UNIVERSITEIT GENT

BIOMATH
Department of Applied Mathematics, Biometrics and Process Control

Mathematical Modelling of the Crystallization Kinetics of Fats: A Methodological View

Peter A. Vanrolleghem, Imogen Foubert and Koen Dewettinck
BIOMATH Dept. Food Technology and Nutrition

93rd AOCS Annual Meeting & Expo,
Montréal, Canada May 5-8 2002

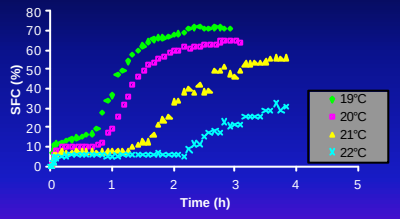
RUG-Biomath, Coupure 653, 9000 Gent, Belgium (e-mail Peter.Vanrolleghem@rug.ac.be)

Why model crystallization kinetics?

- To gain insight in/quantify/predict effects of
  - fat composition
  - differences in fat origin
  - processing conditions
  - temperature (profiles)
  - additives / minor components
- Our focus on:
  - milk fat
  - cocoa butter (Belgian chocolate)

Why model crystallization kinetics? (cont'd)

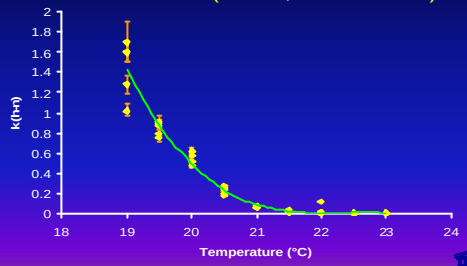
Typical NMR data set



Compare a (small) number of characteristic numbers: e.g. 3 Avrami model parameters (a, k, n)

Example of characteristic number

Temperature effect on crystallization rate constant (EAT-P, Foubert et al.)

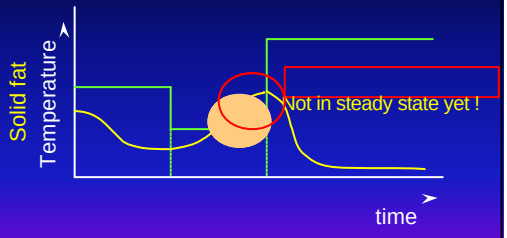


Modelling wish-list

- General:
  - Adequate model to describe/summarize data
  - Meaningful parameters --> physical interpretation
  - Accurate parameter values
  - Statistically correct procedures
  - Confidence information on parameter values
- Specific for our study:
  - Dynamic model for time-varying process conditions

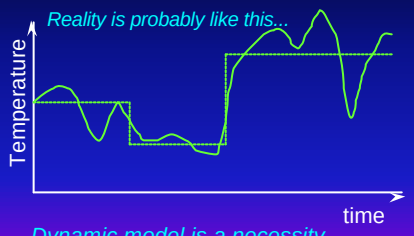
Why a dynamic model ?

Dynamic behaviour is not a sequence of steady states



Why a dynamic model ?

Dynamic behaviour is not a sequence of steady states



Dynamic model is a necessity
--> differential equation model

BIOMATH

Avrami model

- Algebraic equation form

$$f(t) = a * \left(1 - \exp(-k * t^n)\right)$$

- Differential equation form :

$$\frac{df}{dt} = n * k * t^{n-1} * (a - f)$$

$$\frac{df}{dt} = n(T) * k(T) * t^{n(T)-1} * (a(T) - f)$$

k,a = constant

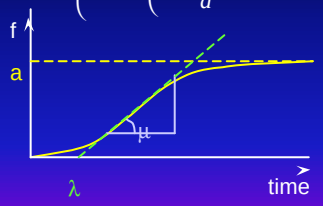
n,k,a = not constant

BIOMATH

Gompertz model (Kloek et al., 2000)

- Algebraic equation form:

$$f(t) = a * \exp\left(-\exp\left(-\frac{m * e}{a} * (t - l) + 1\right)\right)$$



- Differential equation form exists too

BIOMATH

Foubert model (see EAT-P)

- Differential equation form:

$$\frac{dh}{dt} = K * \left(h^n - h\right) \qquad h(t) = \frac{a - f(t)}{f(t)}$$

- Algebraic equation form exists too

Purpose: - select adequate model
- get accurate parameter values

BIOMATH

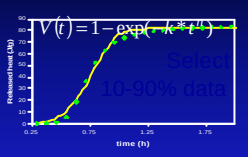
Parameter estimation

- Minimise difference between:
  - data
  - model predictions
- Two approaches:
  - linear regression of transformed model
    - one-step method (on pocket calculator)
  - non-linear regression (with original model)
    - iterative method (e.g. with Excel, Origin, SigmaPlot, WEST)
- Underlying assumptions on measurement errors:
  - independent, normally distributed, constant variance

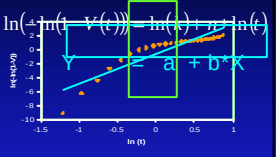
BIOMATH

Linearization of Avrami model

Non-linear regression



Linear regression



Result
(95% confidence intervals)

k = [2.1 – 3.3]

n = [4.3 – 5.5]

Result
(95% confidence intervals)

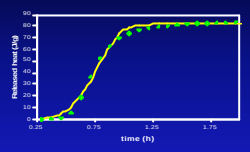
k = [0.4 – 0.8]

n = [3.4 – 5.0]

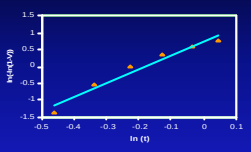
BIOMATH

Linearization of Avrami model

Non-linear regression



Linear regression (10-90%)



Result
(95% confidence intervals)

k = [2.1 – 3.3]

n = [4.3 – 5.5]

Result
(95% confidence intervals)

k = [1.6 – 2.5]

n = [3.3 – 4.9]

BIOMATH

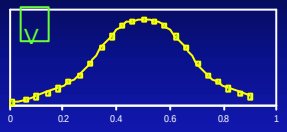
Problems with Linearization

- Data are transformed since the model is rewritten
  e.g V(t) --> ln(-ln(1-V))
- Errors no longer obey statistical requirements for proper linear regression (e.g. normality)

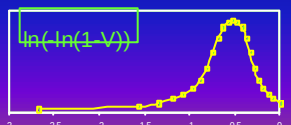
BIOMATH

Problems with Linearization (cont'd)

Measurements of V with *normal* distribution



lead to a *non-normal* distribution of transformed data



BIOMATH

Problems with Linearization (cont'd)

- Data are transformed since the model is rewritten
  e.g V(t) --> ln(-ln(1-V))
- Errors no longer obey statistical requirements for proper linear regression (e.g. normality)

==> biased parameter values

i.e. true value differs from estimated value !

==> no more physical interpretation allowed !

BIOMATH

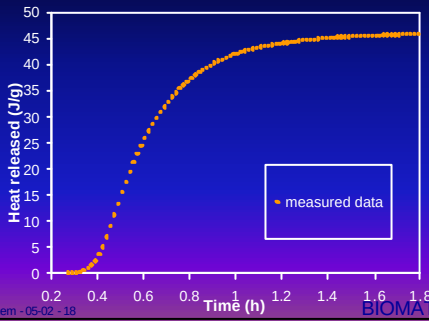
Problems with Linearization (cont'd)

- Calculation of confidence intervals
- Non-linear regression: $f(t) = a * \left(1 - \exp(-k * t^n)\right)$
  – confidence intervals directly on a,k,n
- linear regression: $\ln\left(-\ln\left(1 - V(t)\right)\right) = \ln(k) + n * \ln(t)$
  – confidence interval on n
  – confidence interval on ln(k) --> interval on k ?
  – no information on a because a is hidden in V: $V(t) = \frac{f(t)}{a}$

BIOMATH

Model selection

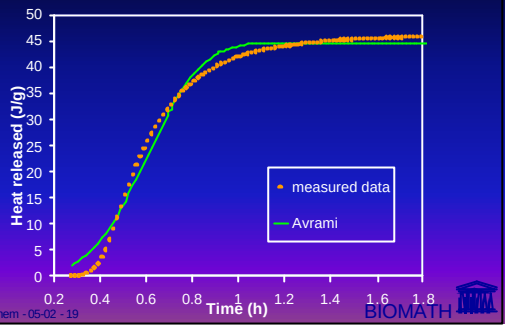
Three models (Avrami, Gompertz, Foubert) are fitted



BIOMATH

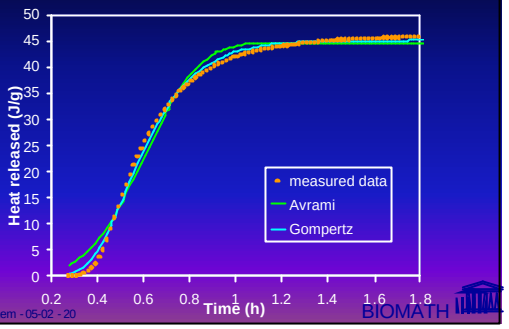
Model selection

Three models (Avrami, Gompertz, Foubert) are fitted



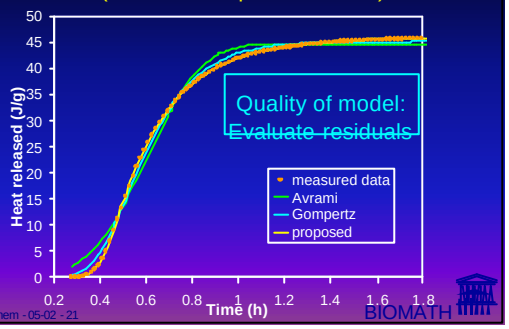
Model selection

Three models (Avrami, Gompertz, Foubert) are fitted



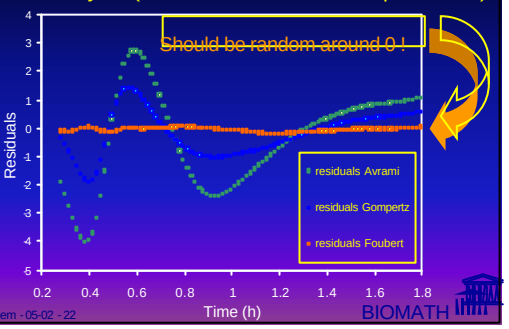
Model selection

Three models (Avrami, Gompertz, Foubert) are fitted



Model selection

Residuals analysis (Residual = data - model prediction)



Residuals analysis

Non-random residuals

⇓

Data are (auto)correlated

= no independent information
= no truly new information per data point

⇓

Calculated confidence limits are too optimistic !

Non-adequate model leads to non-random residuals

Take home

- Modellers wish-list
  - Adequate model to describe/summarize data
  - Meaningful parameters --> physical interpretation
  - Accurate parameter values
  - Confidence information on parameter values
- Linearization
  - biased parameter values
  - difficult to obtain confidence limits on parameters
- Residuals
  - too optimistic confidence limits
- New dynamic models (in differential equation form)