Experimental Measurement of Age Distribution Function

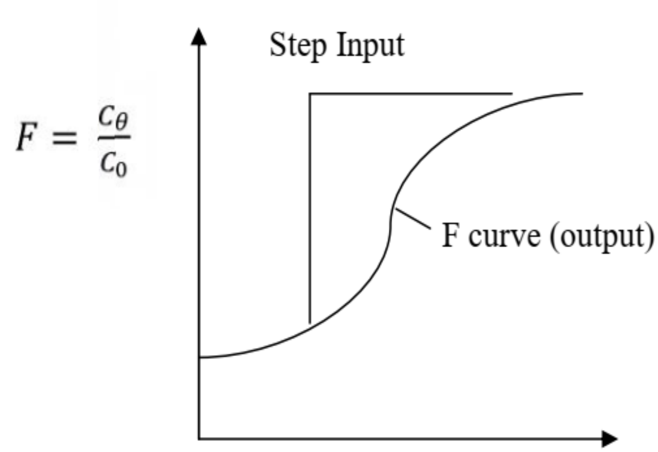
Keywords: RTD, Age Distribution Functions

Experimental Measurement of Age Distribution Function

The experimental determination of the age distribution functions is accomplished for a particular vessel by a stimulus response technique using some sort of tracer material in the inlet fluid stream. The injection is the stimulus and the response is the tracer concentration measured in the outlet stream. The tracer can be a radioactive compound, a colored dye, and electrically conducting salt solution or another material depending on particular situation. It should have the same density, viscosity and other properties as the measure of mixing. Tracer is injected into the inlet stream in some known fashion, such as a step or sudden jump, a pulse, a sine wave other cyclic signal or even a random signal with known properties.

Danckwerts introduced the notation that the dimensionless response to a step injection of tracer be called the F curve and the dimensionless response for an impulse injection be called the C curve.

Suppose that with no tracer initially present a step function (in time) of tracer is introduced into the fluid entering a vessel. Then the dimensionless concentration-time curve for the tracer in the fluid stream leaving the vessel i.e. the F-curve shown in the following figure. F curve rises from 0 to 1.



$$\theta = \frac{t}{\bar{t}}$$

Fig. 13.1: F-Curve

Similarly, suppose an instantaneous pulse or shot of tracer is injected into the entering stream. The dimensionless response of time curve C is as shown in fig. With this choice of variables the area under the C curve is always unity.

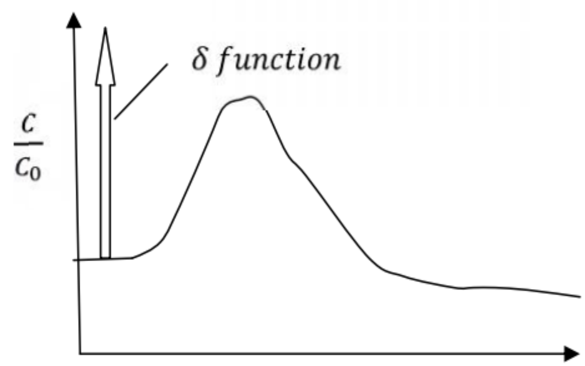


Fig. 13.2: C-Curve

$$\theta = \frac{t}{\bar{t}}$$

$$\int_0^\infty C(\theta)\,d\theta = \int_0^\infty \frac{C(\theta)}{C_0}\,d\theta = 1 \qquad \text{...13.1}$$

$$C^0 = \int_0^\infty C\,d\theta = \frac{1}{\bar{t}}\int_0^\infty C(t)\,dt \qquad \text{...13.2}$$

For a closed vessel there is a simple relationship between the E and C curve and the I and F curve. A closed vessel is defined as one in which there is no back diffusion of any sort at the entrance and exit. Most experimental setups approximately fulfill this requirement since the inlet and outlet pipes are frequently much smaller than the vessel and also the bulk flow is very much larger than any diffusion flux.

Suppose at t=0 an impulse is injected. All tracer elements of fluid have the same starting time for their ages. Thus the outlet concentration –time or C-curve is also a record of the age function of fluid element that entered the vessel at t=0 and left it at t=t, which is precisely the same as E curve.

$$C(\theta) = E(\theta) = \bar{t}E(t) \qquad \ldots 13.3$$

In other words, by injecting a tracer in pulse into fluid flowing into a vessel and determining the ages of fluid element in the output by measuring the tracer concentration in the output stream.

Age Balance:

Amount of Tracer remaining in vessel = Amount of tracer not leaving vessel

or $V\, I(t) = Q\,[1 - F(\theta)]$ $\qquad \ldots 13.4$

or $I(\theta) = 1 - F(\theta)$ $\qquad \ldots 13.5$

Comparing eq. 13.5 with $I(\theta) = \bar{t}I(t) = \int_t^{\infty} E(t')\,dt = 1 - \int_0^{\infty} E(t')dt'd\theta$

$$F(\theta) = \int_0^t E(t')dt' \qquad \ldots 13.6$$

= Fraction of material in exit stream younger than age t

$$F(\theta) = \int_0^t C(\theta)\frac{dt'}{\bar{t}} = \int_0^{\theta} C(\theta')\,d\theta' \qquad \ldots 13.7$$

$$C(\theta) = \frac{dF(\theta)}{d\theta} \qquad \ldots 13.8$$

Age Distribution Functions- Perfect Mixing and Plug Flow

Perfect mixing assumes that the vessel contents are perfectly homogenous and have the same composition as the exit stream. If we consider a step input to a perfectly mixed vessel, a macroscopic material balance gives,

$$\bar{t}\frac{dc}{dt} + c = c_0 \qquad \ldots 13.9$$

$$\frac{c}{c_0} = 1 - e^{-t/\bar{t}} = F(\theta) = 1 - e^{-\theta} \qquad \ldots 13.10$$

Consequently,

$$I(t) = \frac{1}{\bar{t}}[1 - F(\theta)] = \frac{1}{t}e^{-t/\bar{t}}$$

$$I(\theta) = e^{-\theta} \quad \ldots 13.11$$

$$E(t) = \frac{dF(\theta)}{dt} = \frac{1}{t} e^{-t/\bar{t}} \quad \ldots 13.12$$

$$I(\theta) = E(\theta) = e^{-\theta} \quad \ldots 13.13$$

In this case $I(\theta)$ *and* $E(\theta)$ are identical since the fluid within a perfectly mixed vessel has the same composition as that of exit fluid.

The intensity function can be found as follows;

$$\Lambda(t) = \frac{1}{\bar{t}} \frac{E(t)}{I(t)} = \frac{1}{t} \quad \ldots 13.14$$

$$\Lambda(\theta) = 1 \quad \ldots 13.15$$

In plug flow, all material passes through the vessel without any mixing; each fluid element stays in the vessel for exactly the same length of time. For a step input the front or interface between the tracer and non tracer fluids travel down the vessel and come out of the other end in a time equal to main residence time. This $F(\theta)$ curve is a step input curve.

$$F(\theta) = U(t - \bar{t}) \quad \ldots 13.16$$

Where, $U(t - \bar{t}) = 0, t < \bar{t}$

$= 1, t > \bar{t}$

Then,

$$I(t) = \frac{1}{\bar{t}}[1 - F(\theta)] = \frac{1}{t}[1 - u(t - \bar{t})] \quad \ldots 13.17$$

$$or\ E(t) = \frac{dF(\theta)}{dt} = \frac{dU(t-\bar{t})}{dt} = \delta(t - \bar{t}) \quad \ldots 13.18$$

$$I(\theta) = 1 - u(\theta - t) \quad \ldots 13.19$$

$$E(\theta) = \delta(\theta - 1) \quad \ldots 13.20$$

The intensity function for plug flow is,

$$\Lambda(t) = \frac{\delta(t-\bar{t})}{1-U(t-\bar{t})} \quad \ldots 13.21$$

$$\Lambda(t) = 0, 0 \le t < \bar{t} \quad \} \quad \ldots 13.22$$

$$= \quad \infty, t = \bar{t}$$

$$\left.\begin{aligned}\Lambda(\theta) &= 0, 0 \leq \theta < 1 \\ &= \infty, \theta = 1\end{aligned}\right\} \quad \ldots 13.23$$

General Population Balance

Keywords: Models, General Population Balance

General Population Balance

Consider a distribution of entities, $\Psi(x, y, z, \xi_{1,}\xi_2, \dots \dots \dots, \xi_m, t)$, where x, y, z represents ordinary special coordinates, t is the time, and ξ_i represents the ith property of the entity. For example ξ_1 could be the area of a fluid element, or the activity or area of catalyst particle in a reactor.

$$\Psi \Delta x \Delta y \Delta z \Delta \xi_1 \dots \dots \dots \dots \Delta \xi_m \quad \ldots 14.1$$

is the fraction of entities in the geometric volume elements $\Delta V = \Delta x \Delta y \Delta z$, with property values in the range of $\Delta \xi_1, \Delta \xi_2, \dots \dots \dots .. \Delta \xi_m$. Since all the entities must be in some region and have some property value,

$$\int \Psi(x, y, z, \xi_{1,}\xi_2, \dots \dots \dots, \xi_m, t) dx, dy, dz, d\xi_1, \dots . d\xi_m = 1 \quad \ldots 14.2$$

The net generation term will consist of birth and death functions;

$$B = \frac{birth\ of\ entities}{(unit\ time)(unit\ geometric\ volume)(unit\ property\ change)} \quad \ldots 14.3$$

$$D = \frac{death\ of\ entities}{(unit\ time)(unit\ geometric\ volume)(unit\ property\ change)} \quad \ldots 14.4$$

There are z+m independent variables that can be thought of as a z+m dimensional space. For an arbitrary small volume element in this space, R, the balance is,

Accumulation = Net Generation

$$\frac{d}{dt} \int \Psi dR = \int (B - D) dR \quad \ldots 14.5$$

where $dR = dxdydzd\xi_1 \dots \dots . d\xi_m$

For one dimension,

$$\frac{d}{dt}\int_{\alpha(t)}^{b(t)} f(x,t)dx = \int_{\alpha(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}dx + f[b(t),t\}\frac{db(t)}{dt} - f[\alpha(t),t]\frac{d\alpha(t)}{dt}$$

$$= \int_{\alpha(t)}^{b(t)} \left\{\frac{\partial f(x,y)}{\partial t} + \frac{d}{dx}\left[\frac{dx}{dt}f(x,t)\right]\right\} dx \qquad \ldots 14.6$$

$$\int \frac{\partial \Psi}{\partial t} + \frac{\partial(v_x\Psi)}{\partial x} + \frac{\partial(v_y\Psi)}{\partial y} + \frac{\partial(v_z\Psi)}{\partial z} + \sum_{i=1}^{m} \frac{\partial}{\partial \xi_i}\{(v_i\theta) + D - B\}dR = 0 \qquad \ldots 14.7$$

$$\frac{\partial \Psi}{\partial t} + \frac{\partial(v_x\Psi)}{\partial x} + \frac{\partial(v_y\Psi)}{\partial y} + \frac{\partial(v_z\Psi)}{\partial z} + \sum_{i=1}^{m} \frac{\partial}{\partial \xi_i}\{(v_i\theta) + D - B\} = 0 \qquad \ldots 14.8$$

Eq.(14.8) is the general microscopic population balance in x,y,z coordinates.
In many cases the described spatial dependence of $\Psi$is not known and the average values of the properties are required. Then a very useful balance can be obtained from eq. 14.8 by integrating it over. The geometric volume V in vector notation is,

$$\int \left[\frac{\partial \Psi}{\partial t} + \nabla.(\vec{v}\Psi) + \sum_{i=1}^{m} \frac{\partial}{\partial \xi_i}\{(v_i\theta) + D - B\}\right] dxdydz = 0 \qquad \ldots 14.9$$

The geometrically averaged distribution function of interest is

$$\bar{\Psi} = \frac{1}{V}\int \Psi\, dV \qquad \ldots 14.10$$

$$\int \frac{\partial \Psi}{\partial t} dV = \frac{d}{dt}\int \Psi\, dV - \int \nabla.(v_s\Psi)dV = \frac{d}{dt} \qquad \ldots 14.11$$

The second line was obtained by the use of Gaussian divergence theorem. The term v_s stands for the velocity of any part of the surface S, containing the volume V and n is a vector normal to the surface pointing outward. Combing the result we get,

$$\int \left(\frac{\partial \Psi}{\partial t} + \nabla.v\Psi\right) dv = \frac{d}{dt}\int \Psi dv + \int n.(v - v_s)\Psi \qquad \ldots 14.12$$

Finally, the surface integral in eq.(14.12) can be broken up into two parts;

a) The integral over the inlet and outlet pipe surfaces, S_1 and S_2.
b) The integral over the remainder if the surface,$S^1 = S - S_1 - S_2$

$$\int_S n.(v - v_s)\Psi ds = \int_{S_1+S_2} n.(v - v_s)\Psi ds + \int_{S_1} n.(v - v_s)\Psi ds = -\int_{S_1} v_1\Psi_1 ds_1 + \int_{S_2} v_2\Psi_2\, ds_2 + 0 \qquad \ldots 14.13$$

where the minus sign appears in the first term because n was directed outwards and the flow at S_1 is inward.

Putting all of these terms back into Eq. 14.9 as using the definition of Ψ gives the final microscopic balance.

$$\frac{1}{V}\frac{\partial}{\partial t}(V\Psi) + \sum_{i=1}^{m}\frac{\partial(v_i\overline{\Psi})}{\partial \xi_i} + \overline{D} - \overline{B} = \frac{1}{V}[Q_{in}\Psi_{in} - Q_{out}\Psi_{out}] \qquad \ldots 14.14$$

Combined Models

Keywords: Model, Combined Model

Combined Models

We are here concerned with more complicated models using more than one parameter. The one parameter dispersion model works well for simple cases of uniform flow, such as single phase flow in packed beds, and empty tubes, but is not adequate to provide a representative model for stirred tanks, fluidised beds, and the like.

Each of the individual small regions is usually restricted to the simple cases of one of the following:

a) Plug flow regions
b) Perfectly mixed regions
c) Single Parameter dispersion model
d) Dead space

The type of flow streams included are;

- cross flow or exchange of fluid elements
- bypassing flow
- Recycle from end to beginning of a region

In order to understand these concepts consider the model for real stirred tanks proposed by Cholette and Clouteir. It consists of an “active” perfectly mixed region or completely dead region with no transfer to the other region and a certain fraction of the feed bypassing both regions as shown in the fig:

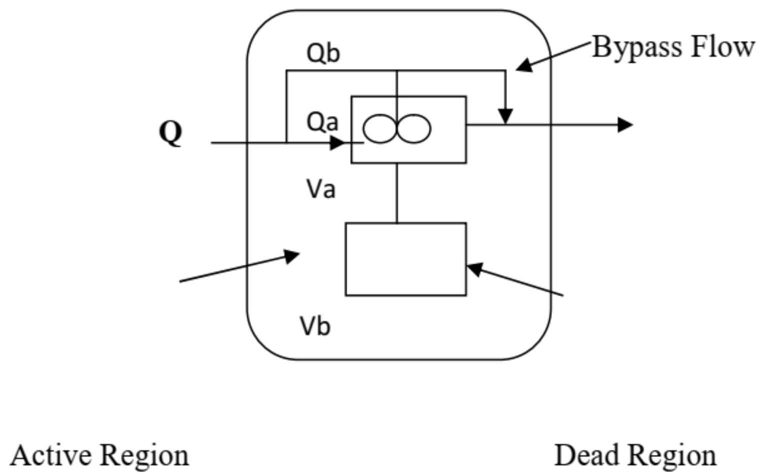


Fig. 15.1: Real stirred tank reactor

- In attempting to represent flow in a real vessel by combined models contain completely stagnant fluid; we meet certain difficulties .Two definitions of dead region are as follows; In a vessel the dead regions are the relatively slow mixing portions of the fluid which we chose to consider tobe completely stagnant. Dead regions contribute to the vessel volume; however we ignore these regions in determining various age distributions.
- Instead of considering fluid in dead regions to be completely stagnant an alternate view imagines that there is a slow interchange or cross flow between the fluid in dead regions and the active fluid passing through the vessel .With this approach, Alder and Hovoka formulated the combined model. This model has four parameters;

  - the cross flow rate
  - the number of units in series
  - two parameters for the ratio of sizes of three types of regions
  - Definition 2 is more representative of the process but the mathematical modeling requirements to solve the models based on it are more complicated.

General Rules among the parameters of combined models;

Levenspiel has devised certain general relations between the parameters of any consistent combined models .

1) Regions in Series

For three regions 1,2,... connected in series, the mean age of the vessel contents is,

$$\bar{t}_I = \bar{t}_{I,1}\frac{V_1}{V} + (\bar{t}_{E,1} + \bar{t}_{I,2})\frac{V_2}{V} + (\bar{t}_{E,1} + \bar{t}_{E,2} + \bar{t}_{I,3})\frac{V_3}{V} + \cdots \qquad \ldots 15.1$$

while the mean age of the fluid in the exit stream is,

$$\bar{t}_E = \bar{t}_{E,1} + \bar{t}_{E,2} + \cdots \ldots \qquad \ldots 15.2$$

2) Regions in Parallel;

For flow regions 1,2,....connected in parallel, the mean age of the vessel is,

$$\bar{t}_I = \bar{t}_{I,1}\frac{Q_1}{Q} + \bar{t}_{I,2}\frac{Q_2}{Q} + \cdots \ldots \qquad \ldots 15.3$$

The mean age of fluid in the exit stream is

$$\bar{t}_E \frac{V_1+V_2+\cdots}{Q} \qquad \ldots 15.4$$

3) Number of parameters in a model

The number of parameters in a model is an indication of its flexibility in filtering a wide variety of processes and suggests the complexity of the mathematical solution.

In general, the number of parameters in a combined model is

$$\begin{bmatrix} Number\ of \\ Parametrs \end{bmatrix} = \sum \begin{bmatrix} Flow\ regions\ in\ excess \\ of\ one \end{bmatrix} + \sum \begin{bmatrix} Flow\ paths\ in\ excess \\ of\ one \end{bmatrix}$$

$$+\sum[Zones\ of\ Cross\ flow] + \sum \begin{bmatrix} Flowregions \\ with\ dispersion \\ \end{bmatrix} -$$

$$\sum[Arbitary\ restrictions\ on\ flow\ and\ volume\ ratios] \qquad \ldots 15.5$$

$$V = \sum V_m + \sum V_p + \sum V_d \qquad \ldots 15.6$$

The volumetric flowrate of streams in parallel are designated by Q_1, Q_2 ....

If Q is the flow rate of fluid to the vessel,

$$Q = Q_1 + Q_2 + \cdots \ldots \ldots \qquad \ldots 15.7$$

Model for Real Stirred Tanks

The F curve can easily be derived by simple material balances. The total flow balance is;

$$Q = Q_a + Q_b \qquad \ldots 15.8$$

$$V = V_a + V_b \qquad \ldots 15.9$$

The concentration of tracer in the entire region is found from the solution of the following differential equation;

$$V_a \frac{dC_a}{dx} = Q_a C_o - Q_a C_a \qquad \ldots 15.10$$

with the initial condition,

$$C_o(0) = 0$$

where

$$C_o = feed\ concentration$$

The solution of eq.(15.10) is

$$\frac{C_a}{C_0} = 1 - e^{-\frac{Q_a t}{V_a}} \qquad \ldots 15.11$$

The outlet response is found from a material balance at the junction of the bypass and active streams.

$$QC_{out} = Q_a C_a + Q_b C_o$$

$$F = \frac{Q_a}{Q}\left(1 - e^{\frac{-Q_a t}{V_a}}\right) + \frac{Q_b}{Q} \qquad \ldots 15.12$$

$$I(\theta) = 1 - F = \frac{Q_a}{Q} e^{\frac{-Q_a V \theta o}{V_a}} \qquad \ldots 15.13$$

The final term in equation (15.13) represents the impulse response that would occur in the outlet caused by bypassing if an inlet impulse were introduced into the system.

The two parameters, the flow split and the volume ratio can be evaluated by tracer experiments. From Cholette and Clouteir model,

$$lnI(\theta) = ln\frac{Q_a}{Q} - \left(\frac{Q_a}{Q}\frac{V}{V_a}\right)\theta \qquad \ldots 15.14$$

Probablistic Models

Keywords: Model, Probabilistic Model

Cholette & Cloutier

They proposed a model for real CSTR. It consists of an active perfectly mixed region and a completely dead region with no transfer of material to the other region and a certain fraction of the feed by passing both regions.

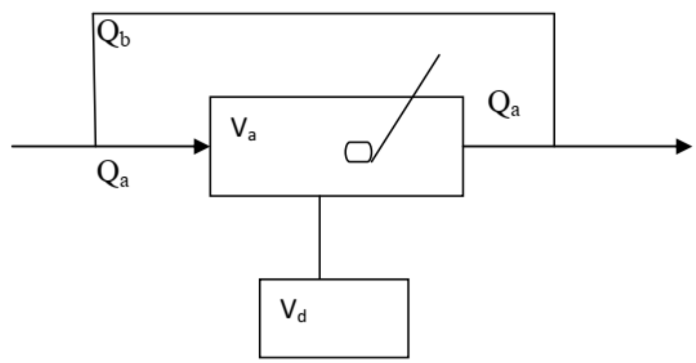


Fig. 16.1: Model of Real CSTR

Probabilistic Approach:

Consider that at t=0 a step input of the tracer is given to the reactor. Instead of determining the concentration of tracer in outlet stream assume that a tracer fluid element enters the system at t=0 either it may go in the by pass stream or to active region. It resides there. For a random time and leaves the system here there are three states or tracer fluid element where it can reside

State -1 Active region

State- 2 Outlet stream

State 3 By pass stream or region

On entering the vessel tracer fluid element may go to state 2 through state 1 or state 3 depending upon the initial probability distribution (EQ 16.1) or we can say that at any time tracer fluid element may be in any of the three states while timer passes continuously. So, logical description of the passage of tracer fluid element through the system may be considered as a discrete state continuous time stochastic process.

Assumptions:

1. Flow rate is steady so λ_i is constant (i=1. 3) where λ is defined as mean frequency of passage of element through the ith stage.
2. At time t element is in the ith stage then the probability that at time $(t+\Delta t)$ it leaces the ith stage and goes to newt i.e. stage 2 is $\lambda_1 \Delta t$
3. At time t tracer fluid element in the ith stage then the probability that at time $(t+\Delta t)$ it remains there is $(1-\lambda_1 \Delta t)$
4. Both the stages (1&3) are considered to behave like a CSTR
5. System considered is a closed vessel

Initial conditions:

$$P1(0) = \begin{cases} a \; ; \; i=1 \\ 1-a \; ; \; i=3 \\ 0 \; ; \; \text{otherwise} \end{cases} \quad \ldots 16.1$$

a = fraction of feed going to active zone

$$= Q_a/Q$$

Our object is to compute P2(t) i.e. probability that the element has left the system P2(t) is the fact F9T) i.e. cumulative residence time of the system.

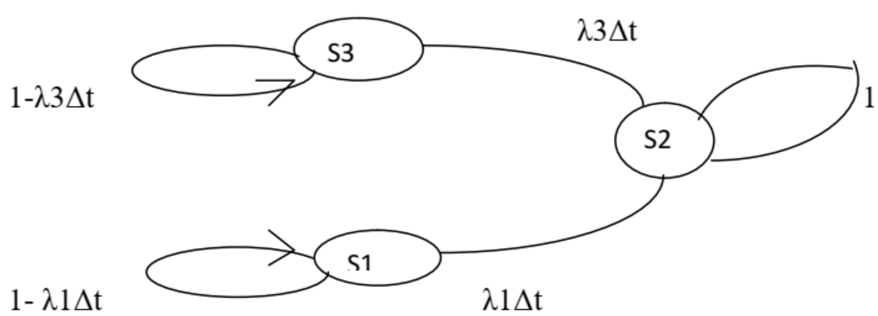


Fig. 16.2: Probabilities of Tracer Element

It is desired to evaluate $P1(t+\Delta t)$ only possibility is that at time t tracer fluid element was in state 1 and in the transition it remain there. Hence

$$P1\,(t+\Delta t) = P1(t)\,(1-\lambda_1 \Delta t)$$

Similarly

$$P3(t+\Delta t) = P3(t)\,(1-\lambda_3\Delta t)$$

$$P2(t+\Delta t) = P2(t) + \lambda_1 P1(t)\,\Delta t + \lambda_1 P3(t)\,\Delta t$$

$$\frac{dP_1(t)}{dx} = -\lambda_1 P_1(t)$$

$$\frac{dP_3(t)}{dx} = -\lambda_3 P_3(t)$$

$$\frac{dP_2(t)}{dx} = \lambda_1 P_1(t) + \lambda_3 P_3(t)$$

Take Laplace transformation

$$S\,L1(s) - a = -\lambda_1 L1(s)$$

$$sL3(s) - (1-a) = -\lambda_3 L3(s)$$

$$s\,L2(s) = \lambda_1 L1(s) + \lambda_3 L3(s)$$

$$L_2(s) = \frac{(1-a)\lambda_3}{s(s+\lambda_3)} + \frac{a\lambda_1}{s(s+\lambda_1)}$$

Taking inverse L.T.

$$P_2(t) = F(t) = (1-a)\left[\,1 - a^{-\lambda_3 t}\right] + a\left[\,1 - a^{-\lambda_1 t}\right]$$

It is reported in the literature that if some the fluid passes through the vessel in a time one tenth of mean residence time of the overall fluid for all practical purposes two fluid can be said to bypass the vessel. So for example if 0.1 is the fraction of the total fluid by passing the vessel then volume of by pass region should be less than or equal to 1/100 of the total volume of the vessel

which is negligibly small Due to this reason volume of bypass stream has not been shown in fig 1 since λ_3 is the ratio of Q_b to volume of bypass region therefore here λ_3 may be assumed as approaching infinity. In case the volume of bypass region is of appreciable magnitude then it may be taken into account accordingly

$$Lim_{\lambda_3 \to \infty}\left[1 - e^{-\lambda_3 t}\right] = U(t)$$

Now Eq. 6 becomes

$$F(t) = (1 - a)U(t) + a[1 - a^{-\lambda_1 t}]$$

$$F(\theta) = \frac{Q_b}{Q} U(\theta) + \frac{Q_a}{Q}\left[1 - e^{-\frac{Q_a}{Q}\frac{V}{V_a}\theta}\right]$$

Hence

$$E(\theta) = \frac{dF(\theta)}{d\theta} = \frac{Q_b}{Q} \partial(\theta) + \left(\frac{Q_a}{Q}\right)^2 \frac{V}{V_a}\left[e^{-\frac{Q_a}{Q}\frac{V}{V_a}\theta}\right]$$

Equations are identical to those obtained by deterministic theory.

Empirical Models

Keywords: Curve Fitting, Linear System, Non-Linear System

Curve Fitting

In experimental work we often face the problem of fitting a curve to data which are subject to error usually a mathematical equation is fitted to experimental data by plotting the data on a graph paper and then passing a straight line through the data points the method has the obvious drawback in that the straight line drawn may not be unique. The method of least squares is probably the most systematic procedure to fit a unique curve through given data points and is widely used in practical computational. It can also be easily implemented on a digital computer.

Let the set of data points be (xi, yi) i= 1, 2, … m and let the curve given by Y= f(x) be fitted to this data at x= xi the experimental value of the ordinate is yi and the corresponding value on the fitting curve is f(x). IF ei is the error of approximation of x=xi then we have

$$ei = yi - f(xi)$$

If we write

$$S = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 \ldots\ldots [y_m - f(x_m)]^2$$

$$= e_1^2 + e_2^2 + \cdots \ldots \ldots \ldots \ldots \ldots . e_m^2$$

Then the method of least squares consists in minimizing S i.e. the sum of the squares of the errors. In the following sections we shall discuses the linear and nonlinear least square fitting to given data.

Fitting a straight Line:

Let Yi represent an experimental value and let Yi be a value from equation

yi= axi +b

Where xi is particular value of the variable assumed free of error. We wish to determine the best values for a and b so that the y's predict the function values that correspond to x values. Let ei = Yi-yi. The least squares criterion requires that

$$S = e_1^2 + e_2^2 + \cdots \ldots \ldots \ldots \ldots \ldots . e_N^2$$

$$= \sum_{i=1}^{N} e_i^2$$

$$= \sum_{i=1}^{N} (Yi - yi)^2$$

$$= \sum_{i=1}^{N} (yi - axi - b)^2$$

Be a minimum. N is the number of x,y pairs. We reach the minimum by proper coice of the parameters a and b so they are the variables of the problem. Ata minimum for S the two partial derivative$\frac{\partial s}{\partial a}$ and $\frac{\partial s}{\partial b}$ will be both be zero. Hence remembering that the xi and Yi are data paints unaffected by our choice of values

$$\frac{\partial s}{\partial a} = 0 = \sum_{i=1}^{N} 2\,(Yi - axi - b)(-xi)$$

$$\frac{\partial s}{\partial b} = 0 = \sum_{i=1}^{N} 2\,(Yi - axi - b)(-1)$$

Dividing equations by -2 and expanding the summation, we get the so calles normal equations:

$$a \sum xi^2 + b \sum xi = \sum xi\,Yi$$

$$a \sum xi \; + b\,N = \sum \; Yi$$

All the summations in equation are from i =1 to i=N. Solving these equation simultaneously gives the values for slope and intercept a & b. Multiplying equation by N and equation by $\sum xi$

$$a\,N \sum xi^2 + b\,N \sum xi = (\sum xi\,Yi)N$$

$$a \sum xi^2 + b\,N \sum xi = \sum Yi \sum xi$$

Subtracting equation

$$a \left[N \sum xi^2 + \left(\sum xi\right)^2\right] = N \sum xiYi - (\sum Yi) \sum xi$$

$$a = \frac{N \sum xiYi - (\sum Yi) \sum xi}{[N \sum xi^2 + (\sum xi)^2]}$$

From equation

$$b\,N = \sum Yi - a \sum xi$$

$$\text{Or } b = \frac{\sum Yi}{N} - a\,\frac{\sum xi}{N}$$

$$b = \frac{\sum Yi}{N} - \frac{\sum xi}{N}\left[\frac{N \sum xiYi - (\sum Yi) \sum xi}{[N \sum xi^2 + (\sum xi)^2]}\right]$$

Non linear curve fitting

We consider a power function a polynomial of te nth degree and an exponential function to fit the given data points

(xi yi) , I = 0, 1, 2 .. m

(a) Power function :

Let $y = ax^c$ be the function to be fitted to the given data taking logarithmic of both sides, we obtain the relation

ln y = ln a +c ln x

which is of the form Y = ao + a1X

where Y = ln y, ao = ln a, a = c and X= log x. Now we can use the same procedure as in the case of straight line.

(b) Polynomial of the nth degree:
Let the polynomial of the nth degree be $Y= a_o +a_1 x +a_2 x^2 a_3 x^3 + \dots \dots$ an x^n

Be fitted to the data points (xi, yi), i=1,2. . . n

We than have

$S= [y_1 – (a_o +a_1 x +a_2 x^2 a_3 x^3 + \dots \dots \text{an } x^n)]^2 + [y_2 – (a_o +a_1 x +a_2 x^2 a_3 x^3 + \dots \dots \text{an } x^n)]^2 +$
$\dots [y_m – (a_o +a_1 x +a_2 x^2 a_3 x^3 + \dots \dots \text{an } x^n)]^2$

For the minimum value of S the partial dericatives of S wrt a_o, a_1 , a_2 a_3 . . . a_m

To be zero

$$\frac{\partial S}{\partial a_0} = 0$$

$$\frac{\partial S}{\partial a_1} = 0$$

$$\frac{\partial S}{\partial a_2} = 0$$

….

…..

….

$$\frac{\partial S}{\partial a_n} = 0$$

$$\frac{\partial s}{\partial a_0} = 0 = \sum [\text{ yi - (ao } + \text{a1 x } + \text{a2 x2 } + \ \dots \dots \text{an xn) }]2\,(-1)$$

$$\frac{\partial s}{\partial a_1} = 0 = \sum [\text{ yi - (ao } + \text{a1 x } + \text{a2 x2 } + \ \dots \dots \text{an xn)}]2\,(-xi)$$

$$\frac{\partial s}{\partial a_2} = 0 = \sum [\text{ yi - (ao } + \text{a1 x } + \text{a2 x2 } + \ \dots \dots \text{an xn) }]2\,(-2x^2)$$

Divide all the equations by -2

$$m\,a_0 + a_1 \sum x_i + a_1 \sum x_i^2 + \dots \dots .+a_n \sum x_i^n = \sum y_i$$

$$a_0 \sum x_i + a_1 \sum x_i^2 + \dots \dots .+a_n \sum x_i^{n+1} = \sum x_i y_i$$

$$a_0 \sum x_i^2 + a_1 \sum x_i^3 + \dots \dots .+a_n \sum x_i^{n+2} = \sum x_i^2 y_i$$

……

…….

$$a_0 \sum x_i^n + a_1 \sum x_i^{n+1} + \dots \dots .+a_n \sum x_i^{2n} = \sum x_i^n y_i$$

These are (n+1) equations in (n+1) unknown and hence can be gives the required polynomial of the nth degree.

(C) Exponential Function:

Let the curve be

$y = a_0\, e^{a1x}$

be fitted to the given data. Then as before taking logarithamics of both sides of equation we get

$\ln y = \ln a_0 + a_1 x$

which can be written in the form

$Z = A + Bx$

Where $z = \ln y$

$A = \ln a_0$

$B = a_1$

The problem therefore reduces to finding a least squares straight line through the given data.

Quantification of Error of linear Regression:

Anyline other than the one computed by $y = a0+a1x$ results in a larger sum of the squares of the residuals thus the line is unique and in terms of our chosen criterion is a best line through the points. A number of additional properties of this fit can be evaluated by examining more closely the way in which residuals were computed.

The derivation of data fitted by expression can be predicted by calculating the standard error of the estimate. The subscript notation y/x designates that the error is for a predicted value of y corresponding to a particular value of x. Also notice that we now divide by n-2 because two data derived estimates a0 and a1 were used to compute Sr thus we have lost two degree of freedom.

The standard deviation quantities the spread of the data quantity the goodness of fit of data. This is particulary useful for comparison of serial regression. To do this we return to the original data and determine the sum of squares around the mean for the dependent variable we can call this total sum of the squares St. This is the amount of spread in the dependent variable that exists prior to regression. After performing the linear regression we can compute Sr which is the sum of the squares of the residuals around the regression line. This represents the spread that remains after regression. The difference between the two quantities or St-Sr quantities the improvement or error reduction due to the straight line model. This difference can be normalized to the total error to yield.

$$r^2 = \frac{S_t - S_r}{S_t}$$

where r is the correlation coefficient and r is the coefficient of determination.

The error of polynomial regression can be quantified by a S.D. error of estimate.

$$S_{y/x} = \left[\frac{S_r}{n-(m+1)}\right]^{\frac{1}{2}}$$

Where m is the order of the polynomial this quantity is divided by n-(m+1) because m+1 data derived coefficients a0 a1…. am used to compute Sr, Thus we have lost m+1 degrees of freedom.

A correlation coefficient is

$$r^2 = \frac{S_t - S_r}{S_t}$$

Algorithm for polynomial Regression:

Step 1 : Input order of polynomial to be fit,m

Step 2 : Input number of data points, n

Step 3 : If n≤m, print out an error message that polynomial regression is impossible and terminate the process. If n>m continue.

Step 4 : compute sums of powers and product as required in polynomial equation

Step 5 : set up these sums of powers and products in the form of an augmented matrix.

Step 6 : Solve the augmented matrix for the coefficients a0, a1, a2. . . . am , using an elimination method

Step 7: Print out the coefficients.

Step 8 : Compute yi and yi- yi

Step 9 : Compute the sum of yi- yi

Step 10 : Compute Sr

Step 11 : Compute S y/x & r2 and r.

Step 12 : Print out the S y/x and r

Multiple Linear Regression:

A useful extension of linear regression is the case where y is a linear function of two or more variables. For example, y might be a linear function of x_1 and x_2. As shown here.

$$y = a_0 + a_1x_1 + a_2x_2$$

such an equation is particularly useful when fitting experimental data were the variable being studied is often a function of two other variables. For this two dimensional case, The regression line becomes a plane.

The best values of coefficients are determined by setting up the sum of the squares of the residuals.

$$Sr = \Sigma_{i=1}^{N}\left[y_1 - \left(a_0 + a_1x_1 + a_2x_2a_3x_3 + \ldots\ldots\ldots a_nx_n\right)\right]2$$

Now diff w.r.t. each of the coefficient

$$\frac{\partial s}{\partial a_0} = -2\,\Sigma[\,\mathrm{yi} - (\mathrm{ao} + \mathrm{a1\ x} + \mathrm{a2\ x2}\,)]2$$

$$\frac{\partial s}{\partial a_1} = -2\,\Sigma[\,\mathrm{yi} - (\mathrm{ao} + \mathrm{a1\ x} + \mathrm{a2\ x2}\,)]2\,(xi)$$

$$\frac{\partial s}{\partial a_2} = -2\,\Sigma[\,\mathrm{yi} - (\mathrm{ao} + \mathrm{a1\ x} + \mathrm{a2\ x2}\,)\,]2\,(x^2)$$

The coefficient yielding the minimum sum of the squares of the residuals are obtained by setting the partial derivatives equal to zero and expecting Eq as a set of simultaneous linear equations

$$n\,a_0 + \sum a_1x_i + \sum a_2x_{2i} = \sum y_i$$

$$a_0\sum x_{1i} + a_1\sum x_{1i}^2 + a_2\sum x_{1i}x_{2i} = \sum x_{1i}y_i$$

$$a_0\sum x_{2i} + a_2\sum x_{2i}^2 + a_1\sum x_{1i}x_{2i} = \sum x_{2i}y_i$$

As a matrix:

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} x_{2i} \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum yi \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{bmatrix}$$

Multiple linear regression for the more general case:

Let the equation be

y = a0+ a1x1 + a2x2 + ... +am xm

The coefficient that minimizes the sum of the squares of the residuals are determined by solving the following equations.

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} & \dots & \dots & \sum x_{mi} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} x_{2i} \dots & \dots & \dots & \sum x_{1i} x_{mi} \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i}^2 & \dots & \dots & \sum x_{2i} x_{mi} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sum x_{mi} & \sum x_{mi} x_{1i} & \sum x_{2i} x_{mi} \dots & \dots & & \sum x_{mi}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ \dots \\ am \end{bmatrix} = \begin{bmatrix} \sum yi \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \\ \dots \\ \dots \\ \sum x_{mi} y_i \end{bmatrix}$$

The standard error of the estimate for multiple linear regressions is formulated as:

$$S_{\frac{y}{x1x2,\dots xm}} = \sqrt{\frac{S_r}{n-(m+1)}}$$

Although there may be certain cases where a variable is linearly related to two or more other variables, multiple linear regression has additional utility in the derivation of power equations of the general form,

$$y = a_0 x_1^{a_1} x_2^{a_2} \dots \dots x_m^{a_m}$$

Such equations are extremely useful when fitting experimental data. In order to use multiple linear regressions, the equation is transformed by taking its logarithm to yield

$$\log y = \log a_0 + a_1 \log x_1 + a_2 \log x_2 \ldots\ldots\ldots + a_m \log x_m$$

$$Y = A_0 + A_1 X_1 + A_2 X_2 + \ldots\ldots\ldots\ldots\ldots + A_m X_m$$

$Y = \log y$ $X_1 = \log x_1$

$A_0 = \log a_0$ $X_2 = \log x_2$

$A_1 = \log a_1$

….

…

$A_m = \log a_m$ $X_m = \log x_m$