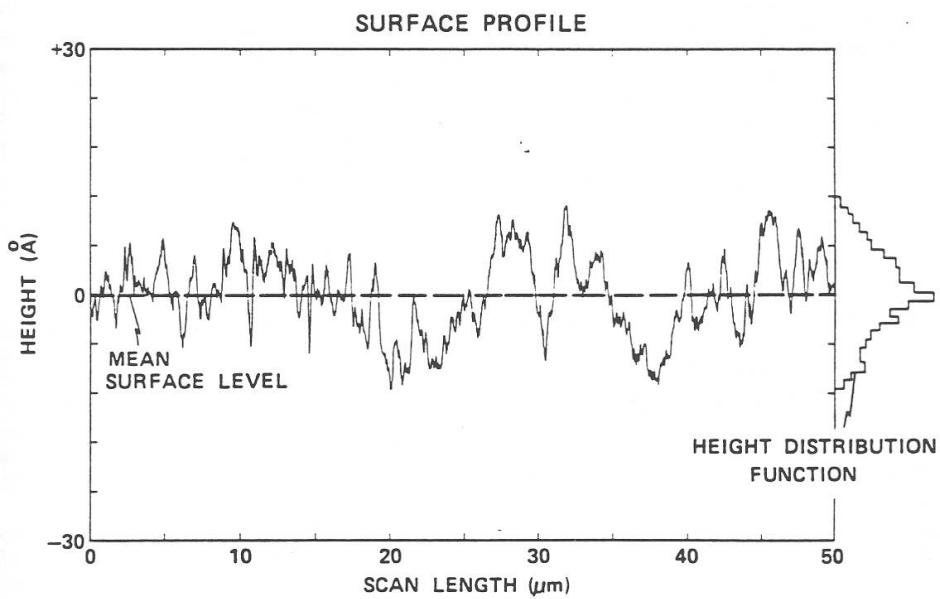


Figure 63 : représentation ("en coupe") d'une surface rugueuse avec mise en évidence des grandeurs statistiques**

Pour cette même surface, on a représenté sur la figure n°64 la loi de distribution (ou densité de probabilité ou loi de probabilité : height distribution function) correspondant à ces différentes valeurs de hauteur δ .



*Figure 64 : vue en coupe d'une surface dont la hauteur quadratique moyenne de la rugosité est $\delta = 4,7 \text{ \AA}$ et histogramme de la densité de probabilité (Height distribution function). Tirée de ***

Si dans le cas particulier de la mesure de rugosité de surface représentée sur la figure n°64 la **densité de probabilité** ne présente pas une allure typique, pour d'autres surfaces comme celle reproduite sur la page suivante, nous voyons apparaître une loi de distribution gaussienne dite aussi loi de distribution normale, avec :

$$p(Z) = A e^{-\frac{Z^2}{b^2}} \quad (190)$$

- A désigne une amplitude maximale qui représente le nombre de points qui se trouvent sur la surface moyenne (mean surface level).
- b est la valeur de Z pour laquelle $p(Z)$ a diminué de $e \approx 2,71828 \dots$ par rapport à sa valeur.

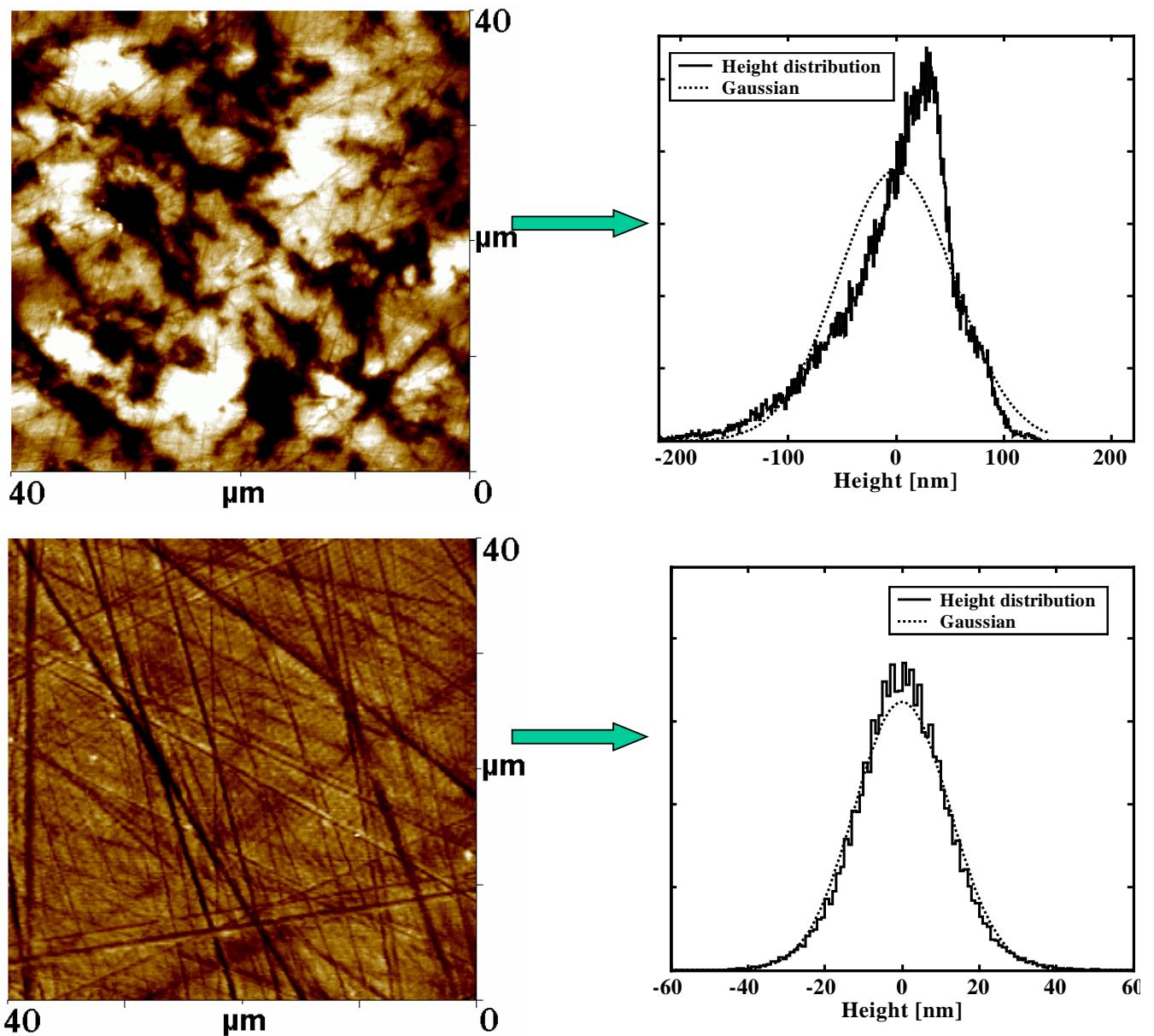
On montre que $b^2 = 2 \delta^2$ dans le cas de cette densité de probabilité gaussienne.

La surface S sous la courbe est donnée par :

$$S = A b \sqrt{\pi}$$

** J.M. Bennett and Lars Mattson, Surface Roughness and Scattering.
Optical Society of America, Washington DL 1989

Figure 65 : Exemples d'images topographiques et de distributions des hauteurs correspondantes



Youssef Haidan et al. Etude topographique d'un étalon de masse en acier inoxydable et optique en champ proche d'une surface en silicium de faible rugosité par SNOM. Journées scientifiques du CNRS "métrologie et capteurs en électromagnétisme" mars 2004 – 33 - 38

6.6. La fonction ou coefficient de corrélation

Pour mesurer le degré de similitude ou de dépendance mutuelle de deux phénomènes aléatoires, on définit une **fonction** ou un coefficient de **corrélation**.

Nous verrons par exemple, que l'émission du rayonnement lors de la désexcitation d'un atome met en jeu la fonction **d'autocorrélation** temporelle du champ électrique de l'onde émise.

Une expression typique de fonction **d'autocorrélation** temporelle de la grandeur $y(t)$ est donnée par :

$$\mathfrak{R}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} y(t) y(t - \tau) dt \quad (191)$$

L'autocorrélation consiste à comparer un signal $y(t)$ avec ce même signal temporel décalé dans le temps de $(t - \tau)$.

On montre que la fonction **d'autocorrélation** prend une valeur maximale quand $\tau = 0$.

L'expression $\mathfrak{R}(0)$ est la puissance du signal.

La fonction **d'autocorrélation** est minimale quand τ tend vers l'infini car, dans ce cas, les observations de $y(t)$ et de $y(t - \tau)$ sont totalement décorrélés.

6.7. Physique microscopique et mécanique quantique

La notion classique de trajectoire n'a aucun sens quand on mesure des noyaux, des électrons, des atomes ou des molécules qui obéissent aux lois physiques de la **mécanique quantique**. En effet, dans ce cas il faut parler de **probabilité de présence** du noyau, de l'électron, etc...

On est donc dans une "physique probabiliste".

La densité de probabilité de présence $dp(x, y, z, t)$ qu'un électron qui se trouve à l'instant t à la position $\vec{r}(x, y, z)$ dans l'élément de volume $dV(x, y, z)$ est donnée par le carré de sa fonction d'onde $\psi(x, y, z)$. Nous reviendrons sur ces concepts dans le chapitre consacré à la spectrométrie moléculaire.