Index

A

absorption spectroscopy, 218
accuracy
  defined, 10, 12
active-experiment approach, 266
active experiments, presumptions for, 265
  control of values and, 265
  measurement of output variables and, 265
  noise variables and, 265
algorithms, 4
algorithms for realizable optimal experimental designs
  exact (or N-Point) D-Optimal designs, 307–309
  sequential composite D-Optimal designs, 313–315
  sequential D-Optimal designs, 310–311
aliasing, 388
all possible subsets, 138
ALS algorithm
  PARAFAC and, 491
alterative hypothesis, 48–49, 122
alternating least-squares (ALS) methods, 491
alternating least-squares (ALS) methods, solution constraints, 493
  known profiles, 494
  nonnegativity, 494
  unimodality, 494
alternating least-squares methods
  PARAFAC / CANDECOMP, 491–494
American Society for Testing Materials (ASTM), 107, 114
analysis of emission spectroscopic data, *241*
analysis of multivariate data in kinetics, *228*
analysis of variance (ANOVA), 28
  analysis of residuals, 30–32
  ANOVA to test for differences between means, 28
  between-sample variation (between-treatment variation), 29–30
  within-sample variation (within-treatment variation), 29
analyte concentrations in unknown solutions, 476
analytical chemistry, 2, 512
ANOVA (analysis of variance), 125
  calculations, summary approach, 31
  table, *32*
apodization
  transfer function and, 397
application of DOE in multivariate calibration, 321–329
  improving quality from historical data, 330–335
application of DOE in multivariate calibration, construction of a calibration set, 321–329
  defining the bounds of the factor space, 327–328
  estimating extinction coefficients, 329–330
  identifying the number of significant factors, 322–324
application of DOE in multivariate calibration, improving quality from historical data
  improving the numerical stability of the data set, 333
  prediction ability, 334
applications, chemometric
  biochemical processes, 449, 449–459
  environmental data, 454
archaeological artifacts
  obsidian, composition of, 356
artificial neural networks (ANN), 477
Association of Official Analytical Chemists (AOAC), 114
AT&T Bell Labs, 294
autocatalysis, 246–248
autocatalytic reaction scheme, *246*
autoscaled data with E-optimal subset (ref Fig 8.31), ***335***
autoscaling, 79, 458

B

backward elimination, 137
  Efroymson's stepwise regression algorithm and, 137
  forward selection and, 137
bandwidth-limited discrete Fourier transform, 388
barycentric coordinate systems, 319
baseline correction, 80
  of Raman emission spectra, *80*
basis vectors, 71–72, 96, 141, 189
  column basis vectors, 71
  row basis vectors, 71
Beer-Lambert's Law, 219, 237, 241, 324, 328

matrix notation, result of data-fitting procedure and, *238*
spectroscopy and, 462
Beer's law, 107
Belousov-Zhabotinsky (BZ) Reaction, 251–253, 256
FKN mechanism, 252
four stages of, 252
Oregonator model, 252
oscillating reactions and, 251–252
oxidation of organic species in, 251
represented b the Oregonator model, *263*
between-run, defined, 10
bias
laboratory, 18
method, 18
systematic error and, 18
biased methods of calibration, 139
bilinear data matrix, 71
bivariate normal distribution
scatter plots of, *56–58*
"black box" principle, *265*, 265, 266, 269
BP Amoco naphtha processing plant, 330
Butterworth filter equation, 396

C

calculation of the concentration profiles: case I, simple mechanisms, 220
calculation of the concentration profiles: case II, complex mechanisms, 241–252
Fourth-Order Runge-Kutta Method in Excel, 242–245
Interesting kinetic examples, 246–252
calculation of the concentration profiles: case III, very complex mechanisms, 253
Arrhenius theory and, 254
changes in ionic strength and, 254–255
changes in pH and, 254–255
experimental inconsistencies and, 253
Eyring theory and, 254
temperature fluctuations and, 253
calibration, 105–166
curvilinear, 112
data sets and, 107–108
internal standards, 156
introduction to, 109
measurement error and measures of prediction error, 114
multivariate, 111
non-zero intercepts and, 110
preprocessing techniques, 156, 156–157
sample sets, 264
selection of calibration and validation samples, 113
sets, 322
software, 159
statistical evaluation of calibration models obtain by least squares, 121–134
univariate, 109, 211
calibration, a practical example, 116
graphical survey of NIR water-methanol data, 116
multivariate, 119–120
univariate, 118, 118–119
calibration, biased methods of, 139–153
example regularization results, 153
other calibration methods, 150–151
partial least squares, 147–149
principal component regression, 140–153
regularization, 151
calibration line, fulcrum and, 129
calibration models, 330
spectral data and, ***324***
spectroscopic, 321
calibration models, statistical evaluation obtained by least squares, 121–134
coefficient of determination and multiple correlation coefficient, 130
confidence interval and hypothesis tests for regression coefficients, 126
hypothesis testing, 122
interference effects and selectivity, 234
interpreting regression in ANOVA tables, 125
leverage and influence, 128
model departures and outliers, 129
partitioning of variance in least-squares solutions, 123–124
prediction confidence intervals, 127
sensitivity and limit of detection, 131–132
calibration, multivariate, 119–120
chemometrics and, 264
calibration, software and, 159
DeLight, 159
GRAMS, 159
MATLAB and, 159–160
Pirouete, 159
PLS_Toolbox, 159
SIMCA, 159
Unscrambler, 159
calibration, standard addition method, 153–154, 153–155
multivariate standard addition method, 155
univariate standard addition method, 154, 154–155
calibration, standardization, 157, 157–159
alternative standardization (transfer) methods, 157
calibration-transfer methods, 157
of instrument response, 158

of predicted values, 157–158
with preprocessing techniques, 159
calibration steps
calibration prediction step, 143
compute a regression vector using calibration samples, 143
compute the projected calibration spectra, 142
estimate the RMSEC, 143
calibration, three-way, 4
hyphenated data and, 475–508
calibration, variable selection, 135–139
all possible subsets, 138
backward elimination, 137
Efroymson's stepwise regression algorithm, 136–137
forward selection, 135
recommendations and precautions, 138
sequential-replacement algorithms, 138
simulated annealing andgeneic algorihm, 138
calibration work, UV range and, 330
cascade of wavelet coefficient vectors from wavelet tree filter banks defining the discrete wavelet transform, *411*
catalog of four eact D-Optimal experimental designs
for the spectroscopic calibration problem, ***327***
central composite designs (CCD), 293
three-factor with axial values and four center points, *294*
for two factors, *293*
central limit theorem, *13*, 45
implications of, 45
Chebyshev polynomials, 396, *404*
chemical analysis
interferences and, 134–135
chemical measurements
chemical properties and, 2
decision making and, 1, 210
physical properties and, 2
statistical properties and, 2
as three-legged platform, 2
chemical properties, 2
equilibria, 2
kinetics, 2
mass balance, 2
stoichemetry, 2
chemometrics, 1–6, 509, 510, 516
analytic chemistry and, 510
biomedical applications of, 514
calibration, validation, and significance testing, 2
in Chemical Abstracts, 511
chemical and biological sciences, 511, 514
chemical data and, 510
chemical measurements and, 2, 509
computing and, 2, 510, 512, 513
environmental forensics and, 512
extraction of chemical information from analytical data, 2
future trends, 509–520
gas chromatography (GC) and, 510
gas chromatography-mass spectrometry (GC-MS), 510
historical development of, 50, 509
hyper spectral images, 514
large data sets and, 514
live cell hyperspectral imaging, 514
mathematical modeling and, 509
as maturing discipline, 511
multivariate calibration methods, 510
multivariate methods of analysis, 510
multivariate statistics and, 509
near infrared spectroscopy and, 510
new instruments and, 513
NIR spectroscopy and, 510, 514
partal least-squares, 510
partial least-squares, 509
pattern recognition and, 509
principal component analysis (PCA), 509
principle component analysis (PCA) and, 510
quantitative biology and, 514
Raman spectroscopy and, 514
smart sensors and, 512
chemometrics citations in *Chemical Abstracts* from 1974–2004, 511
chemometrics, drivers of growth in, 513
chemical and biological sciences, 514
large data sets, 514
chemometrics, reviews of and future trends, 511
food and feed chemistry, 512–513
interesting application areas, 511, 513
process analytical chemistry, 512
spectroscopy, 512
chi-square distribution, 47, 58–59
degrees of freedom and, *48*
chromatography, 140, 477
chromographic profile optimized by ITTFA, *438*
classical and robust tolerance ellipse of the phosphorus data set, *173*
classical PCA (CPCA), 186, 188, 194
classical PCR (CPCR), 196
classical quadratic discriminant analysis (CQDA), 207
classification, 207, 351–354
discriminant analysis and, 207
K-Nearest neighbor, 352
partial least squares (PLS), 352
SIMCA, 353
supervised learning and, 207
classification and pattern recognition, 339–378

clustering and, 347
data preprocessing, 341
introduction to, 339–340
mapping and display, 342–346
practical considerations, 354
classification in high dimensions, 211
dimension-reduction procedure (PCA), 211
MCD as uncomputable, 211
SIMCA method (soft independent modeling of class analogy), 211
classification in low dimensions, 207–210
classical and robust discriminant rules, 207
evaluating the discriminant rules, 208
an example, 209
classification methods, 351
K-nearest neighbor (KNN), 351
partial least squares (PLS), 351
soft independent modeling by class analogy (SIMCA), 351
classification of jet fuels, *364*, 365
closure, 434
cluster analysis, 347
clustering, 345, 347, 363
classification with PCA score plots and, 98
hierarchical, 351
single linkage and, 351
clustering algorithms, 348
FCV, 348
hierarchical, 348
K-means, 348
Patrick Jarvis, 348
similarity matrix, 348
clusters
determining numbers of, *348*
elliptical, 351
identifying, 347
linear, 351
spherical, 351
validity, 347
CM-estimators, 183
coefficient of determination, 131
multiple correlation coefficient and, 130
combined canonical models
independent (process) variables and, 284
mutually dependent (mixture variables) and, 284
combined mixture and process factor space constructed from three mixture variables and one process variable, *286*
components included in product P1, with UV-Inactive components shaded, ***328***
composition (loading) profiles of resolved components in the MCR-ALS analysis of raw augmented data matrix, *460*
compound maps, 462
compounds, kinetic or thermodynamic process and, 435
computation for distance between data cluster and a point, *350*
concentration of Components C1 to C9 used to measure pure-component spectra, ***324***
concentration profiles
for a first-order reaction at constant and increasing temperature, *254*
of an HPLC-DAD data set, *425*
for a reaction as calculated by MATLAB, 221
for a two-component system, 421
concentration windows, 426
confidence intervals, 14, 299
compared with nonoptimal design and D-optimal design, *301*
critical value of the Student statistic, 300
estimate of the measurement error and, 300
hypothesis tests for regression coefficients and, 126–127
influence of on the error of the prediction, *300*
of the mean, equation for, 46
for the regression line and for predictions, illustration of, *128*
variance of the prediction, 300
confidence level, 133
consrained region
in three-component mixtures, *281*
of three mixture variables with lower bounds only, 280
constrained mixture region with upper and lower bounds, 282
constrained mixture spaces, 279–280
constraints, 433–437
application of, 433
chemical properties and, 433
closure, 434
effect on the shape of resolved profiles, *433*
equality and inequality and, 433
hard-modeling constraints: physicochemical models, 435
implementation of, 433
iterative optimization and, 433
known profiles, 435
local rank, 435
mathematical properties and, 433
nonnegativity, 433, 434, 465
selectivity, 435
unimodality, 434
zero-concentration, 435
constraints, hard-modeling, 435
concentration profiles and, *436*
physicochemical models, 435
construction and decomposition of a three-way array via the trilinear PARAFAC model, *480*

construction of a three-way array according to the unconstrained Tucker model, *480*
continuous sine-wave signal with aliased curve reconstructed, *387*
controlled and noise variables, 297
  examples of, ***295***
convergence progress for PARAFAC from DTLD initialization and random initialization, 499
convolution, 397
Cook's distance
  distance measures and, 129
coordinate systems
  barycentric, 271
  barycentric and Descartes, three-dimensional example, *271*
  barycentric and Descartes, two-dimensional example, *270*
  Descartes, 271
  simplex, 271
correlation coefficients between absorbence at selected pairs of wavelengths for sulfamethoxazole training set, ***63***
CPCR
  RPCR and, 199
CQDR, 209
critical values for *F*
  for a one-tailed test, *20*
  for a two-tailed test, 21
curvature effect, 113
curve resolution (CR) method, 421, 423
curvilinear calibration, 112–113
cutoff frequency, 395

D

D-optimal design, 301
data analysis algorithms, 481
  DTLD Algorithm, 481
  GRAM algorithm, 481
  MATLAB programming language and, 481
  PARAFAC algorithm, 481
data matrix
  augmentation arrangement and bilinear model for PCA or MCR-ALS decompositions, *455*
  plotting, 431
data points in subspace of original measurement space, *344*
data preprocessing, 341
  normalization, 342
  raw data into computer ready form, 341
  scaling, 341
data sets, 55
  fundamental modes of vibration, overtones, and combinations, 108
  near infrared spectroscopy and, 107
  solvent interactions, 108
  water-methanol mixtures, 108
data sets from multicomponent systems, *418*
data space
  dividing, 351
data vector, *78*
DATAN data analysis method, 495
datascope, 431
Daubechies wavelets, 407
decision limits, graphical representation of, 132
degrees of freedom, 46, 123–124, 125, 235
dendogram, single-linkage
  of the obsidian data, 357
dendograms, 351
  as intuitive and, *349*
  using single-linkage method developed from gs chromatograms of scent marks of Marmoset monkeys, *349*
denoised signal, *414*
denoising, 413
  of wavelet components using entropy threshhold, *413*
Descartes coordinate systems, 319
design efficiency, 303
Design Expert, 319
design measures, 303
design of experiment (DOE), 264
  software, 320
designer transfer function, 395
designs, full-factorial and simplex-lattice, ***283***
determination limit, 134
DETMAX algorithm, 308
DFT, 400
direct exponential curve-resolution algorithm (DECRA), 495
direct methods,simplex, 225
  simplex algorithm and, 225
direct trilinear decomposition (DTLD), 489
  GRAM algorithm and, 489
discrete, bandwidthlimited Fourier transform of noisy spectrum, *390*
discrete convolution of two functions, *393*
discrete Fourier transforms (DFT), 389, 409
discrete wavelet, 409
discrete wavelet transform (DWT), 409, 410
discriminant analysis
  classical quadratic discriminant analysis (CQDA), 207
  cross-validation and, 208
  evaluating the discriminant rules, 208
  probability of misclassification, 208
distribution
  functions, 41–68
  normal, 14
  volume of, 53

distribution curves
normal and standard normal, *44*
distribution of error in Eigenvalues, 93
Dixon Q test, 33, ***33***
DTLD, 498
algorithm (Appendix 12.2), 503
application, 490, 494–495
application, GRAM and, 490

E

E-Optimal calibrations, ***336***
effect of using a 30-point E-optimal subset during the latent-variable extraction, 334
effects of constrains in feasible bands, *447*
Efroymson's stepwise regression algorithm, 136–137
convergence of algorithm, 137
variable-addition step, 136, 136–137
variable-deletion step, 137
eigenvalues, 70, 74, 75, 76, 141
information and, 345
primary, 74
secondary, 75
eigenvectors, 70, 74, 75, 76, 141, 189
complete set of, 74
eigenspectra and, 73
environmental data
environmental monitoring and, 454
MCR-ALS and, 454
multivariate curve resolution, 454
PCA and, 454
principle component analysis and, 454
environmental monitoring program in Portugal, 455
environmental study of SVOc in surface waters of Portugal, 460
error, type I, 50, 123, 133
error, type II, 50, 123
error, types of
gross error, 10
random error (noise), 10
systematic error, 10
errors, estimating, 235
estimated concentration in mixture samples from GRAM with concatenated matrices, ***486***
estimated concentrations in mixture samples from GRAM with added matrices, ***486***
estimated factors
visual inspection and, 498
estimating extinction coefficients, 329
Euclidean distance between two data points, *350*
Euler's method for numerical integration, *242*
simple, 242
evolving factor analysis (EFA), 423, 426
backward EFA, 424
emergence and decay of compounds and, 424
forward EFA, 424
plot, 424
exact (or N-Point) D-Optimal designs
DETMAX algorithm, 308
Fedorov's algorithm, 307
MD Galil and Kiefer's algorithm, 309
Wynn-MItchell and van Schalkwyk algorithms, 308
Excel, 218
Excel's Solver, 227
algorithm, 234
using for nonlinear fitting of a first-order reaction, *228*
experimental design, 4, 265
catalogs of, 320
controlled factors and, 294
desired properties for, 298–299
full factorial design (FFD), 285
inomplete lattice, 285
for lithium lubricant study, ***278***
mixture and process design, 285
noise factors and, 294
simplex-lattice designs, 272
experimental design, off-the-shelf software packages, 316–320
Design Expert, 319
MATLAB, 316–318
other packages, 319
experimental error
Monte Carlo simulation methods and, 448
resampling methods and, 448
experimental region, 268
experiments, passive, 330
explained varianced by PCA for three data pretreatment methods, ***458***
extensions of three-way methods, 495
extraction of analyte, 476
extreme values
outliers, 34
stragglers, 34

F

F-test for determining the number of factors, 93
Factor Analysis in Chemistry, Malinowsky
history of chemometrics and, 422
factor spaces
constrained mixture spaces, 279–282
defining bonds of, 327
mixture and process spaces, 283–286
one-dimensional, *269*
process, 268
simplex-centroid designs, 275–278

simplex-lattice designs, 272–274
two and three-dimensional in coded variables, *269*
two-dimensional, *268*
factor spaces, mixture, 269–271
Euclidean space and, *269*
mixture variables and, 269
factorial designs
advantages of, 290
disadvantages of, 290
factorial designs, full, 290
three-level, three factor, *292*
three or more levels in, 291
two-level three factor, ***291***
fast Fourier transforms (FFT), 389
algorithms, 389
Cooley-Tukey transform and, 389
Sande-Tukey transform and, 389
FAST-MCD method, 203
Fearn's algorithm for OSC smoothing, 385
Fedorov's algorithm, 307
FFT, 400
figures of merit, analytical, 496
"net analyte signal" (NAS) and, 496
filter
lag, 400
window, *402*
filtering, 380, 400
finite impulse response (FIR) filter, 401
signal processing with a simple moving average filter, *401*
first and second derivatives, 82
spectra of water-methanol mixtures, *117*
first order kinetic single-wavelength experiment and result of a least-squares fit, *223*
fitting parameters to a given measurement, 225
direct methods, 225
Newton-Gauss methods, 225
fitting, with Excel's Solver, *239*
fixed-size image window-evolving-factor analysis(FSIW-EFA), 463
fixed-size moving-window--evolving-factor analysis (FSMW-EFA), 423, 426
flash photolysis, 218
flow reactors, 256
food and feed chemistry, 512
Ford Motor CO., 294
forecasting, 127
forward selection, 136
Fourier domain, 388, 399
Fourier series, 385
Fourier transform, 386, 397, 400
convolution and, 391
correlation and, 391
Parseval's theorem and, 391
properties of, 389–394, ***391***
window function and real part, *393*
Fourier transform infrared spectrometer (FTIR), 70
frequency-domain signal processing, 385–393
bandwidth-limited discrete Fourier transform, 388
Fourier transform, 386
properties of the Fourier transform, 389
sampling theorem and aliasing, 386–387
frequency-domain smoothing, 394, 394–397
with designer transfer functions, 395–397
frequency response of noisy signal and its matched smoother, *398*
FSIW-EFA analysis, 464
fuel spill identification, 358
gas chromatograms, 358
full factorial design (FFD), 283–284
Fundamental Reviews in Chemometrics, 512
fundmental modes of vibration, overtones, and combinations, 108

G

gas chromatography (GC), 70, 330
general equivalence theorem, 304
generalized eigenvalue-eigenvector problem (GEP), 482
generalized M-estimators (GM-estimators), 183
generalized multivariate squared distance, 52
generalized or mahalanobis distances, 52
generalized rank annihilation method (GRAM), 4, 475, 485
generic plot of a variance indicator against a bias measure, *145*
genetic algorithm (GA), 138
genome sequences, chemometrics and, 514
GRAM, 489, 495, 498
algorithm (Appendix 12.1), 502
application, 486–488
trends in predictive performance, *487*
graphical survey of NIR water-methanol data, 116
Grubbs' tests
application to chemical data, 35
critical values of G for, 35
outliers, 34

H

Hadamard transform function, 317
hard-modeling constraints: physicochemical models, 435
harmonious PCR plot for methanol, *149*
Hawkins-Bradu-Kass data set, 187
artificial data set, 186

Heisenberg uncertainty principle, 409
Hessian matrix, 234
heuristic evolving latent projecions (HELP), 429–430
  HELP method, 431
hierarchical models, 514
high-performance liquid chromatography (HPLC), 70
high-speed gas chromatographic profiles of jet fuels, *360*
histogram showing absorbence of sulfamethoxazole training set, 62
Hotelling, 510
HPLC-DAD
  data set, 431
  system, 430
"hyphenated" mass spectrometry-chromatographic methods, 423
hyphenated methods, 477
hypothesis test
  comparison of multivariate means, 59
hypothesis testing, 122–123
  alterative hypothesis, 48–49
  null hypothesis, 48–49
  principles of, 48–49
  statistical inference and, 48–49

I

impulse response function, 399
impulse response function of the seven-point rectangular smoother window function, *396*
incept, nonzero, 119, 462
initial estimates, chemometric methods and, 432
input variables, 268
  mixture variables, 268
  process variables, 268
interferences, 134–135
  chemical, 134–135
  physical, 134–135
  spectral, 134–135
International Chemometric Society, 509
International Union of Pure and Applied Chemistry (IUPAC), 12
Interpreting regression ANOVA tables, 125
iterative resolution methods, 431
iterative target transformation factor analysis (ITFA), 432, 437–438, 477
  process and, 438–439
  target factor analysis (TFA) and, 437
  target vector and, 439

J

Jacobian J
  representation of, *232*
Jacobian J and Marquardt parameter, *234*
Jacobian J matrix, 232–233

K

K-nearest neighbor (KNN), 352, 354
  Bayes classifier and, 352
  Euclidean distance and, 352
kinetic data, analysis of, 246–253
  additional methods, 258
  autocatalysis, 246
  Belousov-Zhabotinsky (BZ) Reaction, 251–253
  flow reactors, 258
  globalization of the analysis, 258–259
  Lotka-Volterra (Sheep and Wolves), 250–251
  measurement techniques and, 256
  model parser, 256
  related issues, 255–258
  soft-modeling methods, 257–258
  Zeroth-Order reaction, 248–249
kinetic modeling of multivariate measurements, nonlinear regression, 217–262
  Beer-Lambert's Law, 219
  calculation of the concentration profiles: case I, simple mechanisms, 220
  calculation of the concentration profiles: case II, complex mechanisms, 241–252
  calculation of the concentration profiles: case III, very complex mechanisms, 253–254
  matrix notation, 219
  model-based non-linear fitting and, 222–240
  multivariate data, 219–220
  related issues, 255–258
kinetic modeling of multivariate measurements, related issues
  flow reactors, 256
  globalization of the analysis, 256
  measurement techniques, 256
  model parser, 256
  other methods, 258
  soft-modeling methods, 257
kinetics of chemical processes
  light-absorption spectroscopy and, 218
kinetiic reactions, 246–252
  autocatalysis, 242–246
  Belousov-Zhabotinsky (BZ) Reaction, 251–252
  Lotka-Volterra (Sheep and Wolves), 250
  zeroth-order reaction, 248–249
Kjeldahl nitrogen titration method, 107

known profiles, 435
Kohonen neural network, 345–346
Kowalski, Bruce
chemometrics and, 509

L

least median of squares (LMS), 183
least squares, 121
classic, 107
estimates, 111
least-squares fit, 222, 237
of emission spectroscopic data, *240*
leave-multiple-out CV (LM) (CV)
Monte Carlo CV (MCCV), 115
leave-one-out cross-validation (LOOCV), 115
leverage and influence, 128
calibration samples and, 128
leverage, in statistics, 128
leverage points, 184, 191
light-absorption spectroscopy
CD (circular dichroism), 218
NIR (near-infrared_ reflectance spectra, 218
UV/Visible spectroscopy, 218
limit of detection, 132
determination limit, 134
multivariate detection limit, 134
univariate decision limit, 132
univariate detection limit, 133
limitation of quantity (LOQ), 114, 134
linear and non-linear parameters, *228*
linear discriminant rule, 209
linear effect, 113
linear models with n standards, diagram of three different types, *109*
linear parameter fitting, 241
linear regression in low dimensions, 176
linear regression with one response variable, 176
linear regression with one response variable, 176
the classical least-squares estimator, 177
the multiple linear regression model, 176
other robust regression estimators, 182, 182–183
an outlier map, 180, 180–181
the robust LTS estimator (least trimmed squares), 178–179
linear regression with several response variables, 183
an example, 185, 194
the multivariate linear regression model, 183–184
principle component regression (PCR), 195
the robust MCD-regression estimator, 184
selecting the number of principle components, 193
local rank analysis methods, 423
evolving factor analysis (EFA), 423
fixed-size moving-window--evolving-factor analysis (FSMW-EFA), 423
images and, 463
local rank constraints, 435–436
local rank maps, 426
of pill images, *464*
location and covariance estimation in low dimensions, 173
the empirical mean and covariance matrix, 173
other robust estimators of location and covariance, 176
the robust MCD estimator, 174–176
location and scale estimation
mean and standard deviation, 169–171
median and median absolute deviation, 171
other robust estimators of location and scale, 171–173
logarithm of the square sum (ssq) of the residual vector r as a function of the rate constant k, *224*
Lotka-Volterra (Sheep and Wolves), 250–251
dynamics of population of predators and preys in closed systems, 250
kinetic population profiles and, 250
MATLAB's Runge-Kuta solver and, 250
predator and prey kinetics, *250*
LTS estimator
regression, scale, and affine equivariant, 179
LTS regression, 183, 197, 198
errors and, 179

M

M-estimators, 183
Mahalanobis, 52, 192
distances, 52, 59, 64
Mahalobis Distances and Probability Densities from Hotelling's T^2 Statistic for Test Samples of Sulfamethoxazole compared with the Training Set, *64*, ***66***
Malinowski's empirical indicator function (IND), 91
Malinowski's RE function, *91*
Manne's resolution theorems, 421, 427
mapping and display, 341
data compression, 343
dimensionality reduction with PCA and, *343*
principle component analysis (PCA) and, 342, 343
variance and, 342
Marquardt parameter mp, 234
mass spectrometer (MS), 70
matched smoothing of noisy signal, *398*

mathematical handling of multiwave absorption data sets, 218
MATLAB, 74, 202, 232, 241, 456, 465, 481
backslash notation, 234
chemometrics and, 218
code, 59, 218
corrocoef functions, 62
cov functions, 62
D-optimal designs and, 317–318
diff command, 60
example 3.1, 61
example 3.2, 61
example 3.3, 62
example 3.4, *63*
example 3.5, 65
example 3.6, 66
example 4.1: eigenvectors and eigenvalues, 75
example 4.2: principal component analysis of a spectroscopic data set, 75–76
example 4.3: principal component analysis using the SVD), 76–77
example 4.4 autoscale of a matrix, 79–80
example 4.5: PCA procedure, 86
example 4.6: function for calculating Malinowski's RE, IND, and REV functions, 92–93
example 4.7: determining the number of significant principal components in a data matrix, 95–96
example 4.8: computation of residual spectra, 100
example 4.9: residual variance analysis, 102
example 7.1, 222, 223, 254
example 7.2, 222
example 7.3a, 226, 227
example 7.3b, 226, 227
example 7.4a, 229
example 7.4b, 230
example 7.5a, 235, 236
example 7.5b, 235, 236–237
example 7.5c, 235, 237
example 7.6a (autocatalysis), 247
example 7.6b, 248
example 7.7a (Zeroth-order kinetics), 249
example 7.7b (Zeroth-order kinetics), 249
example 7.8a, 250
example 7.8a lotka volterra, 251
example 7.8b ode lotka volterra, 251
find command, 61, 65
functions for DECA, 495
hist command, 61
m_dist.m function, 64
mean functions, 62
plot command, 60
Runge-Kutta method and, 242, 246
Statistics Toolbox, 317
zoom command, 60, 65
matrix notation, 218, 219
maximum-likelihood rule, 207
MCD estimates, 184, 187, 193
MCD estimator, 198
MCD regression, 197, 198, 203
an example, 185
MCR-ALS decomposition method, 448
biochemical processes and, 449–452
MD Galil and Kiefer's algorithm, 309
mean
arithmetic, 11
defined, 10
inferences about, 49
population, *13*
mean absolute deviation (MAD), 37
mean centering, 78, 159
effect on a bivariate distribution of data points, graphical illustration, *78*
measurement error and measures of prediction error, 114
measurement techniques, 256
median absolute deviation (MAD), 171
membership probability, 207, 208, 209
mesh representation of pure distribution maps, 465
methanol model at without an intercept term, graphical displays, *120–121*
methanol model with a non-zero intercept, *124–125*
methanol model with a nonzero incept, graphical displays, *122–123*
methanol result from univariate and multivariate calibration of water-methanol mixture, ***118***
middle point of the window, *402*
misclassification probabilities for RQDR and CQDR applied to the fruit data set, ***209***
misclassification probability(MP), 208, 209
misclassification rates, 210
mixture and process factor space for incomplete cubic simplex-lattice design combined with a full factorial design, ***284***
mixture and process factor spaces, 283
symmetric experimental designs and, 283
mixture factor spaces
coordinates in, shown in barycentric and Descartes coordinate systems, ***271***
model
calibration and validation, 198
parser, 256
validation, 114
model-based fitting of kinetic data, 218
methods for nonlinear least-squares fitting, 218

modeling the concentration profiles of the reacting components, 218
model-based non-linear fitting, 222
direct methods, simplex, 225–226
linear and non-linear parameters, 228–229
Newton-Gauss-Levenberg/Marquardt (NGL/M), 230
non-linear fitting using Excel's Solver, 227
nonwhite noise, 237–239
model departures and outliers, 129
residual plot and, 129
model determination showing the bias/variance tradeoff with selection of the metaparameter, *144*
Monte Carlo CV (MCCV)
leave-multiple-out CV (LM)(CV), 115
mother Daubechies wavelet and two shifted and scaled children fitting a noisy signal, *407*
mother wavelet functions, 407, *407*
moving-average filter, noise reduction and, 403
multiblock models, large data sets and, 514
multicollinearity, 194
multicomponent systems
mixtures, 448
processes, 448
multiple correlation coefficient, 131
multiple linear regression (MLR), 169
multiple overlapping absorption bands in the NIR spectrum f water, *89*
multiple rank analyses, 423
multiplicative scatter correction (MSC), 83, *85*
multivariate
distribution, 52
measurements, 225
multivariate absorbence data
Beer-Lambert's Law and, 241
multivariate adaptive regression splines (MARS), 477
multivariate advantages, 3, 52
multivariate calibration, 3, 111–112, 195
methods, 512
multivariate classification and pattern recognition, 4
multivariate curve resolution, 4, 417–467, 417–474
ambiguities and resolution theorems, 418–422
environmental data, 454–460
error in resolution, 446–447
historical background of, 422
local rank and resolution: evolving factor analysis and related techniques, 423–425
MCR-ALS Simultaneous analysis of multiple correlated data matrices, 440–445
range of feasible solutions, 446–447
resolving diverse chemical problems and, 448
self-modeling curve resolution extended to multiway data, 440–445
software, 465–467
spectroscopic images, 461–465
uncertainty in resolution results, 446–447
useful methods and, 441
multivariate curve resolution-alternating least squares (MCR-ALS), 432
multivariate curve resolution-alternating least squares MCR-ALS, 439
multivariate curve resolution-alternating least squares (MCR-ALS), 439
multivariate curve resolution-alternating least squares MCR-ALS, 456–459
procedure, 439
multivariate curve resolution, applications, 448–453
biochemical processes, 449–450
global description of the protein folding process, 453
study of changes in the protein secondary structure, 451–452
multivariate curve resolution, iterative methods, 431–439
constraints based on chemical or mathematical properties, 433–437
constraints, classification, 433–437
constraints, definition, 433–437
constraints equality and inequality, 433–437
generation of initial estimates, 432
iterative target transformation factor analysis (ITFA), 437–438
multivariate curve resolution-alternating least squares (MCR-ALS), 439
multivariate curve resolution, noniterative resolution methods, 426–429
Heuristic Evolving Latenet Projections (HELP), 429–430
other techniques, 429–430
Subwindow Factor Analysis (SFA), 429–430
window factor analysis (WFA), 427–428
multivariate data, 219
parameters and, 229
multivariate dataset, 218
matrix form and, *54*
multivariate detection limit, 134
multivariate distances, 59
"acceptable" and "unacceptable" objects, 64
calculate Mahalanobis distances and probability densities, 64
compute training set mean and variance-covariance matris`, 62
example of, 59–69
graphical review of smx.mat data file, 69

Au: matris not clear

histograms of variables (wavelengths), 69
selection of, 60
multivariate distances, example
step 1: graphical review of smx.mat data file, 60
step 2: selection of variables (wavelengths), 61
step 3: view histograms of selected variables, 61–62
step 4: compute the training set mean and variance-covariance matrix, 62
step 5: calculate Mahalanobis distances and probability densities, 64
step 6: find "acceptable" and "unacceptable" objects, 65
multivariate limit of detection (LOD), 134
multivariate normal distribution, 3, 41–68, 51, 51–52
assumptions, 55
bell-shaped curve and, 51
chi-square distribution, 56
estimation of population parameters from small samples, 54
generalized or Mahalanobis distances, 52
generalized sample variance, 55
graphic illustration of selected bivariate normal distributions, 56
variance-covariance matrix, 53
multiway data analysis approaches, 461

N

N-way data, 478
"net analyte signal" (NAS)
calibration and, 496–497
net analyte signal (NAS) vector, 134
Newton-Gauss algorithm, 233
flow diagram of, *233*
Newton-Gauss-Levenberg/Marquardt (NGL/M), 230–237
algorithm, 230, 234, *235*, 235
fitting procedure, *235*
routine, 235
Newton-Gauss method, 233
NIR absorbence spectra of water-methanol mixtures, *109*
NIR (near-infrared-reflectance spectra), 60, 510
of sulfamethoxazole powder, 59, *60*
NIR spectra projected ino a three-dimensional space formed by the first three principle components, ***335***
NIR water-methanol data set, 141
noise
amplitude, 380
variables, 266
noise removal and the problem of prior information, 380–381
signal estimation and signal detection, 381
noise removal, matched smoothers and, 397
noisy spectrum to be enhanced, *381*
nomadization applied to Raman spectra, *83*
non-zero intercepts, *109*, 110
non-zero mean noise effects, 399
noniterative resolution methods, HELP method, 431
nonlinear fitting, using Excel's Solver, 227
nonlinear multivariate model estimation, chemical kinetics and, 4
nonlinear parameter fitting, 241
nonlinear regression, 477
nonnegativity, 433
nonsymmetric optimal designs, 298–306
optimal vx. equally distanced designs, 299–301
optimality criteria, 298
nonsymmetric optimal designs, design optimality and design efficiency criteria, 302–302
A-Optimality, 306
D-Optimality and D-Efficiency, 304
design measures, 303
E-Optimality, 306
G-Optimality and G-Efficiency, 305
nonwhite noise, 237
normal distribution
plot of, *14*
normalization, 83
null hypothesis, 19, 48–49, 122, 133
acceptance and rejection errors, 122
numerical approximation of the first derivative on near-infrared spectra of water-methanol mixtures, 82
Nyquist criterion, 388

O

octane data set with NIR absorbence spectra
outlier maps and, 205
one variable at a time approach, 288–290
advantages of, 290
applied to a two-factor problem, *289*
disadvantages of, 290
illustration of, *289*
one variable at a time vs. optimal design
bivariate (multivariate) example, 288–289
one-variable at a time approach, advantages of, 290
one-variable at a time approach, disadvantages, 290
one-way data, 476
operating region, 268
optimal composite sequential designs (OCSD), 314
schematic representation of, *314*

for three mixture variables and one process variable, optimal for models having structures in equations, ***321***
for two process variables and two polynomial models, *315*
optimal sequential designs (OSD), 314
ordinary differential equation solvers (ODE-solvers), 245
orthogonal projection approach (OPA), 432
orthogonal signal correction (OSC), 384
other calibration methods, 150
common basis vectors and a generic model, 150, 150–151
continuum regression (CSR), 150
cyckuc subspace regression (CSR), 150
generalized RR (GRR), 150
ridge regression (RR), 150
ridge variations of PCR, PLS, etc., 150
outlier maps, 185, 194, 199
of artificial regression data, *181*
of the biscuit dough data set obtained with RPCR, *201*
of glass data set based on three principal components computed with CPCA and ROBPCA, *195*
PCA analysis results and, 191–192
of robust residuals vs. robust distances of the carriers for the foam data set, ***186***
for the stars data set, *182*
outlier maps, three-dimensional
PCA and regression outlier maps and, 202
outliers, *32*, 168–169
experimental, 33
Grubbs' tests, 34
orthogonal, 191, 194
and outlier map of a three-dimensional data set projected on a robust two-dimensional PCA subspace, *192*
vertical, 184

P

PARAFAC, 481, 494–495, 495, 496
Algorithm (Appendix 12.3), 504
application, 494
relative speed of, 499
PARAFAC / CANDECOMP, 491–494
algorithm, 491
PARAFAC application, 494
solution constraints, 493
tuckals, 493
PARAFAC model, 491, 498
as trilinear, 495
parallel factor analysis (PARAFAC), 4, 455, 475, 478–480
parallel vector analysis (PVA0, 428
parameters
linear, *228*
molar absorptivities, 228
non-linear, *228*
rate-constants, 228
Parseval's theorem, 391
partial least squares (PLS), 112, 140, 147, 147–149, 196
applications, 352
comparison with PCR, 149, 149–150
mathematical procedures, 148–149
method summarized, 352
modeling method, 352
number of basis vectors selection, 149
regression method, 352
software, 352
partial least squares (PLS) regression, 202, 477
classical PLSR and, 202
an example, 204
model calibration and validation, 204
robust PLSR and, 202
partitioning of variance in least-squares solutions, 123
pattern-recognition, 355
pattern-recognition methodology, 341
classification, 341
clustering, 341
mapping and display, 341
pattern-recognition techniques, applications of, 355, 355–371
archaeological artifacts, 356–357
class-membership problem, 355
classification, 355
data preprocessing, 355
feature selection, 355
fuel spill identification, 358–364
mapping and display, 355
PLS models, 355
principle component analysis, 355
self-organizing map (SOM), 355
sorting plastics for recycling, 365–370
taxonomy based on chemical constitution, 371–373
PCR, PLS, and RR harmonious plots for methanol, *152*
PCR, PLS, and simplex harmonious plots for NI analysis of moisture in soy samples, *153*
peak-matching program, 361
pharmaceutical manufacturing process, 43
piecewise direct standardization (PDS), 158
pixels, 461, 462
PCA analysis and, 463
plots

log 19 of singular values vs. estimated concentration for RAFA, 484
NIR spectra of water-methanol mixtures, 86
residual spectra from the training set and the residual spectrum for an unknown w spectrum, *99*
of second and third largest principal components for jet fuel samples. *362*
simulated two-component mixture spectra (solid line) and a spectrum contaminated with a third unknow component, *99*
two largest PLS components in the training set (bees), 373
two largest principle components for jet fuel samples, *361*
two largest principle components of 10 metals in obsidian samples, *356*
two largest principle components of the training set, *369*
PLS, 514
map from training-set data (honeybees), *374*
models, 336
point, measurement and, 302
polynomial least-squares filtering, *404*
polynomial moving-average (Savitsky-Golay) filters, 403
polynomial smoothing, 81–82
polynomial smoothing on near-infrared spectra of water-methanol mixtures, *81*
polynomials
canonical form of, 272
population, 42
in chemistry, 43
finite, 42
infinite. 43
mean, 42
mean vector, 54
standard deviation. 42
population parameters. estimation of
from small samples, 54
population variance, 47
population variance-covariance matrix, 54
possible residual plots with the estimated residual on the y-axis and he estimated concentration on the x-axis, 130
practical considerations in experiment design, 354
choice of classifier, 354
choice of training set. 354
controlling experimental variables, 354
feature selection, 354–355
validity set, 354
precision and accuracy
four common scenarios, *13*
precision, defined, 10. 12
predicted calibration spectra for product P1, 330
predicted standard and mixture concentrating from DTLD and PARAFAC, ***490***
predicted sum of squares (PRESS) statistic, 193–194
prediction confidence intervals, 127–128
prediction set
for honeybee example, ***373***
for jet fuel example, ***359***
for plastics sample, ***368***
prediction-set results (F-values) for Jet Fuel sample, 366
prediction-set samples projected onto the principle component map, *369*
predictive alternating conditional expectations (PACE), 477
preprocessing options, 77
baseline correction, 80
first and second derivatives, 82
mean centering, 77, 78
multiplicative scatter correction (MSC) and, 83
normalization, 83
smoothing and filtering, 81
standard normal variate (SNV) transforms, 83–85
variance scaling, 77, 78
PRESS statistic, 198
principal component analysis (PCA), 69–104, 70
chemometrics and, 3
principal component regression (PCR), 112
principle component analysis (PCA), 70–102, 140–153, 189, 347, 351, 422, 456–459, 514
alternataive formulations of, 77
applied to chromatographic-spectroscopic data, 70
axes defining basis vectors, *344*
basis vectors, 141, 144
classical PCA (CPCA), 185
conclusions, 102
data exploration procedures, 86
example PCR results, 145–146
introduction to, 70
loadings for raw mean-centered augmented data matrix, *459*
mathematical procedures, 142
outlier map, 199
outliers, 193, 198, 199
PCA data exploration procedure, 86
preprocessing options and, 77–85
score plots of the water-methanol mixtures, *88*
scores for raw mean-centered augmented data matrix, *460*
principle component analysis (PCA), basic vectors, 96–98
clustering and classification with PCA score plots, 98

principle component analysis (PCA), influencing factors, 87–88
distribution of error in Eigenvalues, 93
F-test for determining the number of factors, 93
variance and residual variance, 89–92
principle component analysis (PCA) outlier map applying RPCR to the biscuit dough data set
regression outlier map and, *201*
principle component analysis (PCA) outlier map of octane data set obtained with RSIMPLS and SIMPLES
regression outlier map obtained with RSIMPLES and SIMPLES, 205–206
principle component analysis (PCA), residual spectra, 98–101
residual variance analysis, 100–101
principle component map, jet fuel data, *362*, 363
principle component model, 73–76
alternative formulations of, 77
Eigenvectors and Eigenvalues, 74–75
singular-value decomposition, 76
principle component model, diagram for the chromatographic-spectroscopic data set shown in Figure 4.1, *73*
principle component regression (PCR), 140–141, 195, 477
basic vectors, 141
classical PCR, 194
linear regression models and, 194
methanol, *145–146*
method, 196
model calibration and validation, 198
number of principal components and, 198
principle component regression (PCR) applications, multivariate calibration and, 195
principle component regression (PCR), mathematical procedures, 141–142
calibration steps, 142
example PCR results, 145
number of basic vectors, 144
unknown prediction steps, 143
principle component regression (PCR) models, 336
prediction error, estimating, 198
validation of, 198
principle components, selecting the number of, 193
probability distribution function, 43
process analytical chemistry, applications, 512
process variables, transformation formula, 269
products with the respective quantities of the components in each, ***323***
projection of a three-dimensional vector a onto a two-dimensional subspace, *96*
projection pursuit (PP) algorithm
RAPCA (relection algorithm for principal components analysis), 188–189
projection pursuit (PP) techniques, 188
protein folding, *451*, 451
changes in tertiary structure and, 453
global description of the protein folding process, 453
protocols for planning experiments, 3
pulse radiolysis, 218
pulsed gradient spin echo (PGSE), 495
pure-component NIR spectra for water and methanol, ***116***

Q

quadratic discriminant rule (RQDR), 209
quality, illustrated with distributions of important quality characteristics, *296*
QZ algorithm, 486

R

RAFA, 489
application, 483
Raman
images, 462
spectra of plastic, *367*
rank annihilation factor analysis (RAFA), 476
rank annihilation methods, 482–490
direct trilinear decomposition (DTLD), 489
direct trilinear decomposition (DTLD) application, 490
generalized rank annihilation method (GRAM), 485
generalized rank annihilation method (GRAM) application, 486–488
rank annihilation factor analysis (RAFA), 482
rank annihilation factor analysis (RAFA) application, 482–484
RAPCA (relection algorithm for principal components analysis), 189
rate constants, 228
reactions, physical properties and, 2
energy transfer, 2
phase transitions, 2
temperature, 2
recovery of concentration compound by window factor analysis, *428*
"reduced" product formulations after removal of UV-Inactive components, ***326***
reexpressing data in alternate bases to analyze structure, 381–384
projection-based signal analysis as signal processing, 383–384

regression
deepest, 183
depth, 183
outliers, 198
regression analysis notation, 286
regression coefficients, 326
regression model
identifying the type of, 325–326
regularization, 151–152
Tikhonov regularization, 152, 153
regularization results: example, 153
relative speed (in GIgaFLOPs) for the discussed three-way methods), 498
repeatability, defined, 10
reproducibility, defined, 10
residual spectra, 98
residual sum of squares (RSS), 136
residual variance analysis, 100
example, ***102***
residuals, 128
studentized, 129
resolution of a multicomponent chromatographic HPLC-DAD, *420*
resolution results, ambiguity in, 447
response-surface modeling (RSM), 264, 265, 265–286
design of experiments (DOE) and, 265
general scheme of, 265–267
methodology, 267
model fitting and, 265
process or product optimization, 265
regression-analysis-related notation and, 286–287
response-surface modeling (RSM) and experimental design, 256, 263–339
algorithms for realizable optimal experimental designs, 306–315
example: application of DOE in multivariate calibration, 321–329
nonsymmetric optimal designs, 298–306
off-the-shelf software and catalogs of designs of experiments, 316–320
one variable at a time vs.optimal design, 288–290
symmetric optimal designs, 290–294
Taguchi Experimental Design Approach, 294
response-surface modeling (RSM), factor spaces, 268–285
constrained mixture spaces, 279–282
mixture and process spaces, 283–286
mixture factor spaces, 269–271
process factor spaces, 268
simplex-centroid designs, 275–278
simplex-lattice designs, 272–274
results from factor analysis of simulated chromatographic data, ***94***
results from univariate calibration of the water-methanol mixture, ***118***
ridge regression (RR), 112, 140
RMSEC values, 118, 119
RMSEV values, 118, 119
ROBPCA, 193, 197, 199, 202, 203, 211
algorithm, 190, 198
outlier map and, 194
projection pursuit and robust covariance estimation, 189
robust calibration, 159–160, 166–216
introduction to, 168
principle component analysis, 185
software availability, 211
robust calibration, classification, 207–208, 207–211
classification in high dimensions, 211
classification in low dimensions, 207–208
an example (spectra of three cultivars of cantaloupe), 209
in high dimensions, 211
in low dimensions, 207–210
robust calibration, linear regression in low dimensions
linear regression with one response variable, 176
linear regression with several response variables, 183
robust calibration, location and covariance estimation in low dimensions, 173–176
the empirical mean and covariance matrix, 173
other robust estimators of location and covariance, 176
the robust MCD estimator, 174
robust calibration, location and scale estimation, 162–172
the mean and standard deviation, 169
the median and the median absolute deviation, 169
other robust estimators of location and scale, 169
robust calibration, partial least-squares regression, 202–206
classical PLSR, 202
an example, 204
robust PLSR, 203
robust calibration, principle components analysis, 185–194
classical principle components analysis, 185
an example, 194
an outlier map, 191
robust PCA based on a robust covariance estimator, 187

robust PCA based on projection pursuit, 188
robust PCA based on projection pursuit and he MCD, 189
selecting the number of principle components, 193
robust calibration, principle components regression, 194–199
classical PCR, 194
an example, 199
model calibration and validation, 198
robust PCR, 197
robust calibration, software availability, 211
FAST-LTS, 211
FAST-MCD, 211
LIBRA (Library for Robust Analysis), 211
robust discriminant analysis, 210
robust linear discriminant analysis, 208
MCD estimates and, 208
robust LTS-subspace estimator, 191
robust multivariate methods, 4
robust PCA, 4
based on a robust covariance estimator, 187
based on projection pursuit, 188
based on projection pursuit and he MCD, 189–190
robust PCA methods
outliers and, 187
robust PCR, an example (biscuit data set), 198
robust PCR (RPCR) method, 197
regression techniques and, 197
robust PCA and, 197
robust PLC, 4
robust principal component analysis (RAPCA), 209
robust quadratic discriminant analysis (RQDA), 207
robust R-RMSECV curve for biscuit dough data set, 199
robust tolerance ellipses for the fruit data, 210
root mean square eror of cross-validation (RMSECV), 114, 198
root mean square error of calibration (RMSEC), 114
root mean square error of prediction (RMSEP)
computing, 114
RMSEV for validation and, 114
root mean squared error of estimation (RMSEE), 382
RPCR algorithm, 198
RSIMPLS
ROBPCA and, 202
RSIMPLS approach, 203
octane data set with NIR absorbence spectra, 204
Runge-Kutta algorithms
chemical kinetics and, 242
Runge-Kutta method, fourth order
in Excel, 242–245, *244*
in Excel, concentration profiles modeled, *246*

S

S-estimators, 183
S-PLUS software, 202
sample variance, 47
generalized, 55
sample variance-covariance matrix, 54
samples, 43
influential, 128
projected onto two-dimensional space, *343*
sampling and sampling distributions, 42, 42–44
normal distribution, 43–45
simple random sampling, 42
standard normal distribution, 45
sampling theorem, 387
sampling theorem and aliasing, 386–388
sampling theory, 41–68
Savitsky-Golay filtering, *404*
Savitzky-Golay smoothing, 81
scale plots, 410
scatter plot
of absorbence at 1912 nm vs. 2264 for he sulfamethoxazole training set, *63*
of the principal component (PC) scores from the SVD analysis of the water-methanol NIR mixture data, *142*
of the principal component scores from the analysis of the HPLC-UV/visible data set, *96*
score plot and 97.5% tolerance ellipse of the Hawkins-Bradu-Kass data obtained with CPCA and MCD, 187
second loading vector and calibration vector of sucrose for the biscuit dough data set computed with CPCR and RPCR, *200*
selected bivariate normal distributions
graphical illustration of, 56
selection of calibration and validation samples, 113
selectivity, 135
self-modeling curve resolution extended to multiway data
MCR-ALS Simultaneous analysisof multiple correlated data matrices, 440–441
self-modeling curve resolution (SMCR), 477, 479
self-modeling methods, 257–258
self-organizing map (SOM), 345, 351
disadvantages of, 347
distance not preserved in, 347
generated from the training set, *370*

semivolatile organic compounds (SVOC), 455, 456
sensitivity, 131
  analytic, 131
  calibration, 131
  effective, 131
  and limit of detection, 131
sequential D-Optimal experimental design for two process variables and a second-order polynomial, 312
sequential-replacement algorithms, 138
  forward-selection algorithm and, 138
  obtaining, 138
Shell foam data, 185
short-term Fourier transform (STFT), 406
signal estimation and signal detection, 381
  known signal and, 381
  root mean squared error of estimation (RMSEE), 382
signal processing
  convolution property and, 391
  digital filtering and, 4, 379–416
  serial, 400
signal processing and digital filtering, 379–416
  frequency-domain signal processing, 385–393
  frequency-domain smoothing, 394–397
  introduction to, 379
  noise removal and the problem of prior information, 380–381
  reexpressing data in alternate bases to analyze structure, 382–384
  time-domain filtering and smoothing, 398–405
  wavelet-based signal processing, 406–413
signal-to-noise ratio (SNR), 297, 322, 381, 411
  formulas for, 298
significance level, 50
significance testing, 18
  comparison of a sample mean with a certified value, 24–25
  comparison of the means from two samples, 24, 25
  comparison of two methods with different test objects or specimens, 26
  F-test for comparison of variance (precision), 19–21
  one-tailed or two-tailed tess, 24
  student t-test, 21, 22–23
significant factors, identifying the number of, 322
SIMCA method (soft independent modeling of class analogy), 211
SIMCA training-set results for Jet Fuel Sample, 365
similarity values, 348
simple moving average filter ("boxcar averager"), 401
simple-to-use interactive self-modeling analysis (SIMPLISMA), 432
simplex, 280
  algorithm, 226, 228, 234
  defined, 226
  unconstrained, 281
simplex-centroid designs, 275–276
  advantages of, 277
  disadvantages of, 277
  example (lithium lubricant study), 278
simplex-lattice designs, 272
  advantages of, 274
  component contour plots and, 274
  as composite designs, 274
  as D-optimal, 274
  disadvantages of, 274
  example, 275
  examples of for linear, quadratic, full cubic, and special cubic models, *273*
  homogenous mixtures required for, 275
simplex-lattice mixture design, four-component for a cubic polynomial model, ***276***
simplex minimization with three parameters, principle of, *226*
SIMPLISMA, 463, 465
SIMPLS algorithm, 202
simulated annealing (SA) and genetic algorithm, 138
simulated HPLC-UV/visible chromatographic data set showing two overlapping peaks with different UV/visible spectra, *72*
singular-value decomposition (SVD), 75–76, 188, 430
"slow-fast" ambiguity, 237
small sample distributions, 46
  chi-square distribution, 47
  t-distribution, 46
  t-distribution, the, 46
smoothed data
  after inverse Fourier transformation, *394*
  from application of trapezoidal smoother, 397
smoothing, 380, 394
  with designer transfer functions, 395–398
  in the frequency domain with a trapezoidal smoothing function, *396*
  of noisy signal with wavelet transform, *412*
  operation, 383
  time-domain and, 398
  transfer function and, 394
smoothing and filtering, *81*
  signal-to-noise ratio and, 81
soft independent modeling by class analogy (SIMCA), 354, 364
  classification in high dimensions, 353
  quadratic discriminant analysis and, 354

soft modeling, 345
software, 465
  curve-resolution algorithms and methods, 465
software applications, 4
  ANOVA using Excel, 39–40
  MATLAB and, 4
  Microsoft Excel, 37
solution for the integration of ODEs, 222
solvent interactions, 108–109
sorting plastics for recycling, 365
  pattern recognition techniques and, 365
  principle component analysis, ***368***
  Raman spectroscopy, 365
  self-organizing map (SOM), ***368***
spectra, 462
  first-derivative, 116–117
  pixels and, 463
  second derivative, 116–117
  of the UV-active components, ***324***
  of the UV-inactive components, *325*
spectral profiles, 476
spectral windows, 423
spectroscopic-chromatographic data, 70–72
  basic vectors, 71
spectroscopic images, 461–462, 462
  multivariate resolution and, 462
  pharmaceutical qualiity control and, 462
  resolution of, 463
  structure of, *462*
spectroscopic measurement, 256
spectroscopy, 477
  applications of, 512
  chemometrics and, 512
  near infrared spectroscopy and, 512
spectrum, *78*, 195
spectrum, techniques for obtaining, 195
  energy-dispersive x-ray fluorescence spectrometry (ED-XRF), 195
  florescence spectrometry, 195
  NIR (near-infrared_ reflectance spectrometry), 195
  nuclear magnetic resonance (NMR), 195
  ultraviolet sperometry (UV), 195
square sum ssq of the residuals r as a function of two parameters, *224*
squared correlation coefficient (angle cosign) between True and Estimated Profiles from DTLD and PARAFAC, 491
squared correlation coefficient (angle cosign) of predicted and true X-Way Spectra from GRAM with added matrices, ***488***
squared correlation coefficient (angle cosign) of predicted and true X-Way Spectra from GRAM with concatenated matrices, ***488***
squared correlation coefficient (angle cosign) of predicted and true Y-Way Spectra from GRAM with added matrices, ***488***
squared correlation coefficient (angle cosign) of predicted and true Y-Way spectra from GRAM with concatenated matrices, ***487***
standard deviation, 10
  population, 11, *13*
  relative standard (or coefficient of variation), 11
  sample, 11
standard normal distribution, 45
standard normal variate (SNV) preprocessing, *85*, 87
standard normal variate (SNV) transforms, 83
standard residuals of the stars data set based on a classical MLR and a robust LTS estimator, *179*
stars data, 180–181
stars regression data set with classical and robust fit, 177
statistical evaluation of data, 7–40
  analysis of variance and, 27–30
  common terms, 10
  introduction to, 8
  normal distribution, properties of, 14
  outliers, 33–35
  precision and accuracy, 12
  robust estimates of central tendency and spread, 36–38
  significance testing, 18–26
  software and, 38–40
  sources of error, 9
statistical evolution of calibration models obtained by least squares, 121
statistical modeling, 265
statistical properties, 3
  calibration of response signals, 2
  control of interfering factors, 2
  measurement process errors, 2
  modeling of complex multivariate signals, 2
strong filtering of noisy data using quadratic polynomials, *405*
structure of he extended design matrix F for a second-order model with two process variables, ***287***
structure of protein, *450*
student t-test
  paired t-test, 21
  standard t-test, 21
  t-test with nonequal variance, 21
subwindow factor analysis (SFA), 428–430, 429–430
  algorithm, 431
  application of, *430*
sum of squares (ssq), 123–124, 234

summary data for the analysis of Pb in orchard leaves, *49*
summary statistics for NIR calibration of water in water-methanol mixtures using one wavelength and a nonzero intercept, ***125***
sums of squares, 125
symmetric optimal designs, 290
symmetric optimal designs, central composite designs, 293–294
symmetric optimal designs, three or more levels in full factorial designs, 291292
advantages, 290symmetric optimal designs, two-level factorial designs, 290-294
disadvantages, 291
symmetry matrix, *349*
systematic error or bias, 17

T

t-distribution, ***16***
degrees of freedom and, *47*
t-test
comparison of a sample mean with a certified value, 24
comparison of the means from two samples, 24
comparison of the means of two methods with different samples, 24
Taguchi, 296
Experimental Design Approach, 294
methods, 295
orthogonal designs, 319
Taguchi design, 297
for two controlled factors and three noise factors, *297*
Taguchi, Genichi, *294*
quality engineering and, 294
target factor analysis (TFA) and, 437
taxonomy based on chemical constitution
Africanized and European honeybees, 371
gas chromatography and, 371
principle component analysis, 372
three-factor principal component model for NIR spectra of water-methanol mixtures, diagram of, *142*
three-way calibration, advantages of, 475
three-way calibration with hyphenated data, 475–499
alternating least-squares methods, 491–494
background, 476
caveats, 497
extensions of three-way methods, 495
figures of merit, 496
introduction to, 475
nomenclature of three-way data, 478
rank annihilation methods, 482–490
three-way data, 477
models of, 478–479
three-way data analysis, 443–444, 464, 477
calibration with, 477
constraints and, 445
examples of, 481
MCR-ALS decomposition method, 445
nontrilinear, 442
trilinear, 442
unfolding and, 441
unfolding or matricization, *441*
three-way data, bilinear models for
unfolded PCA and unfolded MCR, *444*
three-way data, nomenclature of
data as orders, 478
data as ways, 478
time-domain filtering and smoothing, 398–405
filtering, 400–402
polynomial moving-average (Savitsky-Golay) filters, 403–405
smoothing, 398–399
time-domain processing of noisy data with moving-average filter, *402*
time-domain smoothing of noisy data, *399*
training set
for the honeybee sample, ***372***
jet fuel sample, *358*
for plastics sample, ***368***
transfer functions for FIR filters as a function of window size, *403*
translated experimental design for product P1, ***329***
translation of coded factor levels to real experimental levels for product P1, ***328***
trilinearity constraint in MCR bilinear models for three-way data, *446*
true, best, and worst estimated X-way and Y-way analyte profiles from GRAM, *487*
"true values", 114
TUCK alternating least squares, 493
Tuckals, 493
Tucker3 model, 479–480, 493
two-way data, 476
two-way data matrix, augmentation of, *441*
types of observations based on heir robust distance (RD) and their Residual Distance (RD, ***185***)
types of observations based on score distance (SD) and orthogonal distance (OD), ***193***

U

uncertainty in resolution results, experimental error and, 448
uncertainty in resolution results, sources of, 446

ambiguity of recovered profiles, 446
experimental noise of the data, 446
unimodality, 434
Union of Pure and Applied Chemistry (IUPAC), 131
univariate calibration, 109, 118, 118
with an intercept term, 119
without an intercept term, 118, 118–119
univariate decision limit, 132
univariate detection limit, 133
univariate hypotheses testing, 48
inferences about means, 49
inferences about variance and the F-distribution, 51
univariate MCd, 193
univariate normal distribution, 3
unknown prediction steps
prediction step, 143
validation step, 143
UV absorption, 323
UV-inactive components, 323
UV spectra, 323
of all nine pure components, *324*
UV spectroscopy, 322
UV/visible spectrometer, 70

V

v-fold CV, 115
validation set, 208
variables, 10, 11, 44
between-sample means, 28
collinear, 344
continuous random, 43
discrete random, 43
heteroscedastic, 130
homoscedastic, 130
residual variance and, *89*
scaling, 78–79
selection, 135
within-sample means, 28
variance and the f-distribution
inferences about, 51
variance-covariance matrix, 53
viscose data set and standardized residuals obtained with different estimators of location and scale, ***170***
Visual Basic, 218

W

water-methanol mixtures, 108
wavelengths, 462
wavelet, 406
components resulting for discrete wavelet transform, *410*
sine basis functions for orthogonal transforms and, 408
truncation and, 412
wavelet analysis
Fourier analysis and, 408
wavelet-based signal processing, 406–413
discrete wavelet transform, 409–410
smoothing and denoising with wavelets, 411–413
wavelet function, 406–407
wavelet function, time and frequency localizations of, 408
wavelet family, 407
vanishing moments and, 407
wavelet functions, 411
coiflet wavelet families, 411
Daubechies wavelet families, 411
symlet wavelet families, 411
time and frequency localizations of, 408
weak filtering of noisy data using quadratic polynomials, *405*
white noise, 222, 397
defined, 237
Wilson-Hilferty transformation, 192
window factor analysis (WFA), 427–428
steps of, 427–428
within-run, defined, 10
Wold, Herman
partial least squares (PLS) and, 352, 510
Wold, Svante
chemometrics and, 509
partial least squares (PLS) and, 510
Wynn-MItchell and van Schalkwyk algorithms, 308

X

X-way and Y-way profiles used to construct data sets for each example, *481*
Xerox, 294

Z

Zeroth-Order reaction, 248–249
catalyst and, 248
macroscopically observed, 248