ENVIRONMENTAL RISK ANALYSIS

(EENV 3351)

CHAPTER 3:
Risk Quantification

Instructor:
Dr. Yunes Mogheir
**Introduction**

- **Risk quantification** is the next step after the formulation of the problem and the analysis of different uncertainties that may cause failure (risk identification).

- Because, by definition, risk analysis is related to uncertainties, quantification of risks should be based on methodologies that may take **into account uncertainties**.

- Static reliability analysis is first considered, when loads and resistances are supposed to be constant at a given time.
Introduction

The stochastic approach is relatively well established, fitting of probability laws and analysing the dependencies between random variables need large quantities of data which are not always available.

If load and resistance are assumed to be independent, direct integration may be applied to quantify risk and reliability.

Available data may be used to determine extreme values and the risk of exceedance, such as the hydrologic risk.

Another possibility for quantifying risks is the formulation of stochastic differential equations. Monte Carlo simulation is a powerful technique for numerical representation of the system and subsequent risk quantification.
Introduction

The stochastic approach is relatively well established, fitting of probability laws and analysing the dependencies between random variables need large quantities of data which are not always available.

If load and resistance are assumed to be independent, direct integration may be applied to quantify risk and reliability.

Available data may be used to determine extreme values and the risk of exceedance, such as the hydrologic risk.

Another possibility for quantifying risks is the formulation of stochastic differential equations. Monte Carlo simulation is a powerful technique for numerical representation of the system and subsequent risk quantification.
3.1 Stochastic Approach

3.1.1 Direct Evaluation

* Loads and resistances are considered here to be positive scalars (not vector), applied on a single component of a system and are assumed to be time-independent (steady).

* The above approach is generally known as *static reliability analysis*

* Let us consider the load or exposure *C* as a random variable *L*. In this case *uncertainties* associated with the estimation of *L* are quantified by means of *probabilistic methods*. 
Suppose now that both $l$ and $r$ are positive random variables and that probabilistic methods are utilised to quantify risk. According to Eq. (2.4), in the general case, the risk $P_F$ may be computed as follows

\[
p_F = P(L > R) = \int_0^\infty \left( \int_0^\ell f_{LR} (\ell, r) dr \right) d\ell
\]

Introducing the conditional probability $f_{LR}(\ell/r) = f_{LR}(\ell, r) / f_R(r)$ into Eq. (3.1) we obtain

\[
p_F = \int_0^\infty f_{LR}(\ell/r) \left( \int_0^\ell f_R(r) dr \right) d\ell
\]
Let \( L \) and \( R \) be independent random variables.

\[
f_{LR}(\ell/r) = f_L(\ell).
\]

Fig. (3.1), the integral takes non-zero values in the overlap between the two curves \( f_L(\ell) \) and \( f_R(r) \), and Eq. (3.2) yields

\[
p_F = \int f_L(\ell) \left( \int f_R(r) \, dr \right) d\ell
\]

(3.3)

\( F_L(\ell), f_L(\ell) \): load
\( F_R(r), f_R(r) \): resistance

Fig. 3.1 Direct quantification of risk by use of probability distribution functions.
Introducing the distribution function $F_R(r)$ into Eq. (3.3), where $F_R(\ell) = \int_0^\ell f_R(r) \, dr$, the following equation is obtained

$$p_F = P(L > R) = \int_0^\infty F_R(\ell) f_L(\ell) \, d\ell$$

(3.4)

-The successive steps for quantifying risk by means of Eq. (3.4) are shown in Fig. (3.2).
Fig. 3.2 Steps for direct quantification of risk.
Example 3.1

Assume that both load and resistance are exponentially distributed, i.e.

\[ f_R(r) = \lambda_R e^{-\lambda_R r} , \quad r > 0 \]  \hspace{1cm} (3.5)

\[ f_L(\ell) = \lambda_L e^{-\lambda_L \ell} , \quad \ell > 0 \]  \hspace{1cm} (3.6)

By introducing expressions (3.5) and (3.6) into the Eq. (3.4), the risk is calculated as

\[ p_F = P(L > R) = \int_{0}^{\infty} F_R(\ell) f_L(\ell) d\ell \]

\[ = \int_{0}^{\infty} \left[ 1 - e^{-\lambda_R \ell} \right] \lambda_L e^{-\lambda_L \ell} d\ell \]

\[ = \frac{\lambda_R}{\lambda_L + \lambda_R} = \frac{1}{1 + \left( \frac{\lambda_L}{\lambda_R} \right)} = \frac{1}{1 + \frac{E(R)}{E(L)}} = \frac{1}{1 + \frac{R_0}{L_0}} \]  \hspace{1cm} (3.7)
where \( E(R) = R_0 = 1/\lambda_R \) and \( E(\ell) = L_0 = 1/\lambda_L \) are the mean values of \( R \) and \( L \). The result (3.7), shown graphically in Fig.(3.3), indicates that, for exponential distributions of \( L \) and \( R \), a unique relation exists between the risk of failure and the ratio \( R_0/L_0 \). The latter is called central safety factor.
Margin of Safety

Risk may be considered as a \textit{performance index of the system}, \textit{(Chapter 5)} where more details about the management of risk are given.

Other \textit{performance variables and performance indices of the system are the margin of safety and the safety factor}.

Because L and R are random variables, M is also a random variable with probability density distribution function $f_M(m)$. The condition for an incident or failure to happen is written as follows:

$$M \leq 0 \quad \text{or} \quad R - L \leq 0$$
Margin of Safety

\[ p_F = \int_{-\infty}^{0} f_M(m) \, dm = F_M(0) \]  

(3.10)

\( p_F \) represents the area below the curve \( f_M(m) \), for \( m < 0 \).

Fig. 3.4  Probability density distribution of safety margin \( M = R - L \).
Example 3.2
Find the risk and the reliability in terms of the safety margin $M$, if we assume that the load $L$ and resistance $R$ are normally distributed, i.e.

$$L = N(L_0, \sigma_L^2) \quad \text{and} \quad R = N(R_0, \sigma_R^2)$$

Sol,

- If $L$ and $R$ are independent, then

the safety margin $M = R - L$ is also a normal random variable, with mean value $M_0 = R_0 - L_0$ and variance $\sigma_M^2 = \sigma_L^2 + \sigma_R^2$.

- The non-dimensional variable (standardized normal distribution)

$$M' = (M - M_0)/\sigma_M$$

is a normal probability density with zero mean and unit variance

$$M' = N(0, 1)$$
The risk may be calculated as:

\[ p_F = P(M < 0) = P(R - L < 0) \]

Subtracting \( M_0 = R_0 - L_0 \) from both sides of the inequality dividing both sides by \( \sigma_M \) we obtain

\[ p_F = P\left( \frac{M - M_0}{\sigma_M} < -\frac{M_0}{\sigma_M} \right) = \Phi\left( -\frac{M_0}{\sigma_M} \right) \]

where \( \Phi \) is the cumulative normal function or the normal distribution function.

\[ \Phi(x) = \left( \frac{1}{\sqrt{2\pi}} \right) \int_{-\infty}^{x} \exp\left( -\frac{x^2}{2} \right) dx \]

Numerical values of \( \Phi \) are given in statistical handbooks and reported in Tab.(3.1).

From (3.13) we have

\[ p_F = \Phi\left( -\frac{M_0}{\sigma_M} \right) = 1 - \Phi\left( \frac{M_0}{\sigma_M} \right) = 1 - \Phi\left( \frac{M_0}{\sigma_M} \right) = 1 - \Phi\left( \frac{R_0 - L_0}{\sqrt{\sigma^2_L + \sigma^2_R}} \right) \]

(3.14)
Tab. 3.1 The cumulative normal function \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2) dt \).

<table>
<thead>
<tr>
<th>( x )</th>
<th>.005</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0</td>
<td>.5000</td>
<td>.5040</td>
<td>.5080</td>
<td>.5120</td>
<td>.5160</td>
<td>.5199</td>
<td>.5239</td>
<td>.5279</td>
<td>.5319</td>
<td>.5359</td>
</tr>
<tr>
<td>.1</td>
<td>.5398</td>
<td>.5438</td>
<td>.5478</td>
<td>.5517</td>
<td>.5557</td>
<td>.5596</td>
<td>.5636</td>
<td>.5675</td>
<td>.5714</td>
<td>.5753</td>
</tr>
<tr>
<td>.2</td>
<td>.5793</td>
<td>.5832</td>
<td>.5871</td>
<td>.5910</td>
<td>.5948</td>
<td>.5987</td>
<td>.6026</td>
<td>.6064</td>
<td>.6103</td>
<td>.6141</td>
</tr>
<tr>
<td>.3</td>
<td>.6179</td>
<td>.6217</td>
<td>.6255</td>
<td>.6293</td>
<td>.6331</td>
<td>.6368</td>
<td>.6406</td>
<td>.6443</td>
<td>.6480</td>
<td>.6517</td>
</tr>
<tr>
<td>.4</td>
<td>.6554</td>
<td>.6591</td>
<td>.6628</td>
<td>.6664</td>
<td>.6700</td>
<td>.6736</td>
<td>.6772</td>
<td>.6808</td>
<td>.6844</td>
<td>.6879</td>
</tr>
<tr>
<td>.5</td>
<td>.6915</td>
<td>.6950</td>
<td>.6985</td>
<td>.7019</td>
<td>.7054</td>
<td>.7088</td>
<td>.7123</td>
<td>.7157</td>
<td>.7190</td>
<td>.7224</td>
</tr>
<tr>
<td>.6</td>
<td>.7257</td>
<td>.7291</td>
<td>.7324</td>
<td>.7357</td>
<td>.7389</td>
<td>.7422</td>
<td>.7454</td>
<td>.7486</td>
<td>.7517</td>
<td>.7549</td>
</tr>
<tr>
<td>.7</td>
<td>.7580</td>
<td>.7611</td>
<td>.7642</td>
<td>.7673</td>
<td>.7704</td>
<td>.7734</td>
<td>.7764</td>
<td>.7794</td>
<td>.7823</td>
<td>.7852</td>
</tr>
<tr>
<td>.8</td>
<td>.8078</td>
<td>.8106</td>
<td>.8133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.8159</td>
<td>.8186</td>
<td>.8212</td>
<td>.8238</td>
<td>.8264</td>
<td>.8289</td>
<td>.8315</td>
<td>.8340</td>
<td>.8365</td>
<td>.8389</td>
</tr>
<tr>
<td>1.0</td>
<td>.8413</td>
<td>.8438</td>
<td>.8461</td>
<td>.8485</td>
<td>.8508</td>
<td>.8531</td>
<td>.8554</td>
<td>.8577</td>
<td>.8599</td>
<td>.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>.8643</td>
<td>.8665</td>
<td>.8686</td>
<td>.8708</td>
<td>.8729</td>
<td>.8749</td>
<td>.8770</td>
<td>.8790</td>
<td>.8810</td>
<td>.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>.8849</td>
<td>.8869</td>
<td>.8888</td>
<td>.8907</td>
<td>.8925</td>
<td>.8944</td>
<td>.8962</td>
<td>.8980</td>
<td>.8997</td>
<td>.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>.9032</td>
<td>.9049</td>
<td>.9066</td>
<td>.9082</td>
<td>.9099</td>
<td>.9115</td>
<td>.9131</td>
<td>.9147</td>
<td>.9162</td>
<td>.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>.9192</td>
<td>.9207</td>
<td>.9222</td>
<td>.9236</td>
<td>.9251</td>
<td>.9265</td>
<td>.9279</td>
<td>.9292</td>
<td>.9306</td>
<td>.9319</td>
</tr>
<tr>
<td>1.5</td>
<td>.9332</td>
<td>.9345</td>
<td>.9357</td>
<td>.9370</td>
<td>.9382</td>
<td>.9394</td>
<td>.9406</td>
<td>.9418</td>
<td>.9429</td>
<td>.9441</td>
</tr>
<tr>
<td>1.6</td>
<td>.9452</td>
<td>.9463</td>
<td>.9474</td>
<td>.9484</td>
<td>.9495</td>
<td>.9505</td>
<td>.9515</td>
<td>.9525</td>
<td>.9535</td>
<td>.9545</td>
</tr>
<tr>
<td>1.7</td>
<td>.9554</td>
<td>.9564</td>
<td>.9573</td>
<td>.9582</td>
<td>.9591</td>
<td>.9599</td>
<td>.9608</td>
<td>.9616</td>
<td>.9625</td>
<td>.9633</td>
</tr>
<tr>
<td>1.8</td>
<td>.9641</td>
<td>.9649</td>
<td>.9656</td>
<td>.9664</td>
<td>.9671</td>
<td>.9678</td>
<td>.9686</td>
<td>.9693</td>
<td>.9699</td>
<td>.9706</td>
</tr>
<tr>
<td>1.9</td>
<td>.9713</td>
<td>.9719</td>
<td>.9726</td>
<td>.9732</td>
<td>.9738</td>
<td>.9744</td>
<td>.9750</td>
<td>.9756</td>
<td>.9761</td>
<td>.9767</td>
</tr>
<tr>
<td>2.0</td>
<td>.9772</td>
<td>.9778</td>
<td>.9783</td>
<td>.9788</td>
<td>.9793</td>
<td>.9798</td>
<td>.9803</td>
<td>.9808</td>
<td>.9812</td>
<td>.9817</td>
</tr>
</tbody>
</table>
a numerical application is given for the case of river pollution.

According to the environmental quality standards, which have been recommended by the environmental protection agency, there is pollution in the river when the pollutant concentration \( C(t) \) exceeds a critical concentration \( C_0 \).

The permitted risk of exceeding the standards, i.e. the risk which does not affect the ecosystem, is supposed to be 10%.

This means that we should have

\[
p_F = P(C(t) > C_0) \leq 0.10
\]
If we multiply both terms in the above inequality by the river flow rate $Q(t)$ we obtain

where

$M_L(t) = C(t)Q(t)$ is the pollutant load (mass per time)

$M_R(t) = C_0Q(t)$ is the allowed pollutant mass rate or the resistant load (mass per time)

Suppose now that both loads $M_L(t)$ and $M_R(t)$ are normally distributed.

$$M_R = N(45, 7.5^2) \quad M_L = N(35, 5^2).$$

$$p_F = 1 - \Phi \left( \frac{M_{R0} - M_{L0}}{\sqrt{\sigma_{M_L}^2 + \sigma_{M_R}^2}} \right)$$

$$= 1 - \Phi \left( \frac{45 - 35}{\sqrt{7.5^2 + 5^2}} \right)$$

$$1 - \Phi(0.864) = 0.19 > 10\%$$

there is risk of pollution.
the reliability $R_c = 1 - p_F$ of the system is

$$R_c = \Phi\left(\frac{R_0 - L_0}{\sqrt{\sigma_L^2 + \sigma_R^2}}\right)$$

It is seen that both risk and reliability are functions of the coefficient

$safety \ margin \ \ M_0 = R_0 - L_0$

$reliability \ index.$

$$\beta = \frac{R_0 - L_0}{\sqrt{\sigma_L^2 + \sigma_R^2}}$$

$$\sigma_M = \sqrt{\sigma_L^2 + \sigma_R^2}.$$
Table 3.2: Risk $p_F$ as a function of the reliability index $\beta$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>0</th>
<th>1.28</th>
<th>2.33</th>
<th>3.10</th>
<th>3.72</th>
<th>4.25</th>
<th>4.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_F$</td>
<td>0.5</td>
<td>$10^{-1}$</td>
<td>$10^{-2}$</td>
<td>$10^{-3}$</td>
<td>$10^{-4}$</td>
<td>$10^{-5}$</td>
<td>$10^{-6}$</td>
</tr>
</tbody>
</table>

The risk is a function of both the relative position of $F_R(r)$ and $F_L(\ell)$.

$$\beta = \frac{R_0 - L_0}{\sqrt{\sigma_L^2 + \sigma_R^2}}$$
The Safety Factor

- Another performance index, very well known in engineering, is the safety factor $Z = R/L$,

- When $L$ and $R$ are random variables, $Z$ is also a random variable with probability density distribution function $f_z(z)$.

- The failure condition is $Z < 1$ and the risk is specified by $P(Z < 1)$.
risk is represented by the area under $f_Z(z)$ for $0 < Z < 1$. This area may be determined by means of the following integral

$$p_F = \int_0^1 f_Z(z) \, dz = F_Z(1.0)$$

(3.17)

Fig. 3.5 Probability density distribution of the safety factor $Z = R/L$. 
Example 3.3

- Find the risk and reliability when L and R are independent, log-normal random variables.

- A random variable may also called "variate." R and L are log-normal variates when their logarithmic transformations

\[ X = \ln R \quad \text{and} \quad Y = \ln L \]

follow normal distributions with parameters \( \mu_{\ln R}, \sigma^2_{\ln R} \) and \( \mu_{\ln L}, \sigma^2_{\ln L} \).

\[
\begin{align*}
    f_R(r) &= \frac{1}{r (\sqrt{2\pi}) \sigma_{\ln R}} \exp \left[ 0.5 \left( \frac{\ln r - \mu_{\ln R}}{\sigma_{\ln R}} \right)^2 \right] \\
    f_L(\ell) &= \frac{1}{\ell (\sqrt{2\pi}) \sigma_{\ln L}} \exp \left[ 0.5 \left( \frac{\ln \ell - \mu_{\ln L}}{\sigma_{\ln L}} \right)^2 \right]
\end{align*}
\]
The risk is
\[ p_F = P\left( Z = \frac{R}{L} < 1 \right) = P\left( \ln\left\{ \frac{R}{L} \right\} < 0 \right) \]  
(3.18)

The random variable \( \ln Z = \ln L - \ln R \) is also a normal variate with parameters

\[ \mu_{\ln Z} = \mu_{\ln R} - \mu_{\ln L} \]  
(3.19)

and

\[ \sigma^2_{\ln Z} = \sigma^2_{\ln R} + \sigma^2_{\ln L} \]  
(3.20)

The random variable \( Z' = \frac{\ln Z - \mu_{\ln Z}}{\sigma_{\ln Z}} \) is a normal variate with zero mean and unit variance. By means of Eq. (3.18) we obtain

\[ p_F = P\left( Z' < -\frac{\mu_{\ln Z}}{\sigma_{\ln Z}} \right) = \Phi\left( -\frac{\mu_{\ln Z}}{\sigma_{\ln Z}} \right) = 1 - \Phi\left( \frac{\mu_{\ln Z}}{\sigma_{\ln Z}} \right) \]

Taking into account the relations (3.20) and (3.21) the following result is obtained

\[ p_F = 1 - \Phi\left( \frac{\mu_{\ln Z}}{\sigma_{\ln Z}} \right) = 1 - \Phi\left( \frac{\mu_{\ln R} - \mu_{\ln L}}{\sqrt{\sigma^2_{\ln R} + \sigma^2_{\ln L}}} \right) \]

(3.21)
As a numerical example, find the risk if $R$ and $L$ are log-normal variates with the same parameters as in the previous example (3.2), i.e.

$$\mu_R = 45, \sigma_R = 7.5 \quad \text{and} \quad \mu_L = 35, \sigma_L = 5.0.$$ 

Between the corresponding mean values and variances of the variates $\ln R$, $R$ and $\ln L$, $L$ the following relations apply (Ang and Tang, 1975)

$$\mu_{\ln R} = \frac{1}{2} \ln\left[ \frac{\mu_R}{\sigma_R^2} \right] \quad \sigma_{\ln R}^2 = \ln\left[ \frac{\left( \frac{\sigma_R}{\mu_R} \right)^2 + 1}{\left( \frac{\sigma_R}{\mu_R} \right)^2} \right]$$

and

$$\mu_{\ln L} = \frac{1}{2} \ln\left[ \frac{\mu_L}{\sigma_L^2} \right] \quad \sigma_{\ln L}^2 = \ln\left[ \frac{\left( \frac{\sigma_L}{\mu_L} \right)^2 + 1}{\left( \frac{\sigma_L}{\mu_L} \right)^2} \right]$$
For

\[ \mu_R = 45, \sigma_R = 7.5 \text{ and } \mu_L = 35, \sigma_L = 5.0 \]

we find that

\[ \mu_{lnR} = 3.793, \sigma_{lnR}^2 = 0.027 \text{ and } \mu_{lnL} = 3.545, \sigma_{lnL}^2 = 0.02 \]

From Eq. (3.21) it follows that

\[ p_F = 1 - \Phi \left( \frac{3.793 - 3.545}{\sqrt{0.027 + 0.02}} \right) = 1 - \Phi(1.138) \approx 0.13 \]

This result may be compared with Eq. (3.14), which expresses the risk for normal variates in terms of the reliability index \( \beta \).
As a numerical example, find the risk if $R$ and $L$ are log-normal variates with the same parameters as in the previous example (3.2), i.e.

$$\mu_R = 45, \sigma_R = 7.5 \text{ and } \mu_L = 35, \sigma_L = 5.0.$$ 

Between the corresponding mean values and variances of the variates $\ln R$, $R$ and $\ln L$, $L$ the following relations apply (Ang and Tang, 1975)

$$\mu_{\ln R} = \frac{1}{2} \ln\left[\frac{\mu_R}{1 + \left(\frac{\sigma_R}{\mu_R}\right)^2}\right] \quad \text{and} \quad \sigma_{\ln R}^2 = \ln\left[\left(\frac{\sigma_R}{\mu_R}\right)^2 + 1\right]$$

$$\mu_{\ln L} = \frac{1}{2} \ln\left[\frac{\mu_L}{1 + \left(\frac{\sigma_L}{\mu_L}\right)^2}\right] \quad \text{and} \quad \sigma_{\ln L}^2 = \ln\left[\left(\frac{\sigma_L}{\mu_L}\right)^2 + 1\right]$$
3.1.2 Second-Moment Formulation

In most cases the probability distribution functions of load and resistance are not known and it is very difficult to obtain complete information about them without substantial further effort.

Usually the available data are scarce and only estimation can be made about the first and second moments of the probability distributions.

\[ E(\ell), E(r) \text{ and variances } \sigma_L^2 = \text{Var}(\ell), \sigma_R^2 = \text{Var}(r) \]

Consider the reduced variables

\[ L' = \frac{L - E(L)}{\sigma_L}, \quad R' = \frac{R - E(R)}{\sigma_R} \]
The failure condition of the system may be expressed in terms of a performance index (safety margin $M$).

\[ L' = \frac{L - E(L)}{\sigma_L}, \quad R' = \frac{R - E(R)}{\sigma_R} \quad (3.22) \]

The critical condition is

\[ M = R - L = 0 \quad (3.23) \]

Then we obtain the limit-state equation

\[ \sigma_R R' - \sigma_L L' + E(R) - E(L) = 0 \quad (3.24) \]

As shown in Fig. (3.6) the geometrical representation of Eq. (3.24) in the space of reduced variates $L'$ and $R'$ is a straight line. This line divides the plane in two parts.
(1) the upper part, where $M < 0$, represents the failure state of the system and

(2) the rest of the plane indicates safe conditions ($M > 0$).

All points on the straight line correspond to values of $R$ and $L$ for which the failure or critical condition is valid (failure line).
The distance $d$ between the origin and the failure line is a measure of the reliability of the system. By means of analytical geometry we obtain:

$$d = \frac{E(R) - E(L)}{\sqrt{\sigma_R^2 + \sigma_L^2}}$$

the distance $d$ is equal to the reliability index $\beta$.

In this case the reliability $R_e$ and the risk $P_F$ may be computed in terms of the distance $d$ as follows:

$$R_e = \Phi(d) \quad \text{and} \quad P_F = 1 - \Phi(d)$$

where $\Phi$ is the cumulative normal function.
The second moment method can be generalized when \( R \) and \( L \) are functions of other variables \( X_i \) of the system.

Take for example the case of a dam break due to broken of the spillway or to an accidental breach.

The load is the inflow rate \( Q_L \), which is a function of the peak flow upstream \( Q_p \).

Resistance is the flow capacity of the spillway \( Q_C \), which is a function of the width \( b \) and the elevation \( z \) of the crest and also the water level \( h \).

\[
Q_C = C \sqrt{2g} b(h - z)^{3/2} \text{ with } C = \text{constant.}
\]
The performance function or state function in this case is

\[ g(X) = Q_L - Q_c = f(Q_p) - C \sqrt{2g b (h - z)^{3/2}} \]

The safe state when \( g(X) > 0 \)  

the failure state when \( g(X) < 0 \)

If \( f_X(X) \) is the joint probability density distribution function, the reliability \( R_e \) and risk \( p_F \) may be computed by

\[ R_e = \int_{g(X)>0} f_X(X) \, dx \quad \text{and} \quad p_F = \int_{g(X)<0} f_X(X) \, dx \]
Frequency Analysis of Data

- The aim is to quantify extreme values of physical variables and associated risks, has been applied traditionally in hydrology.

- The main objective in this case is the definition of the hydrologic risk or the evaluation of the probability of hydrologic exceedance.

- Water engineering structures are designed to operate for a certain period of time. During this life-time period, they should be reliable, i.e. fulfill their purpose and withstand the applied loads.
Frequency Analysis of Data

- Water supply system is designed to satisfy the demand for water supply over 35 years of operation.

- This should be the case irrespective of uncertainties related to various operating conditions of the system, such as increase of population or availability of water supply.

- A flood levee is provided to resist to the largest flood over 50 years of life-time

To evaluate hydrologic risks, or more generally risks caused by natural phenomena, it is important to have large time series information about extreme values of loads and resistances

- Time series of maximum annual flow rate in a river,
- Maximum annual precipitation of a given duration
- Maximum daily pollutant concentration in a river
The *maximum observed annual values* \((X_1, X_2, \ldots, X_N)\) over \(N\) years may be considered to be independent random variables and have the same probability distribution function \(F_X(x)\).
If $x_0$ is a characteristic value, then the annual hydrologic risk or annual risk of exceedance is the probability that the maximum annual value $X$ exceeds $x_0$, i.e. (Fig. 3.8)

$$P(X \geq x_0) = p_F$$

(3.27)

$X = \text{maximum annual value}$

the risk of having a hydrologic failure or incident and it is easily computed in terms of the probability distribution of $X$.

However, most of the time this probability distribution is not known and in any case it is difficult to be estimated.

Fig. 3.8 Annual hydrologic risk of exceedance.
Another method has been developed to evaluate hydrologic risk, which does not make reference to any specific probability distribution of hydrologic loads X. This is the method of the return period.

Let \( T \) a characteristic time period called interval of occurrence or return period to be defined as the number of years until the considered load X equals or exceeds on average a specified value \( x \) only once.

For example the 50-yr flood is by definition the flood which may occur on average only once in 50 years. This does not imply necessarily that the above flood will occur only after 50 years: it may occur next year or several times in the next 50 years or not at all for 100 years.
- Of course, the probability of exceedance of $x$ depends on the interval of time being considered, but the fundamental result is that the annual risk of exceedance is equal to $1/T$.

$$P(X \geq x_T) = p_F = \frac{1}{T}$$

For example the probability that the 50-yr flood may be exceeded in a given year is $1/50$.

-The above result expressed by Eq.(3.28) is the consequence of two main assumptions:
(i) occurrences of random variables $X$ are independent
(ii) the hydrologic events are time invariant.
$P(X \geq x_T) = \text{annual hydrologic risk} = 1/T$

Let $A$ be the exceedance event, i.e.

$A = X \geq x_T$

$\overline{A} = X < x_T$

$T = \frac{V}{\mu} = \text{E}(V)$

$V = T_1 \quad V = T_2 \quad V = T_3$

Time (years)
This means that the probability that A may occur k times in the n future years follows a binomial distribution law \( B(n, k, p_F) \), where

\[
P(A \text{ occurs } k \text{ times over } n \text{ years}) = \binom{n}{k} p_F^k (1 - p_F)^{n-k} = \frac{n!}{k!(n-k)!} p_F^k (1 - p_F)^{n-k} = B(n, k, p_F)
\]

The reliability of the system over n-years is the probability that the T-yr event is not exceeded in that period

\[
R_{e,n} = P(X < x_T)_n
\]

\[
R_{e,n} = P(X < x_T)_n = (P(X < x_T))^n = (1 - P(X \geq x_T))^n
\]

The n-year risk \( p_{F,n} \) is equal to \( 1 - R_{e,n} \). From Eq. (3.35) and using (3.28) we obtain

\[
p_{F,n} = 1 - R_{e,n} = 1 - (1 - P(X \geq x_T))^n = 1 - (1 - \frac{1}{T})^n
\]  

(3.36)

the risk of the occurrence of \( x_T \) at least once in n years.
Example 3.4

Find the risk of occurrence in a period of 5 years of a flood with a return period of 20 years. From (3.36) we have

\[ p_{F,5} = 1 - (1 - \frac{1}{20})^5 = 22\%. \]

Tab. (3.3) gives the risk of exceedance of the T-yr flood in a particular period of n years.

Tab. 3.3 Risk of exceedance (%) of the T-yr flood in n years.

<table>
<thead>
<tr>
<th>T</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33</td>
<td>19</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>63</td>
<td>41</td>
<td>22</td>
<td>10</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>87</td>
<td>65</td>
<td>40</td>
<td>18</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>98</td>
<td>88</td>
<td>64</td>
<td>33</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>99.4</td>
<td>87</td>
<td>60</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>99.6</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>96</td>
</tr>
</tbody>
</table>
**Example 3.5**

The maximum daily concentration of nitrates in a river has been found to follow a probability density distribution of exponential type. The mean daily concentration is 6.08 mg/l and the maximum allowed (for this particular example) is 14 mg/l. It will be assumed that values of daily concentrations are statistically independent of each other. Find

(a) the risk of pollution in a single day
(b) the return period in days
(c) the risk of pollution in the 5 next days
(d) the risk of pollution for the first time on the fifth day after starting measuring
(e) the risk of having exactly one exceedance in 5 days

The general form of an exponential probability density distribution is

\[ f_C(c) = \lambda e^{-\lambda c} \]

where \( c \) is the maximum daily concentration in mg/l and \( \lambda \) a constant equal to \( 1/E(C) \). \( E(C) \) is the mean value of \( C \), equal to 6.08 mg/l. We have, \( \lambda = 1/6.08 \).
(a) the risk of pollution in a single day is the probability that $C$ exceeds the critical value $C_0 = 14$ mg/l. In this example, the probability distribution is known, so that by use of Eq. (3.27) we get

$$p_F = P(C \geq C_0) = P(C \geq 14) = \int_{14}^{\infty} \lambda e^{-\lambda c} dc = e^{-14\lambda} = e^{-\frac{14}{6.08}} = 0.10$$

(b) according to Eq. (3.28) the return period $T$ in days is given by

$$T = \frac{1}{p_F} = \frac{1}{0.10} = 10 \text{ days}$$

(c) the risk of pollution in 5 consecutive days is given by Eq. (3.36) or by Tab. (3.3) and has the following value

$$p_{F,n} = 1 - \left(1 - \frac{1}{T}\right)^n = 1 - (1 - 0.10)^5 = 1 - 0.90^5 = 1 - 0.59 = 0.41$$
This is the probability to have at least one 10-day pollution event in 5 days. We can see that the same result may be obtained if we use Eq. (3.31) to compute the probability that the time $V$ for having an exceedance will be less than or equal to 5 days, i.e.

$$p_{F,n} = P(V \leq 5) = \sum_{k=1}^{5} (0.10)(0.90)^{k-1}$$

$$= 0.10 + (0.10)(0.90) + (0.10)(0.90)^2 + (0.10)(0.90)^3 + (0.10)(0.90)^4$$

$$= 0.10(1 + 0.90 + 0.81 + 0.729 + 0.656) = (0.10)(4.095) = 0.41$$

(d) the probability to have exceedance in the fifth day is

$$P(V=5) = (0.10)(0.90)^4 = 0.0656$$

(e) using of the binomial distribution law (3.33) we have

$$\binom{5}{1}(0.10)(0.90)^{5-1} = \frac{5!}{1!(5-1)!}(0.10)(0.90)^4 = 5(0.10)(0.656) = 0.328$$
Probability Distribution of Extremes

-By use of the extreme-value theory (Gumbel) has shown that in a series of extreme values \( X_1, X_2, \ldots, X_N \), the probability that \( X \) will be less than the \( T \)-yr value \( x_T \) is given by:

\[
P \left( X < x_T \right) = e^{-e^{-y}} = 1 - \frac{1}{T}
\]  

(3.37)

where

\[
y = a \left( x_T - x_0 \right)
\]  

(3.38)

\[
a = \pi / (\sqrt{6} \sigma_X)
\]  

(3.39)

\[
x_0 = \text{E}(X) - 0.45 \sigma_X
\]  

(3.40)

\[
\text{E}(X) = \text{(mean value of } X_i \text{)} = \frac{\Sigma X_i}{N}
\]  

(3.41)

\[
\sigma_X = \text{standard deviation given by } \sigma_X = \sqrt{\frac{\Sigma (X - X_i)^2}{N - 1}}
\]  

(3.42)
Probability Distribution of Extremes

From Eqs. (3.37) and (3.38) we have

\[ x_T = x_0 - \frac{\ln \left\{ \ln \left( \frac{1}{1-(1/T)} \right) \right\}}{a} \]  

(3.43)

For a given sample of values \( X_1, X_2, \ldots, X_n \),
1) first the parameters \( E(X) \) and \( x_0 \) are estimated by use of Eqs (3.41) and (3.42) and
2) then the values of \( x_0, a \) and \( x_T \) for a given return period \( T \) are calculated, as given by (3.40), (3.39) and (3.43).
Analysis of Frequency

- If N observations are given, from which a certain hydrologic event is found to be exceeded m times, then the frequency:

\[ f = \frac{m}{N} \]

- The return period T may be estimated by use of Eq. (3.36) as the inverse of the exceedance probability, i.e.

\[ T = \frac{1}{f} = \frac{N}{m} \]

- In practice, data may be arranged in descending order, starting from the higher maximum annual value.

- If r is the rank of the event being considered and N the total number of observations, then this event has been exceeded r times over N observations. Accordingly the return period T should be:

\[ T = \frac{N}{r} \]
Example 3.6
Rainfall data are available from a meteorological station located in a Mediterranean island. Rainfall is characterized by its duration in min and height in mm. Although the data extend over a limited period of time, which is only 7 years, find the rainfall heights of return period 10, 20 and 50 years for rainfall durations of 15 min and 1 hour respectively (Ganoulis, 1994).

The following Tab. (3.4) summarizes, in descending order, data of maximum annual rainfall heights for given rainfall durations.

<table>
<thead>
<tr>
<th>Rainfall Duration (min)</th>
<th>5'</th>
<th>10'</th>
<th>15'</th>
<th>30'</th>
<th>60'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>16.4</td>
<td>20.3</td>
<td>35.0</td>
<td>50.7</td>
</tr>
<tr>
<td>2</td>
<td>9.5</td>
<td>14.9</td>
<td>18.9</td>
<td>24.3</td>
<td>31.0</td>
</tr>
<tr>
<td>3</td>
<td>8.7</td>
<td>14.3</td>
<td>16.7</td>
<td>22.5</td>
<td>28.2</td>
</tr>
<tr>
<td>4</td>
<td>7.6</td>
<td>13.6</td>
<td>16.1</td>
<td>20.8</td>
<td>27.9</td>
</tr>
<tr>
<td>5</td>
<td>7.4</td>
<td>11.6</td>
<td>14.6</td>
<td>20.5</td>
<td>26.7</td>
</tr>
<tr>
<td>6</td>
<td>7.3</td>
<td>11.4</td>
<td>13.3</td>
<td>18.3</td>
<td>25.4</td>
</tr>
<tr>
<td>7</td>
<td>6.9</td>
<td>10.7</td>
<td>12.6</td>
<td>17.9</td>
<td>25.4</td>
</tr>
</tbody>
</table>
From Eq. (3.45), with \(N=7\) and \(r\) the rank of rainfall, for every maximum annual height, the corresponding return period \(T\) is defined. Tab. (3.5) shows the results for every group of rainfalls having the same duration.

**Tab. 3.5** Observed maximum annual rainfall height \(h(t)\) in mm as function of the return period \(T\) (yr) for rainfall durations of 5, 10, 15, 30 and 60 min.

<table>
<thead>
<tr>
<th>Rainfall Duration (t) (min)</th>
<th>T years</th>
<th>1</th>
<th>1.17</th>
<th>1.4</th>
<th>1.75</th>
<th>2.3</th>
<th>3.5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td>6.9</td>
<td>7.3</td>
<td>7.4</td>
<td>7.6</td>
<td>8.7</td>
<td>9.5</td>
<td>10.0</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>10.7</td>
<td>11.4</td>
<td>11.6</td>
<td>13.6</td>
<td>14.3</td>
<td>14.9</td>
<td>16.4</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>12.6</td>
<td>13.3</td>
<td>14.6</td>
<td>16.1</td>
<td>16.7</td>
<td>18.9</td>
<td>20.3</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>17.9</td>
<td>18.3</td>
<td>20.5</td>
<td>20.8</td>
<td>22.5</td>
<td>24.3</td>
<td>35.0</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>25.4</td>
<td>25.4</td>
<td>26.7</td>
<td>27.9</td>
<td>28.2</td>
<td>31.0</td>
<td>50.7</td>
</tr>
</tbody>
</table>
Tab. 3.5  Observed maximum annual rainfall height $h(t)$ in mm as function of the return period $T$ (yr) for rainfall durations of 5, 10, 15, 30 and 60 min.

<table>
<thead>
<tr>
<th>T years</th>
<th>1</th>
<th>1.17</th>
<th>1.4</th>
<th>1.75</th>
<th>2.3</th>
<th>3.5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall Duration t(min)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.9</td>
<td>7.3</td>
<td>7.4</td>
<td>7.6</td>
<td>8.7</td>
<td>9.5</td>
<td>10.0</td>
</tr>
<tr>
<td>10</td>
<td>10.7</td>
<td>11.4</td>
<td>11.6</td>
<td>13.6</td>
<td>14.3</td>
<td>14.9</td>
<td>16.4</td>
</tr>
<tr>
<td>30</td>
<td>17.9</td>
<td>18.3</td>
<td>20.5</td>
<td>20.8</td>
<td>22.5</td>
<td>24.3</td>
<td>35.0</td>
</tr>
<tr>
<td>60</td>
<td>25.4</td>
<td>25.4</td>
<td>26.7</td>
<td>27.9</td>
<td>28.2</td>
<td>31.0</td>
<td>50.7</td>
</tr>
</tbody>
</table>
From Eq. (3.45), with $N=7$ and $r$ the rank of rainfall, for every maximum annual height, the corresponding return period $T$ is defined. Tab. (3.5) shows the results for every group of rainfalls having the same duration.

![Graph showing rainfall height (mm) versus the return period $T$ (yr) for 15 min rainfall duration.](image)

*Fig. 3.10* Rainfall height (mm) versus the return period $T$ (yr) for 15 min rainfall duration.
Another possibility for finding the relation between $h_T$ and $T$ is the use of a probability distribution function of extreme values, such as Gumbel's law, given by Eq. (3.37). The necessary steps are as follows

(1) using the available data $X_i$ which, for this example, are the maximum annual rainfall heights $h_i$, as given in Tab. (3.4), the mean value $E(h)$ and the variance $\sigma_h$ are computed by applying of Eqs. (3.41) and (3.42)

(2) the parameters $a$ and $h_0$ are estimated by means of Eqs. (3.39) and (3.40)

(3) for every value of return period $T$ the corresponding $T$-yr value of the rainfall height $h_T$ is calculated from Eq. (3.43)

-The maximum height for return periods T=5, 10 and 50 years, which is useful for the design of hydraulic structures, may be estimated from Fig. (3.10).

- It is seen that the Gumbel's distribution law tends to underestimate, at least in this case, the rainfall height for a given return period.

Fig. 3.10 Rainfall height (mm) versus the return period $T$ (yr) for 15 min rainfall duration.
- Data on rainfall height given in Tab. (3.4) have been transformed to rainfall intensity by means of Eq. (3.46) and the results are presented in descending order, in Tab. (3.6).

<table>
<thead>
<tr>
<th>Rank</th>
<th>Duration</th>
<th>5'</th>
<th>10'</th>
<th>15'</th>
<th>30'</th>
<th>60'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>120.0</td>
<td>98.4</td>
<td>81.2</td>
<td>70.0</td>
<td>50.7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>114.0</td>
<td>89.4</td>
<td>75.6</td>
<td>48.6</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>104.4</td>
<td>85.8</td>
<td>66.8</td>
<td>45.0</td>
<td>28.2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>91.2</td>
<td>81.6</td>
<td>64.4</td>
<td>41.6</td>
<td>27.9</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>88.8</td>
<td>69.6</td>
<td>58.4</td>
<td>41.0</td>
<td>26.7</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>87.6</td>
<td>68.4</td>
<td>53.2</td>
<td>36.6</td>
<td>25.4</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>82.8</td>
<td>64.2</td>
<td>50.4</td>
<td>35.8</td>
<td>25.4</td>
</tr>
</tbody>
</table>
By means of Eq.(3.45), the rainfall intensities in Tab. (3.6) are expressed in terms of return period and the results are given in Tab. (3.7).

Tab. 3.7 Observed maximum annual rainfall intensity i (mm/h) as a function of the return period T (yr) for rainfall durations of 5, 10, 15, 30 and 60 min.

<table>
<thead>
<tr>
<th>T years t (min)</th>
<th>1</th>
<th>1.17</th>
<th>1.4</th>
<th>1.75</th>
<th>2.3</th>
<th>3.5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>82.8</td>
<td>87.6</td>
<td>88.8</td>
<td>91.2</td>
<td>104.4</td>
<td>114.0</td>
<td>120.0</td>
</tr>
<tr>
<td>10</td>
<td>64.2</td>
<td>68.4</td>
<td>69.6</td>
<td>81.6</td>
<td>85.8</td>
<td>89.4</td>
<td>98.4</td>
</tr>
<tr>
<td>15</td>
<td>50.4</td>
<td>53.2</td>
<td>58.4</td>
<td>64.4</td>
<td>66.8</td>
<td>75.6</td>
<td>81.2</td>
</tr>
<tr>
<td>30</td>
<td>35.8</td>
<td>36.6</td>
<td>41.0</td>
<td>41.6</td>
<td>45.0</td>
<td>48.6</td>
<td>70.0</td>
</tr>
<tr>
<td>60</td>
<td>25.4</td>
<td>25.4</td>
<td>26.7</td>
<td>27.9</td>
<td>28.2</td>
<td>31.0</td>
<td>50.7</td>
</tr>
</tbody>
</table>
Rainfall intensity - duration curves are given in Fig. (3.11) in natural (a) and log-log (b) coordinates.
Comparison between the Gumbel distribution, and best fit of data is shown in Fig. (3.13) for rainfalls of 15-min and 1-hour duration.
The above example may be applied in areas where flooding due to intensive precipitation may occur.

Two different approaches have been developed to protect urban areas from storm waters.

First is based on the principle that storm water should be kept away from the city and be evacuated as quickly as possible (dams and reservoirs may be designed upstream of urban area to contain the volume of floodwater OR streams may be deviated from urban areas by channels and tunnels, which divert stormwater far from the city).

The second approach provides for storm water to be stored in places within the urban area. The volume of water is then evacuated slowly by the storm-sewer system of the city. The basins are integrated into the activities of the city and can be used as gardens, picnic areas, kindergartens, or stadiums with sport facilities.
- For designing such facilities it is necessary to estimate rainfall intensities and heights of return period 20 or 50 years. As has been shown in the example (3.3), the available rainfall data may be taken into account in two different ways:

(1) using a statistical analysis to quantify the maximum intensity of rainfall of a given duration and return period, and

(2) fitting the data by a probability distribution law of extremes, such as the Gumbel's distribution.
Stochastic modelling

- is a general methodology allowing to introduce probabilities in order to simulate systems, which are subject to uncertainties.

- Such systems may be physical hydrological (coastal areas, rivers, aquifers) or technological engineering systems, (dams, water distribution systems or wastewater treatment plants)

- Stochastic modelling aims to quantify uncertainties in order to assess the risk, which is associated with the use or operation of the system.

- In stochastic modelling for reliability analysis of water quality, the hydrological and water quality variables or parameters are considered as random or stochastic.

-- the stochastic approach is compared with the classical deterministic engineering modelling method.
Deterministic modelling

To understand the stochastic approach it is better to be compared with the classical deterministic engineering modelling method.

A physical hydrological system may be characterized by a set of physical parameters, noted by the vector \( a = \{ a_1, a_2, \ldots, a_k \} \).

According to the deterministic approach, there is a unique relation between the deterministic output \( Y_d \) of the system, the input \( x \) and the vector of system parameters \( a \).

\[
y_d = f(x, a)
\]  

(3.48)
-Deterministic modelling

- More precisely Yd is the model conditional deterministic solution. This solution is subject to two main conditions:
  -(i) the model assumptions, which may be satisfactory for a given set of objectives
  -(ii) the current state-of-knowledge in the field

--The actual or unconditional or true value y may be written in the form:

\[
y = y_d + \varepsilon_d \tag{3.49}
\]

-where \( \varepsilon_d \) is the deterministic deviation or error which, as stated in Chapter 2, paragraph 2.2, is caused by:
  -(a) the intrinsic randomness of the system (aleatory uncertainties)
  -(b) the epistemic or man-induced uncertainties
Deterministic modelling

Epistemic uncertainties include model and parameter uncertainties. These are due to:

- model imperfection and
- the finitness of data (physical meaning of parameters, precision of instruments,
  human errors in obtaining the value of coefficients a
  numerical errors when the relation (3.48) is computed, usually by integration of partial differential equations.

As a typical example, let us consider the flow in a two-dimensional confined aquifer shown in Fig. (3.14).

In this example, the hygrological system is characterized by a single parameter, which is the transmissivity T(m2/s).
- Deterministic modelling

- Inputs are the pumping and recharge flow rates $q_p$ (m$^3$/s/m$^2$) shown in Fig. (3.14), and the boundary conditions (values of piezometric height $h$ (m) or its normal derivative $dh/dn$ along parts of the boundary $S$)

- The output vector may be composed either by both the groundwater flow velocity and the piezometric height $h$, or by the latter only.

*Fig. 3.14 Flow in a two-dimensional confined aquifer.*
Deterministic modelling

- The deterministic model of the problem, based on a set of assumptions and the actual state of knowledge, is composed by the Darcy’s law and the mass continuity equation.

-- Analytical solution for the piezometric height at every point M inside the aquifer is:

\[
h = \int_{S} \left( h^* \frac{\partial h}{\partial n} - h \frac{\partial h^*}{\partial n} \right) dS + \sum_{p} \frac{q_p}{2\pi T} \ln \frac{1}{r_{M_p}}
\]

- This is the deterministic solution of the model, which is a special case of Eq. (3.48) having the form:

\[
h_d = f(q_p, h_s, (\partial h/\partial n)_s, T)
\]
Deterministic modelling

- If in the above solution the integral is computed numerically (Boundary Element Method), then the deterministic solution will contain numerical errors and also any error related with the estimation of parameter T (epistemic uncertainties).

- It will also contain errors due to the intrinsic randomness of the system (spatial variability of the transmissivity T and uncertainties on inputs).

- The deterministic approach is not able to take into account these aleatory uncertainties which, can be handled by the stochastic approach (Fig. 3.15).
Stochastic modelling

In the stochastic approach the physical parameters of the hydrological system A and the inputs X are considered to be *random or stochastic variables, having certain probability distribution functions.*

By use of physical conservation laws (usually in the form of partial differential equations) or with empirical statistical analysis of the available data, a model is first formulated in order to describe the system.

This model, as in the case of a deterministic approach, is subject to two main conditions

(i) the model assumptions, which may correspond to a given set of objectives, and

(ii) the current state-of-knowledge in the field
Stochastic modelling

- In the stochastic approach the physical parameters of the hydrological system $A$ and the inputs $X$ are considered to be *random or stochastic variables*, having *certain probability distribution functions*.

- By use of physical conservation laws (usually in the form of partial differential equations) or with empirical statistical analysis of the available data, a model is first formulated in order to describe the system.

- This model, as in the case of a deterministic approach, is subject to two main conditions
  - (i) the model assumptions, which may correspond to a given set of objectives, and
  - (ii) the current state-of-knowledge in the field
- **Stochastic modelling**

- Using this model the objective of the stochastic approach is to determine the probability distribution law of the dependent variable $Y$ in the form:

\[
P (Y \leq y) = F (X, A, y) \tag{3.50}
\]

The model conditional stochastic value $Y$ may be expressed as the sum between the expected value $<Y>$, given by the probability law (3.50) and the *stochastic deviation* or error $\varepsilon_S$ in the form

\[
Y = <Y> + \varepsilon_S \tag{3.51}
\]

Excepting some specific cases of linear problems, we generally have

\[
y_d \neq <Y> \tag{3.52}
\]
Stochastic modelling

- Using this model the objective of the stochastic approach is to determine the probability distribution law of the dependent variable \( Y \) in the form:

\[
P(Y \leq y) = F(X, A, y)
\]  
(3.50)

The model conditional stochastic value \( Y \) may be expressed as the sum between the expected value \( \langle Y \rangle \), given by the probability law (3.50) and the stochastic deviation or error \( \varepsilon_s \) in the form

\[
Y = \langle Y \rangle + \varepsilon_s
\]  
(3.51)

Excepting some specific cases of linear problems, we generally have

\[
y_d \neq \langle Y \rangle
\]  
(3.52)

By comparing the relations (3.49) and (3.51) and setting, for a given realization \( y \) of the stochastic variable \( Y \), the equality \( y = Y \) one obtains

\[
\varepsilon_d = \langle Y \rangle - y_d + \varepsilon_s
\]  
(3.53)
- **Stochastic modelling**

- Excluding all numerical, experimental and model structural errors, the deviations $\delta d$ and $\delta S$ are due only to the physical uncertainties of the system.

- In this case, it is seen from *Eq. (3.51) that* the stochastic approach furnishes a formal procedure for computing $\delta S$, because both $Y$ and $<Y>$ may be evaluated.

- Also by use of stochastic techniques, some measures of $\delta S$ such as the standard deviation or the confidence interval may be estimated.

- This can be interpreted as an advantage *of* the stochastic approach over the deterministic analysis.
**Stochastic modelling**

Applying the stochastic approach to the example of aquifer flow shown in Fig. (3.14), the transmissivity $T$ or the related hydraulic conductivity $K$ may considered to follow a log-normal probability density distribution in the form:

$$f(k) = \frac{1}{k\sqrt{2\pi}\sigma} \exp\left[-\frac{(\ln k - \mu)^2}{2\sigma^2}\right] \quad k > 0, \quad \sigma > 0, \quad -\infty < \mu < \infty$$

If the inputs may be considered as deterministic or random variables, because of the stochastic character of the parameter $K$, the output variable $h$ should have a probability distribution function. If this is evaluated in form of Eq. (3.50), then the aleatory uncertainties due to the random variation of aquifer parameter should be evaluated.

The state-of–knowledge stochastic modelling of aquifer flow includes multivariate normal distributions and exponential correlation function for the hydraulic conductivity random field.
Stochastic modelling

Fig. 3.15 Qualitative relation between modelling improvement and various uncertainties.
Stochastic modelling

One important question in the stochastic modelling of hydrological systems is the change in the spatial heterogeneity scales. (Furthermore, various methods and tools have been extensively used in the past for stochastic simulation, such as:

- Time series analysis, filtering, krigging
- Stochastic differential equations
- Spectral analysis
- Taylor series and perturbation analysis
- Monte-Carlo simulation
-**Monte Carlo Simulation**

- This is a general simulation technique which may be applied when some random variables are related with deterministic functional relationships.

- In the Monte Carlo method several possible realisations of a random variable would be produced, from which the statistical properties of the variable, such as mean value and variance, are obtained.

- The main point of the technique is to generate samples having a prescribed probability distribution functions.
Monte Carlo Simulation

For reliability computations, the Monte Carlo simulation technique proceeds in three steps:

(1) generation of synthetic samples of random numbers, following specified probability distributions. This may be done for input variables, loads and resistances

(2) simulation of the system by means of a model, where values of generated random variables are taken into account

(3) reliability assessment of the system by counting the number of satisfactory realisations over the total number of realisations. Thus, the probability of success, or the system reliability, may be estimated.
Monte Carlo Simulation

The Monte Carlo simulation technique is a powerful tool, capable of representing complex systems with a non-linear structure.

It is equivalent to the experimental methodology, in which testing of a system is performed by repetition of experiments.

Therefore, the Monte Carlo simulation technique suffers from some drawbacks as any experimental method: lack of insight in the structure of the system and difficulty in making synthesis of the results.

Also, for complex systems, a considerable amount of computing may be necessary and sometimes inconsistent result could be obtained because of sampling variabilities.
- **Assignment 3**

1. The demand for urban drinking water and the supply for drinking water are variable and follow probability distribution with mean values and standard deviation given by the following table:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean Value</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Supply</td>
<td>5</td>
<td>0.75</td>
</tr>
</tbody>
</table>

- Calculate the risk and the reliability of the water supply system assuming that both variables follow
  - a.1) normal distribution
  - a.2) log-normal distribution
-Assignment 3

-2. A groundwater quality monitoring wells recorded the annual maximum pollution concentration (NO3) in Gaza is given in the following table:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>180</td>
<td>234</td>
<td>179</td>
<td>189</td>
<td>435</td>
<td>345</td>
<td>321</td>
<td>432</td>
<td>147</td>
<td>185</td>
</tr>
<tr>
<td>Value</td>
<td>147</td>
<td>254</td>
<td>139</td>
<td>152</td>
<td>305</td>
<td>307</td>
<td>322</td>
<td>225</td>
<td>167</td>
<td>175</td>
</tr>
</tbody>
</table>

-compute:
-(2.1) the risk of pollution in a single year
-(2.2) the return period in years
-(2.3) the risk of pollution in the 5 next years
-(2.4) the risk of pollution for the first time on the fifth year after starting measuring
-(2.5) the risk of having exactly one exceedance in 5 years
-(2.6) compare between the Gumbel distribution and best fit distribution?
3. In your words compare between stochastic and deterministic modeling?