

Recent improvements in the definition and identification of error models for chemical degradation data

Johannes Ranke
Scientific consultant
jrwb.de

Piacenza, 4 September 2019

Outline

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only
Coupled fits

Conclusions

1 Introduction

Outline

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

1 Introduction

2 Results

- Parent only
- Coupled fits

Outline

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

1 Introduction

2 Results

- Parent only
- Coupled fits

3 Conclusions

Observations

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only
Coupled fits

Conclusions

- Ordinary least squares assumes constant variance

Observations

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only
Coupled fits

Conclusions

- Ordinary least squares assumes constant variance
- Metabolite residues sometimes have different variance from parent residues

Residuals over time

Error models
for chemical
degradation
data

J. Ranke

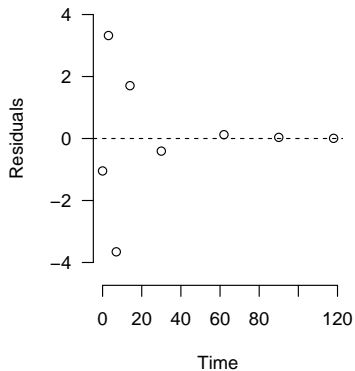
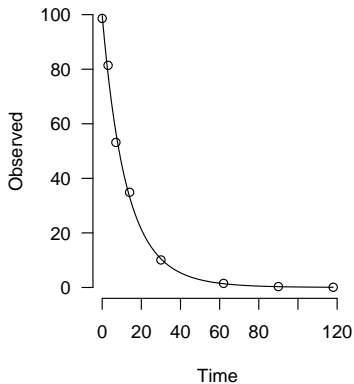
Introduction

Results

Parent only

Coupled fits

Conclusions



Does variance depend on the residue level?

Error models
for chemical
degradation
data

J. Ranke

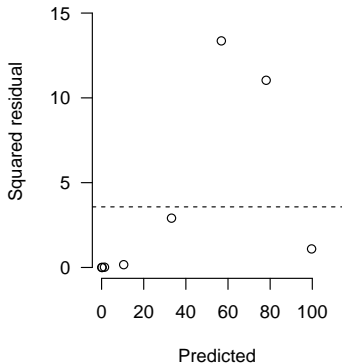
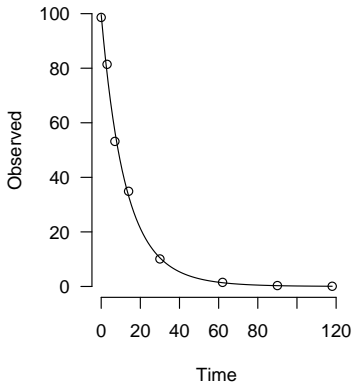
Introduction

Results

Parent only

Coupled fits

Conclusions



Does variance depend on the observed variable?

Error models
for chemical
degradation
data

J. Ranke

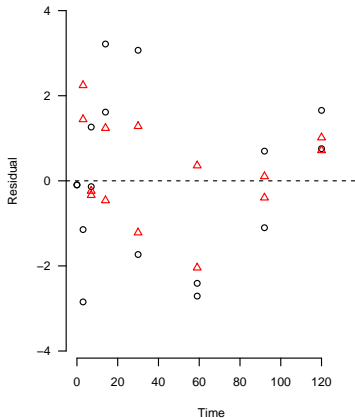
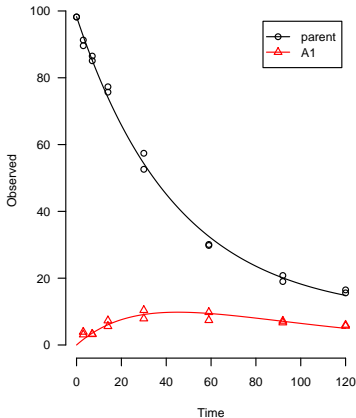
Introduction

Results

Parent only

Coupled fits

Conclusions



Error models for degradation experiments

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only
Coupled fits

Conclusions

- Constant variance

Error models for degradation experiments

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

- Constant variance
- Variance by variable (Gao *et al.* 2011)

Error models for degradation experiments

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

- Constant variance
- Variance by variable (Gao *et al.* 2011)
- Variance by residue level?

A two-component error model from analytical chemistry

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

$$\sigma(y) = \sqrt{\sigma_{\text{low}}^2 + y^2 \text{rsd}_{\text{high}}^2}$$

$\sigma(y)$	Standard deviation of residuals as a function of the magnitude of the observed value
σ_{low}	Standard deviation for small observed values
rsd_{high}	Relative standard deviation for large values

How to fit the different error models

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

	OLS	IRLS	Other
Constant variance	✓	-	-
Variance by variable	-	✓	✓
Two-component variance	-	(✓)	✓

OLS: Ordinary Least Squares

IRLS: Iteratively Reweighted Least Squares

Other: Direct or stepwise maximisation of the likelihood

Selection criterion for error models

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

- FOCUS χ^2 error level assumes constant variance (which is then expressed as a relative error)

Selection criterion for error models

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

- FOCUS χ^2 error level assumes constant variance (which is then expressed as a relative error)
- Akaike Information Criterion (AIC) depends on likelihood and number of parameters

Implementations in mkin

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

February 2018 (mkin 0.9.47.2):
Fitting the two-component error model by IRLS

July 2019 (mkin 0.9.45.5):
Use more general likelihood-based algorithms

<https://pkgdown.jrwb.de/mkin/news>

Test with datasets from Risk Assessment Report (RAR)
documents

- 12 soil datasets, 11 with metabolites

Test with datasets from Risk Assessment Report (RAR)
documents

- 12 soil datasets, 11 with metabolites
- 6 water sediment datasets without metabolites

Soil datasets - parent only

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

- Constant variance and two-component error model for Simple First-Order (SFO), Dual First-Order in Parallel (DFOP) and Hockey Stick (HS)

Soil datasets - parent only

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

- Constant variance and two-component error model for Simple First-Order (SFO), Dual First-Order in Parallel (DFOP) and Hockey Stick (HS)
- Two-component error model has lower AIC in 4 out of 12 cases

Soil 6: SFO

Error models
for chemical
degradation
data

J. Ranke

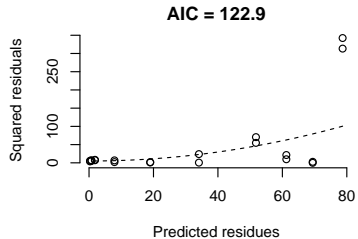
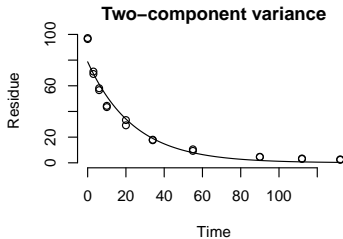
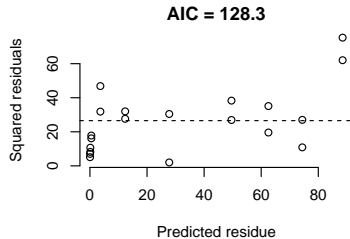
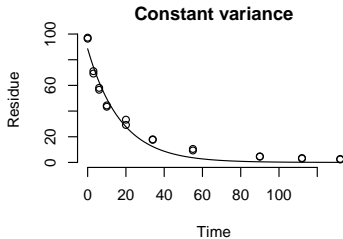
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 6: DFOP

Error models
for chemical
degradation
data

J. Ranke

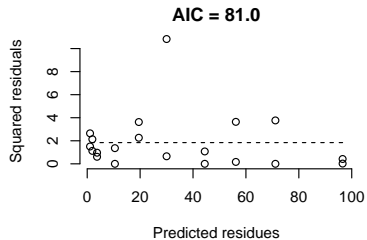
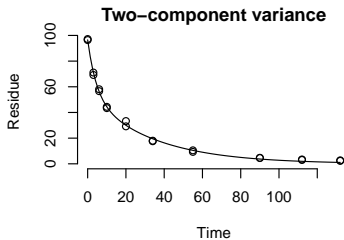
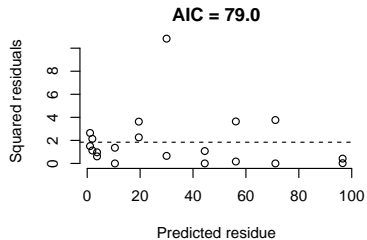
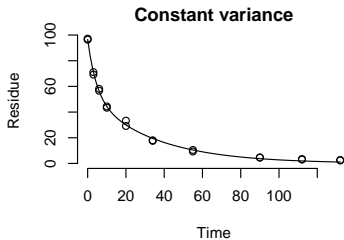
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 10: SFO

Error models
for chemical
degradation
data

J. Ranke

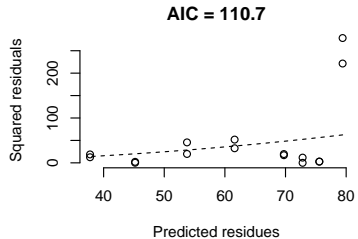
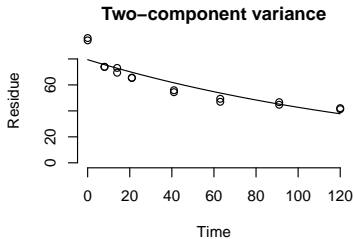
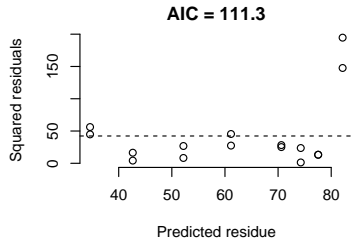
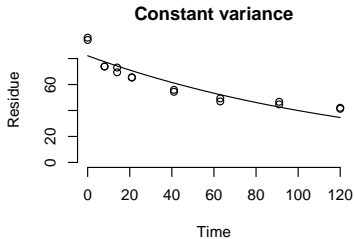
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 10: DFOP

Error models
for chemical
degradation
data

J. Ranke

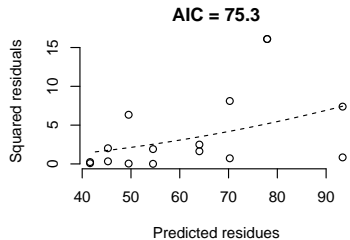
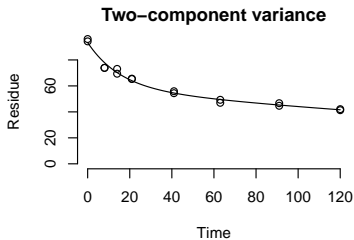
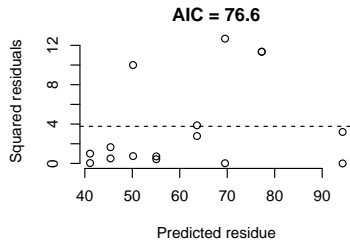
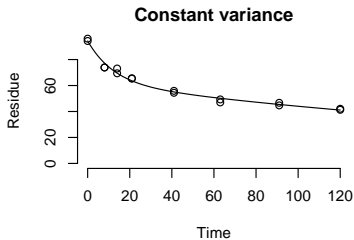
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 2: Variance by variable

Error models
for chemical
degradation
data

J. Ranke

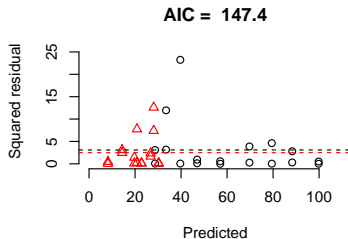
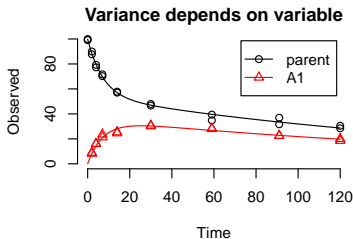
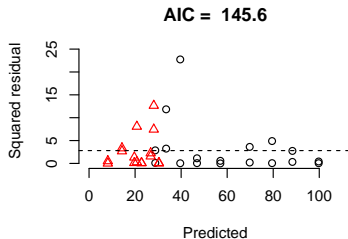
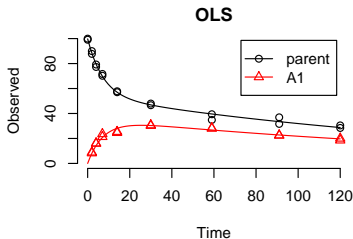
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 2: Two-component error model

Error models
for chemical
degradation
data

J. Ranke

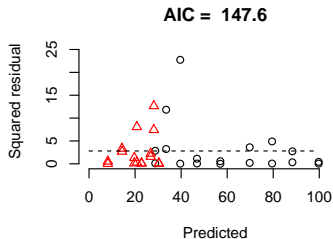
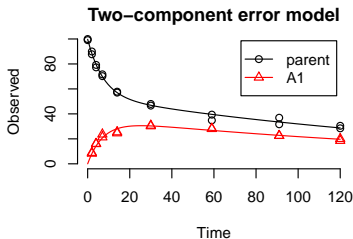
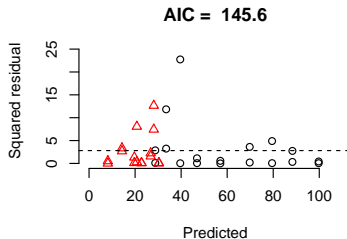
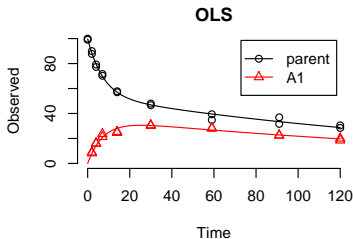
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 4: Variance by variable

Error models
for chemical
degradation
data

J. Ranke

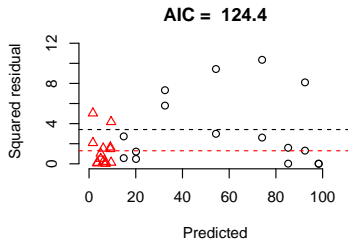
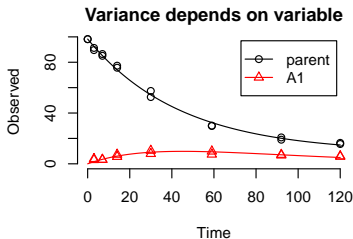
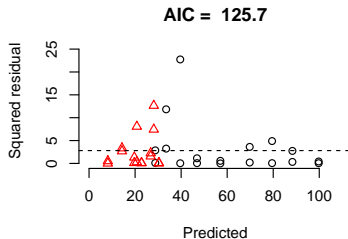
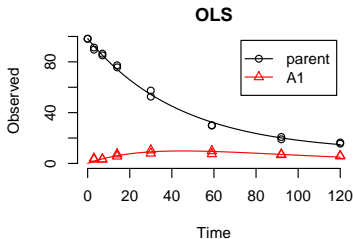
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 4: Two-component error model

Error models
for chemical
degradation
data

J. Ranke

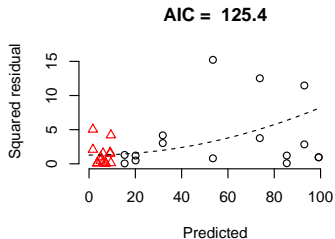
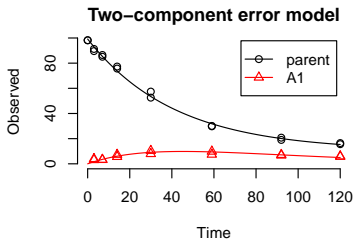
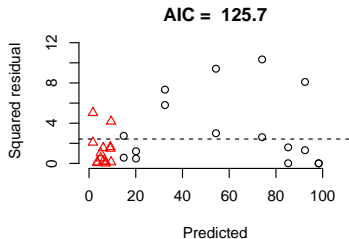
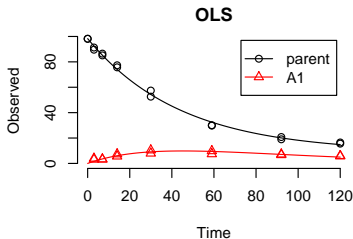
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 3: Variance by variable

Error models
for chemical
degradation
data

J. Ranke

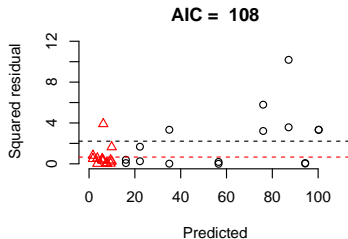
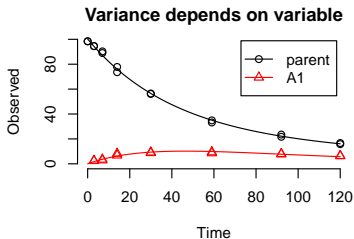
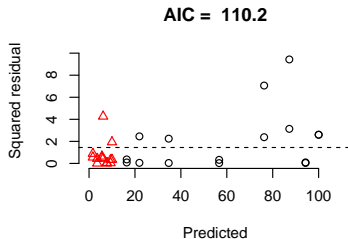
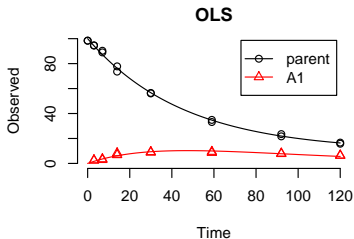
Introduction

Results

Parent only

Coupled fits

Conclusions



Soil 3: Two-component error model

Error models
for chemical
degradation
data

J. Ranke

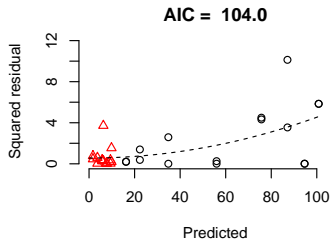
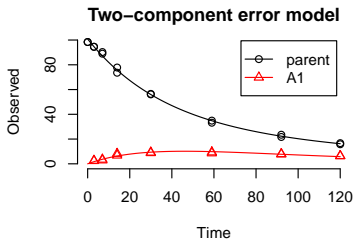
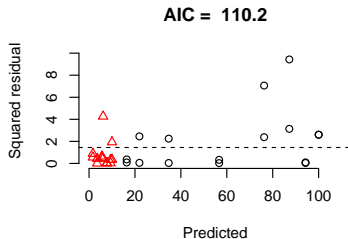
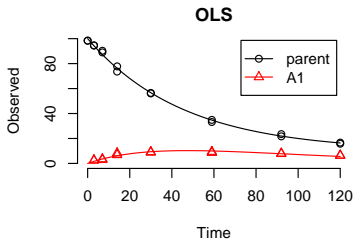
Introduction

Results

Parent only

Coupled fits

Conclusions



Conclusions

- For many datasets, the two-component error model provides a better representation of the error structure

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

Conclusions

- For many datasets, the two-component error model provides a better representation of the error structure
- Metabolite rate constants, k_2 values of biphasic kinetics and their p-values can become lower or higher

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

Conclusions

- For many datasets, the two-component error model provides a better representation of the error structure
- Metabolite rate constants, k_2 values of biphasic kinetics and their p-values can become lower or higher
- The variance around the initial residue level of 100% is often overestimated

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

Conclusions

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only

Coupled fits

Conclusions

- For many datasets, the two-component error model provides a better representation of the error structure
- Metabolite rate constants, k_2 values of biphasic kinetics and their p-values can become lower or higher
- The variance around the initial residue level of 100% is often overestimated

Conclusions

Error models
for chemical
degradation
data

J. Ranke

Introduction

Results

Parent only
Coupled fits

Conclusions

- For many datasets, the two-component error model provides a better representation of the error structure
- Metabolite rate constants, k_2 values of biphasic kinetics and their p-values can become lower or higher
- The variance around the initial residue level of 100% is often overestimated



Financial support and feedback by the Federal Environmental Agency is gratefully acknowledged!