Modelling

Why do we need modelling?

1. To enable us to create a realistic representation of some natural object or effect. A description is a model.
2. To provide a parameterised representation of some natural object or effect.

Imagine that a member of the school is missing, and we have to provide the police with a description

Missing Person Form

1. Height: *tall*
2. Features: *handsome*

Data can be:
Categorical: red, green, handsome
Ordinal: short, medium, tall
Interval: 3°C, 10°C
Ratio: 3ft 6in, 6ft 3in.

Missing Person Form

1. Height: *6ft 1in*
2. Weight: *11st 7lbs*
3. Features: *handsome*

Data can be:
Categorical: red, green, handsome
Ordinal: short, medium, tall
Interval: 3°C, 10°C
Ratio: 3ft 6in, 6ft 3in.

Realistic Representation

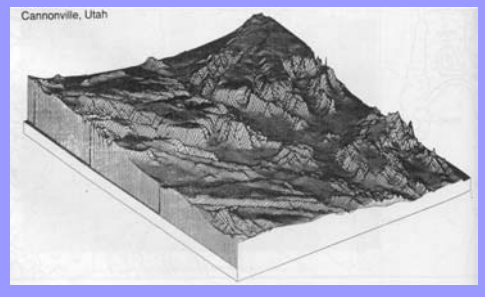
Measurements made by hand or instrument can be converted into a representation of the original object.

The conversion can be quite primitive with values processed point by point.

More sophisticated processing takes groups of points and processes them together with the assumption that something is known about their relationship.

Realistic Representation

Simple processing. The elevations at each point on a grid are simply joined by straight lines.



More Realistic Representation

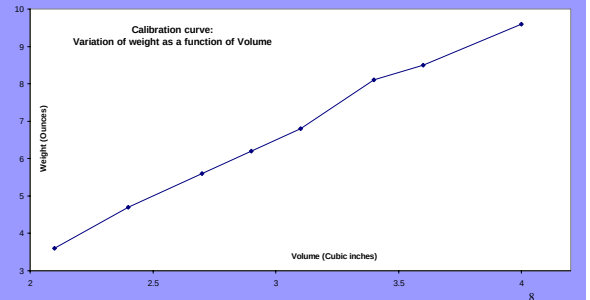
The elevations are processed in groups using splines to create a 'realistic' image. The values of the parameters used in the processing are of no interest – only the final image is wanted.

Determination of the volume of irregular objects of a given material

1. Big objects are heavier than small ones.
2. Determine the weight of some regular objects of known volume.
3. Plot a graph of volume against weight.

Analogue Representation

This calibration curve enables the weight and volume of objects to be related.

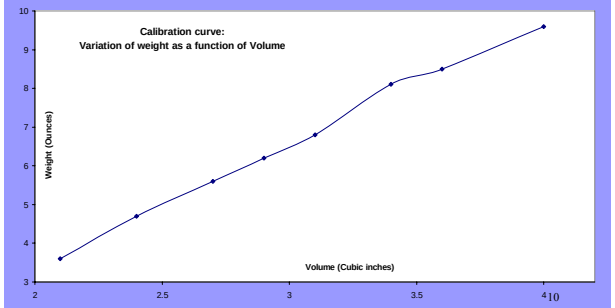


Determining the volume of an object

Weigh the irregular object, and determine its volume from the graph

Analogue Representation

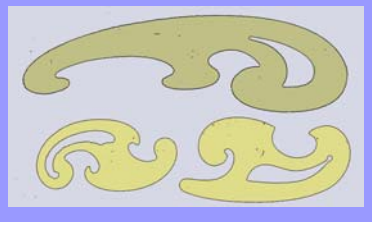
Few phenomena show sharp discontinuities – the points might be better joined with a curve.



10

Some Antique French Curves

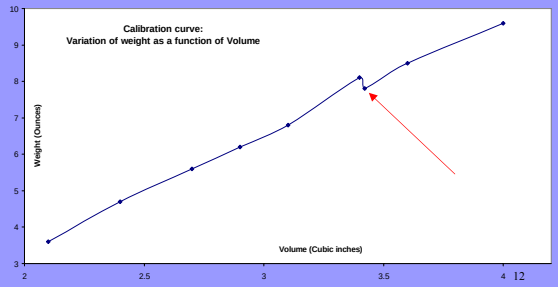
Smooth curves can be drawn by hand using templates (French Curves), or mathematically using splines. A large number of parameters are used in computing the curve, but they are never looked at.



11

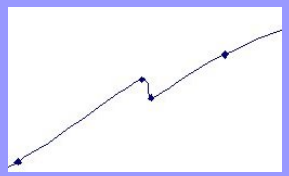
Analogue Representations

A new observation is added to the curve – clearly something is wrong.



12

A Use of French Curves



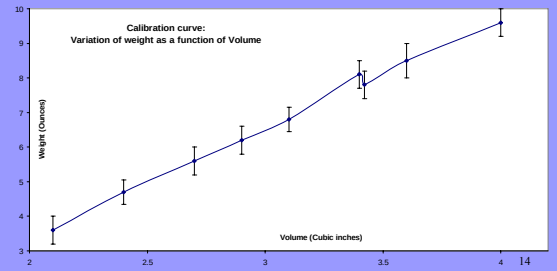
A substantial number of parameters are needed to fit a smooth curve (spline) through these points – is there any point in trying?

We need to see the error bars on the observations

13

Analogue Representations

When we add error bars, it becomes evident that the fluctuations may be artefacts of the measurements.

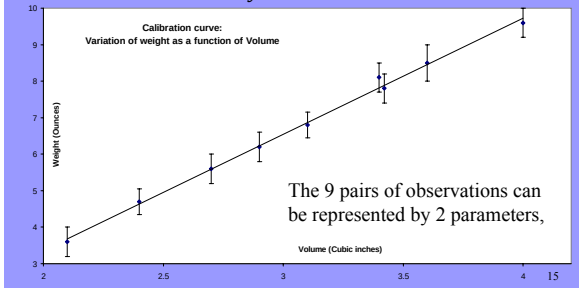


14

Analogue Representations

In fact, a straight trend line passes quite close to all the observations. A good digital representation of the data is

$y = 3.2x-3.0$

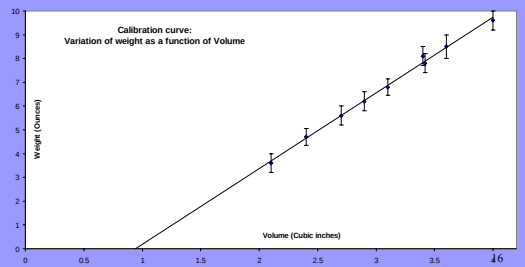


The 9 pairs of observations can be represented by 2 parameters,

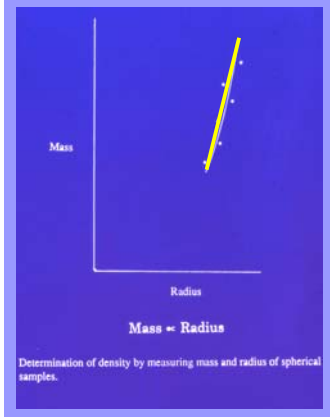
15

Cautions

Extrapolation to small volume samples shows that they have negative weights! Is there a zero-point error, a non-linearity with the instruments, serious under-estimates of the errors or is the relationship really not linear?



16



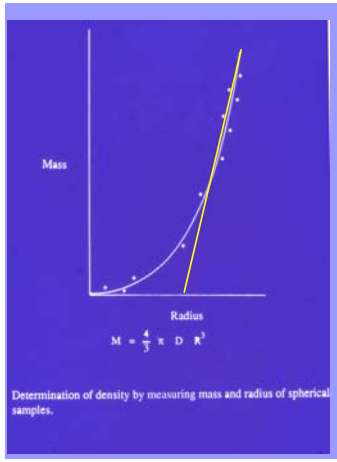
Data Extent

Big spheres are heavier than small ones.

Try plotting weight against radius.

The observational space should be well covered to reduce the risk of using an invalid model

17



Data Extent

High and low order data help confirm/refute the basic model.

A straight line is a poor model

18

Applications in Crystallography

Uses of calibration curves in crystallography:

- Dead time corrections for quantum counters
- Spatial correction files for ccds with tapers
- Dark current correction
- Temperature indications from thermocouples

Strength in Numbers

You can put any kind of curve through one point.

A straight line will pass through 2 points, but so will a sine wave or circle.

The more observations one has, the less the chance of fitting an inappropriate function.

Don't forget that the points have errors.

The better ones knowledge of the model, the less observations needed to parameterise it.

Some Numbers

Imagine a cell 10x10x10 Angstrom

Its volume is 1,000 cubic Angstrom

The volume of an atom is 20 Cubic Angstrom

Thus the cell contains 50 atoms

Atoms can be represented by 9 parameters each.

Thus there are 500 parameters to be optimised.

More Numbers

For a 10 x 10 x 10 A cell, the observable h,k & l max are all greater than 10

and h,k & l min are all less than –10

Thus total number of reflections is 20 x 20 x 10

Number of observations/number of parameters = 4000/500 = 8

This is equivalent to putting a straight line through 16 observations.

Conclusions –what have we learned?

A parameterised model may be a very efficient way of summarising data.

Parameterised models may be concealed in other representations of data.

Over-parameterisation may simply model errors in the data.

Under-parameterisation may fail to represent the data.

Neglect of data from some part of the observational space may lead to problems being over looked.

Demo – size of a piece of paper

Refinement

Why do we need refinement?

1. To enable us to use a model to smooth out errors in the data.
2. To enable us to estimate values for data we cannot measure.
3. To provide a parameterised model for an event or object.

Crystallographic Refinement

To-day, refinement in crystallography usually implies least-squares or a related technique.

However, for difficult problems it is useful to remember modelling and Fourier refinement .

These have different sensitivities to errors on the data and errors in the model.

Lots of Observations

In the experiment we looked at earlier, we saw that the observations could be represented by a straight line.

A straight line can be defined by only two error-free points.

If there are errors, more observations may help to reduce their influence, or enable us to estimate their effect upon the equation (redundancy).

More points near the ends of the line may help define it more precisely (completeness & leverage).

Linear Least Squares

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = y_1$$

The observable quantity *y* is linearly dependent upon the unknown quantities *x*.

We make the observation under the conditions *a*.

We can generally choose the values of a.

In P -1, the structure factor expression is

$$\sum f_n \text{ Cos } 2\pi (hx_n + ky_n + lz_n) = F_{hkl}$$

which is similar to

$$\sum a_n x_n = Y$$

i.e. $a_1x_1 + a_2x_2 = Y$

The Cos term is non-linear

To carry out least squares, we must linearise the expression and rearrange it.

A new Fc equals the original Fc plus a correction.

$$\sum_{i}^{unknowns} a_i \,.\, x_i = y$$

$$\sum_{i}^{atoms} f_i e^{-T} Cos\left(h.x\right) = Fc \approx Fo$$

$$Fc + \sum_{i}^{parameters} \frac{\partial Fc}{\partial x_i} . \delta x_i = F'c \approx Fo$$

$$\sum_{i}^{parameters} \frac{\partial Fc}{\partial x}_i . \delta x_i \approx Fo - Fc$$

Problems with Crystallographic LSQ

The right-hand side is not *Fo*, it is *Fo-Fc*, with *Fc* computed from a model

The left-hand side is a derivative, computed from a model. We cannot choose it.

$$\sum_{i}^{parameters} \frac{\partial Fc}{\partial x}_i . \delta x_i \approx Fo - Fc$$

Problems with Crystallographic LSQ

Because Fc and its derivative occur in the expression, we must have a starting model.

$$\sum_{i}^{parameters} \frac{\partial Fc}{\partial x}_i . \delta x_i \approx Fo - Fc$$

To improve the model we apply shifts δx in order to minimise:

$$\sum (Fo - Fc)^2 \quad \text{or} \quad \sum \left(Fo^2 - Fc^2\right)^2$$

Supplementary Observations

It is not necessary to restrict the observations to just the X-ray observations.

We can also use any other quantities we know or can observe and which are functions of the structural parameters.

These observations are used to set up

Equations of Restraint (Restraints)

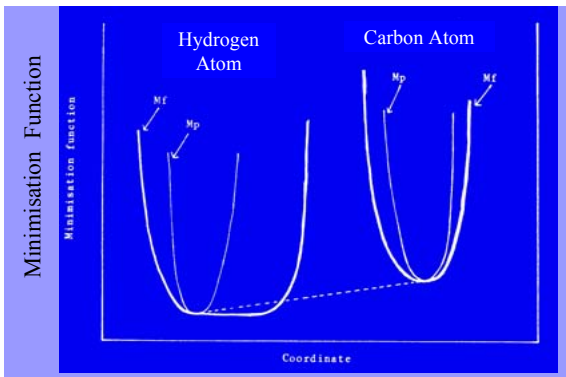
$$\sum_{1}^{n} f_n Cos2\pi(hx_n + ky_n + lz_n) = F_h$$

Compare with

$$\sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2} = d_{a,b}$$

That is

Function of parameters = known value



BOLD = X-ray minima

Faint = restraint minima

Distance Restraint

GEOMETRIC RESTRAINTS

Absolute: We assign a numerical value to a geometrical molecular parameter.

Similarity: We require related quantities to converge to an average value.

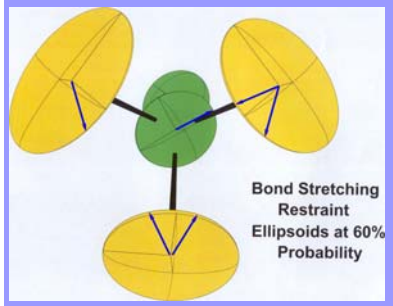
THERMAL RESTRAINTS

Bond Stretching: The components of U along a bond direction are similar.

Rigid Group: The components of U in non-bonded directions are similar.

Similarity: The components of U_{ij} of adjacent atoms are similar.

Thermal Restraints



The components of the adp in selected directions can be set to be similar.

Is Using Restraints Cheating?

It depends upon how they are used.

They can enable a crystallographer to build upon earlier work if the current experiment is not self-sufficient.

They can also be used to support a preconceived idea.

Demo

Demo - Queen Victoria's Penny