Available online at www.joac.info

ISSN: 2278-1862

Journal of Applicable Chemistry

2015, 4(5): 1313-1428

(International Peer Reviewed Journal)

Research tutorial (ResT)

[Computational] TensorLab (CTLab)

Part I: 3D-surfaces and 2D-contours in OmniMetricS

K. Ramakrishna[1] and R. Sambasiva Rao[2*]

1. Department of Chemistry, Gitam Institute of Science, Gitam University,
Visakhapatnam, 530 017, **INDIA**
2. School of Chemistry, Andhra University, Visakhapatnam 530 003, **INDIA**

Email: karipeddirk@gmail.com, rsr.chem@gmail.com

Accepted on 20th September2015

To
Mother Nature
Creator/ preserver/ assimilator of knowledge/ Information/ Data systems (Kids)

Conspectus

Background:Tensorial representation of numerical (integer, floating point) values, also called m-way data, is the basic input in data-information-knowledge cycle. Graphical output of primary data as scatter diagrams to model fitting, residual trend exploitation for better explanation are not only coveted visual appreciation tools, but explore misfits, unexplained/unexplored information etc. The reliable software no doubt result in parameters/ statistics of model mostly in the traditional mode or in limited cases with expert system (ES) driven inferences.

3D-surfaces and 2D-contours: The profiles of linear, full quadratic, polynomial, exponential, transcendental functions in two variables as 3D-surfaces, 2D-contours with gradients, rotating view angle (i.e. keeping one of the variable constant) such that surface reduces to 2D- plots are detailed.The depth and breadth of the response surface modelling strategy is highlighted.The popular graphics of data sets in vector/ matrix form using today's state-of-art-profiles are incorporated in appendices.

m-Way data generation:Zero to third order instruments, variation of influential experimental variables and external environmental factors in the interacting as well as non-interacting chemical systems generate one- to multi-way through 5-way data tensors. They are usually modelled with multi-variate methods in the unfolded modeor as they are. The possible unfolding modes and ill effect on end results are discussed. The nomenclature of numerical data under different heads like vector, matrices, tensors, multi-ways are brought under the same roof.

Applications in Omnimetrics:The role of multi-dimensional graphics of raw experimental data in exploratory data analysis and in various phases of multi-variate-multi-response-linear-nonlinear parametric-/non-parametric-/ free-variable models in prime disciplines viz. environment, foodomics, medical diagnosis, pharma industry, physical chemistry reported during last two years are incorporated. The typical case studies incorporated in this research tutorial (HOT Ice: hands on tutorial for intelligent *chemical education) include applications in quality of natural water (streams, estuaries etc.)/ artificial water reservoirs, water quality tap water, potable watertreatment,waste water treatment plants, aerosols, metabolic profiling, pollutants like PAHs, adulterants in lime fruits/fruit juices, wine discrimination andclinical analysis of urine. The DOM (dissolved organic matter) in different phases of environment drew attention of multi-variate methods and graphics. The estimation of rate/ equilibrium constants of metal ligand systems and outersphere/ innersphere complexes are improved by multivariate chemometric methods and contour diagrams. The excitation-emission fluorescence spectroscopy, 2D-NMR, HPLC-DAD etc. are prime second order hyphenated instrumental techniques employed in these studies.The 3D surfaces and corresponding contour diagrams in basic and inter-disciplinary chemical research are reproduced from research literature in the appendices.*

A brief note of a few multi-way chemometric methods are incorporated with a tinge of knowledge base/necessary conditions /failure instances and remedial measures. The statistical/fuzzy distributions, chaotic series, multiple global functions not discussed here will be described in a separate communication.

Keywords:3D-surface, 2D-contour, Tensorial data, Local/global extrema, iso-surfaces/contours, Kriging, Neural networks, Interpolation, omnimetrics, experimental design, Environment, Dietetometrics, physical chemistry, chemical physics, chemical biology.

Contents

3D-Surfaces & 2D-Contoursin OmniMetricS

INTRODUCTION

In yester-centuries, human race gathered data about themselves (inside and/or outside) and surroundings through theirfive senses. The progress in science, engineering and technology brought simple as possible (SAP) instruments like pH/ DO/ conductivity/calori-/ colori- meters, microscope to multi-channel-hyphenated-self-adaptive instruments and nanoscope. The highest and ultimate instruments to date include those in Mars exploration, Boson discovery, probing into dark-matter/dark-energy and largest telescopes, spectrometers to discover news in the galaxy and universe.

Remembering and reproducing what has been seen/felt is a natural bliss to humans through trivial activity of brain. The representation of data as attributes, nominals, numbers, pictures and images was a step forward in giving emphasis to next level of understanding than simple oral reproduction in toto. Mathematics, statistics, fuzzy sets etc. are the core in fitting data in curve fitting, modelling to grasp essential trends rather than sheer reproduction, knowledge extraction through model information are now basic tools in understanding, predicting, controlling and mimicking bio-/geo-/oceanographic-/atmospheric-/chemical-/bio-chemical- phenomenon/ rare events or even calamities. Algebraic-, differential-, matrix-, tensorial-, symbolic-computations are the corner stone in deriving information from primary data in most of disciplines. Euclidean geometric representation of data, model parameters, information etc. is a high impact output for grasping broad picture (or trends) on one hand, and probing deep into what is not easily graspable from bundles of numerical data. The man is the best pattern recogniser even in three dimensions and all intelligent machine learning techniques (of course manmade) are marching fast towards this novel simple goal in nature. The graphical representation in two dimensional Cartesian spaces is a popular well-practiced tool for exploratory data analysis. A detailed account of 3D- surfaces, 2D-contours from scatter plots, model development, trend analysis deserves attention, as unfolding 3-way tenors to sets of 2-way matrices and analysis is non-optimal/ misleading or even wrong.

The present day interdisciplinary tasks involve thousands of correlated, redundant, conflicting, non- orthogonal variables on different scales with missing values and so on. Thus, data/ dimension reduction techniques without losing information involve more number crunching, of course, in a crutch mode, as it not possible to probe into every tiny detail. The theoretical proof the methods in scaling up operation are the guidelines in this pursuit. Here, too multi-dimensional graphics have a key role throughout data processing. The 3D- display of total electron density, ESP, HOMO/LUMOs in computational quantum chemistry gave a new life to organic chemicals, bio-materials compared to yester years ball-and-stick models. The display is computationally intensive job and computer is indispensable.

The prime goal of any laboratory or industrial experiment is to obtain drugs with minimum side effects, conversion of harmful materials and pollutants into non-hazardous ones, preparing materials of desired characteristics (semiconductors to superconductors) and developing products useful in transportations and enjoyment with optimum cost.A major part of chemistry is devoted to isolation of organic compounds and other materials through inexpensive routes from abundantly available natural resources and synthesis. In order to achieve the desired characteristics trillions of chemical experiments are performed by chemists varying physical, chemical and biological factors.

Some of the important chemical tasks pertinent in industry as well as training a chemist for twenty first century to face the challenges of biological and environmental tasks are (a) complete/partial separation of a mixture of compounds into their constituents (isomeric compounds, metal ions, pollutants), (b) quantitation in partially (if not completely) unknown matrices in presence of interferences, (c) partial synthesis of organic or inorganic compounds, their structure elucidation and (d) reactivity and stability of different chemical species. In an attempt to achieve one or more of these goals destructive or non-destructive chemical techniques are adopted employing instruments of increasing sophistication. The raw data obtained in these venture range from a zero order (scalar), first order (vector), second order (matrix), third or a multi-order (tensor) instruments. When the size of the data (chart 1.1) is small, a careful inspection ofthe numbers results in chemical information. But with the explosion of the data with computerized instruments, adopting statistical methods is a rule (a natural way) rather than exception

The 3D-surfaces and 2D-contours in chemical sciences [1-125] is the main focus of this research tutorial. The basics of multi-dimensional graphics and typicalpublished literature were described earlier [57]. In continuation of our efforts in Omnimetrics [63, 65-73] including nature inspired algorithms [70-72], the current research tutorial in chemical sciences includestensorial data display, isocontours/ 2D-profiles/ surfacesfor simplest possible model (z = 0) to polynomials and transcendental/exponential functions. The salient research results during the last couple years in Envirometrics, Dietetometrics etc. using PARAFAC model and Ex-Em_Fl spectroscopy are described with a focus on multi-variate analysis. The popular 2D-graphic displays are summarized citing researches in environmental monitoring.The profiles of probabilistic, fuzzy logic and chaotic frame will be described elsewhere[113].

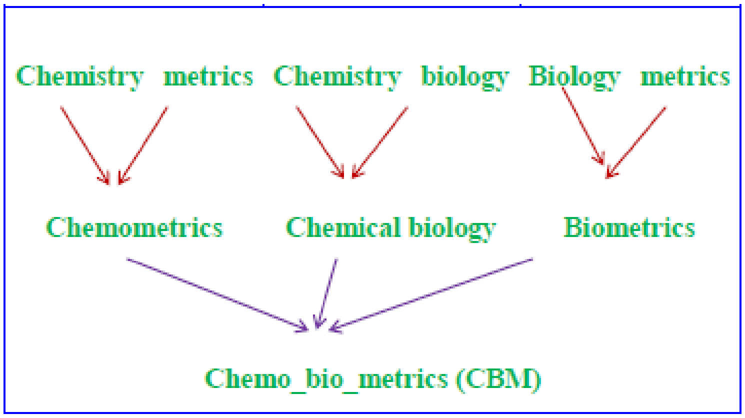
Chart 1.1: Observations & Data

- Observation
  - Human senses: Eye, ...
- Instruments
  - Single
  - Hyphenated
  - Satellite
  - Sky Lab
- Nature
- Simulation
  - Mathematical Functions
  - Wavelets, Ridgelets
  - Statistical distributions
  - Fuzzy sets
  - Possibility profiles

Chart 1.1b: **Challenging disciplines**

- Proteomics
- Foodomics
- Metabolomics
- Petroleomics
- Proteins
- Glycoproteins
- Metabolites
- Petrochemical products

- Chemistry
- Physics
- Biology
- Paleology
- Future



2. Tensorial (or multi-way) data

Tensor operations started a century ago in the hands of mathematicians. It took nearly five decades to enter into applied science in a small way. Physicists started using tensor algebra earlier to others. The geometric representation of tensorial data was brought to the forefront from Kowalski's school [16,33]. Differentiation of matrix equations was in higher strata compared to long algebraic derivations. Hamilton's geometric algebra was in archives for over a century. Instead, Maxwell differential equations have become bread winning tools. The geometric representation leave alone visualization/comprehension/integration in the brain like a mother tongue, native/learned logic for most of end users is a pie in the sky. It is worth practicable paradigm to visualize process, physical/chemical/biological model and appropriate mathematical equivalent and the best possible solution effortlessly just like distinguishing airplane, train, bullock cart, bicycle etc. and seeing their form and components. In the next generation if not in coming decade, this should be the common sense base in any discipline just like blind touch keying in of English alphabets. One should visualize effortlessly (with comfort) algebra, tensorial manipulation, 3D-graphic profiles, algorithmic sequences of steps in both or any direction, although now, it is with well learned expert professionals.Particularly in chemistry, Rasmus Bro' s seminal Ph.D. thesis [114]and follow up dissertations [115-117] opened a new era for tensorial manipulations in curve resolution and calibrationfrom 1990s and now reached a stage of a tool in the hands of Omnimetricians to comfortably arrive at understandable graphic output through the glasses of their core disciplines.

Data to intelligence cycle

The different categories of collections of method_Bases, data structures, and implementation in MATLAB environment are briefed in Appendix-A1. The concerted efforts in instrumentation, generation of m-way data tensors, nano-electronics to nano-chemistry, information science, signal processing, data massaging, data driven models, nature inspired algorithms on one hand and transferring/translating them into chemometric frame changed the mold of measurement science (Appendix-A2 to appendix-A6).Unfolding of tensorial data, models for different tasks in mathematical, chemical domains, typical chemometric models in m-way supervised and unsupervised analyses follows mostly in tabular and object oriented format.The combination (including fusion) of outputs of different instrumental techniques dramatically enhanced the reliability of classification/discrimination, calibration/prediction even in presence of unknown interferents/matrices of foodstuffs/ beverages as compared to using a single analytical technique that to classical univariate calibration of single pure analyte using one-variable-at-a-time(OVAT) approach without experimental design at any stage. However, the combination of data from several techniques is not straightforward and is challenge for chemometricians to train armatures into skilled personnel and more importantly impart formal education to undergraduates with a tinge of handling real life datasets in tutorials session. This should be accompanied by hands-on-tutorials (HOT) [113] intelligent chemical education (ICE) with simple as possible datasets and white box approach software with a focus of tensorial notation and corresponding geometric interpretation.

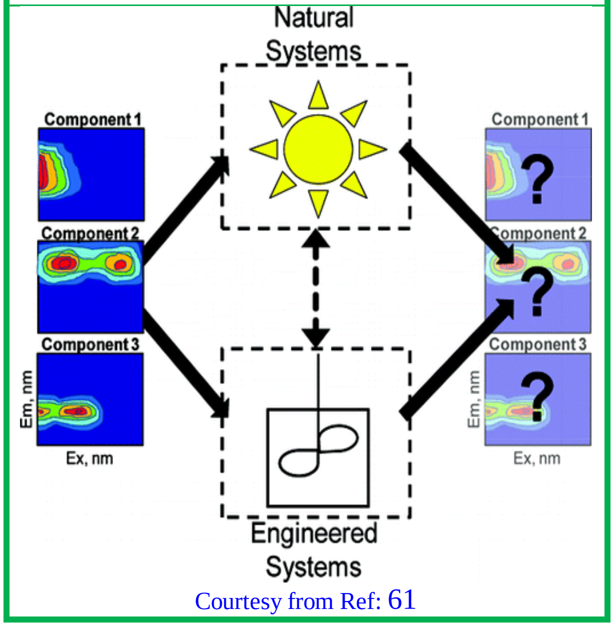
Tensor Laboratory for Chemistry (TLC): It is a conglomeration of mathematics, translation of chemical tasks into mathematical space, hyphenated instrumental data acquisition and solution methods.The mathematical component consists of tensor algebra and solution methods.Tensor algebra relieves the drudgery of representation of multi-way data, mathematical formulae, their traditional derivations, decomposition into lower dimensional sizes, solution methods, accumulation operations (addition/multiplication etc.). A chemist immediately remembers thin layer chromatography which is a popular experimental procedure (simple enough requiring a glass place, coating material, a small roller) of puristorganic chemists of last century. This simple, no high tech, is a powerful separation technique to detect and later determine the mixture of compounds. The state-of-art-of TLC of 21st century is high ended imbibing expert systems, technology etc. Today, 2D-LC, HPLC, UPLC etc. does the same, of course, many more intricate separations not possible by simple TLC.

3.Applications of multi-dimensional surfaces and contours in Omnimetrics

The chemical tasks are broadly divided into non-interacting and interacting chemical systems. The case studies that follow are those reported in literature in different disciplines of chemical research. Hyphenated instruments and multi-variate calibration/curve resolution methods revolutionised the understanding of chemical systems in natural environment (fig.1)

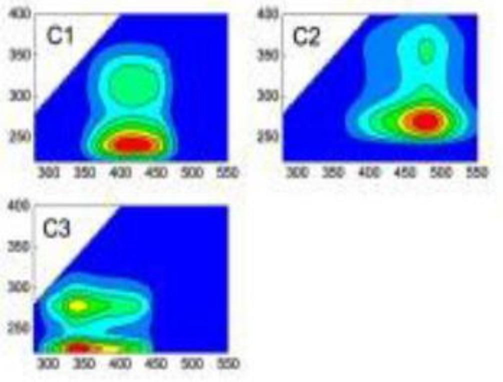
Natural organic matter in Songhua River water: Shao et al. [87]detectedtwo humic-like (C1 and C2) and one protein-like (C3) components in NOM fractions of Songhua River from PARAFAC analysis of ex-em-fluorescencespectra (fig. 3.2). These components give foul smell to ultrafiltration membrane.

Fig.3.1: Chemometric approach to natural systems



Courtesy from Ref: 61

Fig.3.2: Ex-Em-Fl spectra of NOM in Songhua River



Courtesy from Ref:87

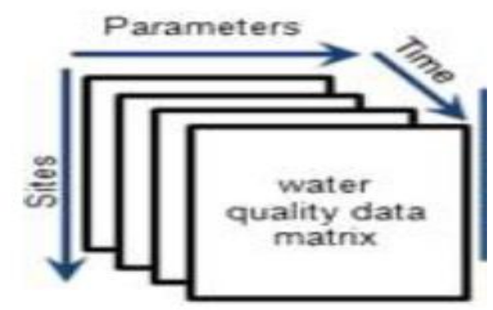
Artificial water reservoir:Cid et al. [14]applied PARAAC, Tucker-3 and matrix augmentation _PCA, N-way-PCA models for 3-way water quality data (sampling sites × parameters × sampling time) in Midwest of Argentina (chart 3.1). The first PARAFAC factor implies organic pollution with seasonality and the second one has lead pollution information. A plot of PARAFAC loadings as a function of the UTM (Universal Transverse Mercator) coordinates depicts polluting sources and areas in the reservoir.

Chart 3.1: 3-way water quality data in Argentina with PARAFAC model

Surface water samples

- # points of the water reservoir : 38
- # sampling campaigns :eleven
- Period : October 1998 to June 2000
- Seasons: warm wet, cold dry season

X(38,30,11)



Water quality parameters

- DO, conductivity, pH
- COD, BOD
- Viable aerobic bacteria
- Total coliform bacteria
- Ammonium, nitrate, nitrite
- TDS, alkalinity, hardness
- bicarbonate, chloride
- sulphate
- Calcium, magnesium
- fluoride
- Sodium, potassium, iron
- Aluminum, silica, phosphate
- Sulfide, arsenic, chromium
- Lead, cadmium

Methods_chemomet

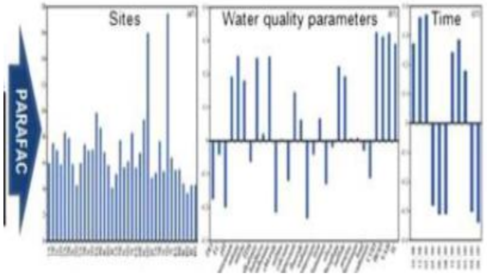
- PCA
  - Matrix augmentation
  - N-way-
- Tucker3
- PARAFAC

Inf.Bits.

− Tucker3 model not appropriate for the dataset

⊥ MA-PCA accounts for organic pollution parameters

\+ PARAFAC (two-factor) model projects best picture



Courtesy from Ref:

Dissolved organic matter (DOM): Soil organic matter (SOM)/ dissolved organic matter (DOM) is a major reservoir of reduced carbon. It has a vital role in many ecosystem processes viz. nutrient supply, metal solubulization, sediments, carbon sequestration and it's cycling in the coastal-shelf-ocean boundaries. The production, structure/ composition and transformation of watershed-derived dissolved organic matter (DOM) have been increasing over several decades. The wetland structure and climate change have definite roles on it. However, it carries information on changes in sea and landscape modifications over long periods of time. The beneficial applications include in assessing formation of disinfection byproducts in chlorination/treatment of drinking water sources. The complexity, however, is a hurdle to monitor its concentrations and character during drinking/ waste water treatment processes. Further, the distribution and transformations of biogeochemical changes in the atmospheric boundary involve multitude of processes and as such probing into natural organic natter (NOM) in atmospheric aerosols remained to be a challenge. In this context, the classical univariate/ multi-variate curve resolution/calibration/prediction analytical/instrumental/chemometric tools are inadequate/ invulnerable to arrive at conclusive trustworthy inferences. In this decade, parallel factor analysis (PARAFAC, PARAFAC2), quadrilinear decomposition, residual tri-linearization etc. of fluorescence excitation–emission (EEFs) of 3-way/4-way tensors threw light on correlating mathematical factors with chemical moieties and probing into concentration profiles in presence of unknown interferents in the test samples. Further, knowledge extraction from blind source apportionment of factors/ clustering methods drew the attention of cross-disciplinary research scientists.

Ishii and Boyer [61] reviewed results of studies on analysis of humic-like components in DOM of natural systems with PARAFAC from published literature after the year 2000. It is opined that fluorescence spectroscopy with PARAFAC is a weak predictive tool.Mendoza and Zika [88] resolved fluorescent chemical moieties inDOM from near-shore to off-shore coastal-shelf-ocean boundary on the southwest (SW) Florida by PARAFAC model (chart 3.2). The humic- and protein- components of different origin are detected in DOM here.

Chart 3.2: DOM analysis with PARAFAC model

Samples from cruises ▶ May and October 2008 ▶ January and April 2009	Model.Inf.Bits. - Model not appropriate for the dataset ⊥ Accounting for organic pollution parameters + (Two-factor) model projects best picture	Models Tucker3 MA-PCA PARAFAC

Inferences of PARAFAC resolution

Humic-like components	component 1	Natural
	component 2	anthropogenic
Unidentified component	component 3	
Tyrosine-like	component 5	

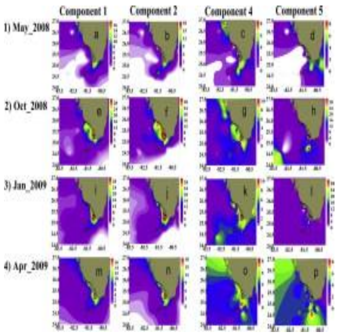
Output

○ Spatial and temporal distribution maps for each components on the shelf

○ Biological, chemical, and physical processes controlling DOM variability

Inf.Bits.

- Evidence of offshore transport of land-derived materials.



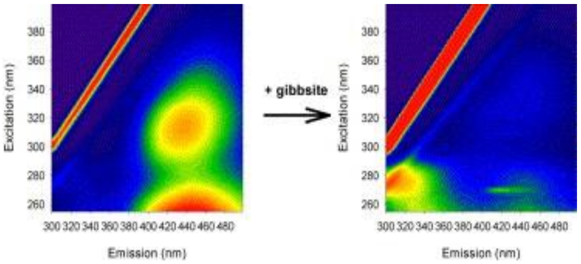
Courtesy from Ref:88

Banaitis et al. [85] studieddissolved organic matter (DOM) from soil organic horizons and tree leaf tissues with ex-em-fluorescence spectral data modeling with PARAFAC (chart 3.3).

Chart 3.3: DOM in soil organic matter

Inferences of PARAFAC resolution

		peak location	
	$$$-like	excitation (nm)	Emission (nm)
1	Tryptophan	<255	342
2	Tyrosine	276	312
3	Humic-substance	255	456
4		309	426
5		255	401



Courtesy from Ref:85

Inf.Bits.

\+ PARAFAC of 3-way (Ex-Em-Fluorescence) data quantitatively describe the chemical fractionation of DOM with mineral surfaces

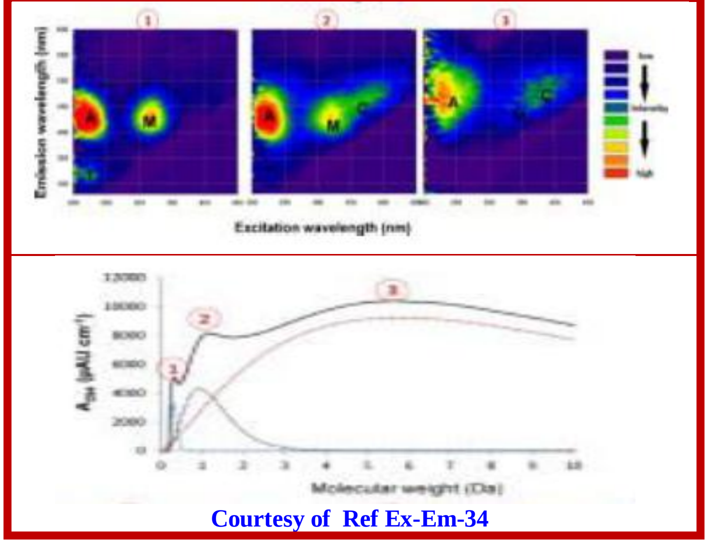
Cuss and Guéguen [13] correlated fluorescing components with molecular weight fractions/size of dissolved organic matter (chart 3.4).

Chart 3.4: Correlation of fluorescence with molecular size

Instruments	Experimental Procedure
1) Ex-Em-Fl **2)** UV-Vis	▶ Asymmetrical flow field-flow fractionation

Chemometric_methods	
☐ PARAFAC	☐ Statistical fractogram deconvolution

Tyrosine/polyphenol-like fluorescence	smallest size group (relative molecular weight = 310 ± 10 Da)
Microbial humic-like and terrestrial visible humic-like fluorescence	Intermediate size group (1600 ± 150 Da
Terrestrial fulvic-like and tryptophan/polyphenol-like fluorescence	Largest size group (4300 ± 660 Da).



Courtesy of Ref Ex-Em-34

Cuss and Guéguen [57] studied size and optical properties of leachates of unfractionated and AF4-fractionated DOMs over 72–96 h to probe into molecular weight distributions (chart 3.5).

Chart 3.5: AF4-fractionated DOMs

Instruments	Chemometric_methods	
1) Ex-Em-Fl **2)** UV-VIs	☐ PARAFAC	☐ PCA of proportional component loadings

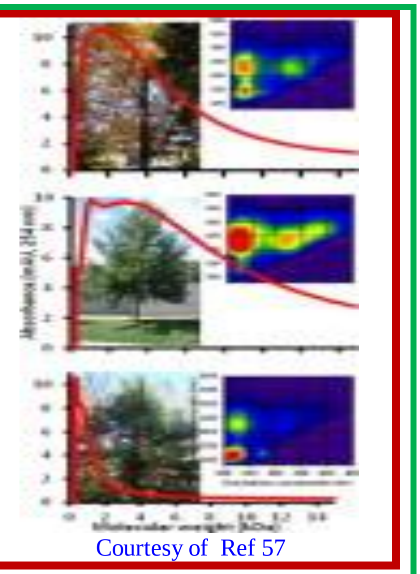
Information.Bits.

- 🔔 Unfractionated or fractionated leachates are highly correlated with PC_loadings

#Loading	R^2
1	0.85
2	0.95

Inference.Bits.

- 🔔 Differences in optical (fluorescent) characteristics depend upon species composition



Courtesy of Ref 57

Meng et al. [111] attributed the four factors of PARAFAC model of water samples from Zhujiang River to amnioacids and humic substances (Chart3.6).

Chart 3.6: PARAFAC model for river water samples

Objective

- Spatio_temporal variations of anthropogenically impacted DOM urbanized Zhujiang River

Water samples

- Upstream
- Urbanized area
- Downstream

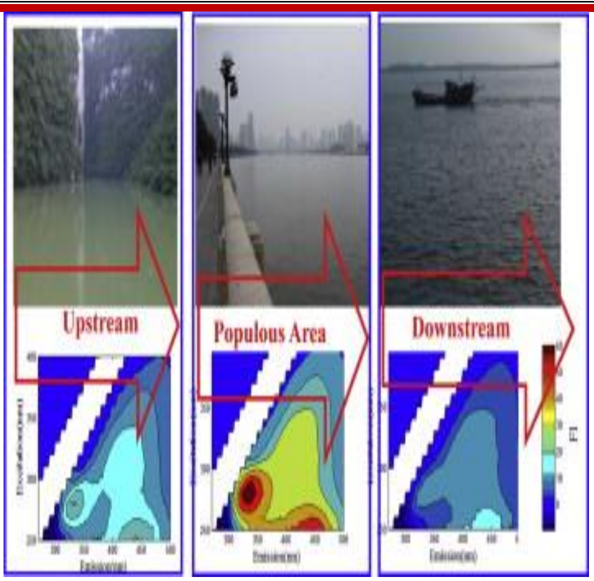
Period

- Different days of one year

Inference Bits.

- Four-factor PARAFAC

#factor	R^2
C1	Tryptophan-like
C3	Tyrosine-like proteins
C5	Anthropogenic humic substances



Courtesy of Ref 111

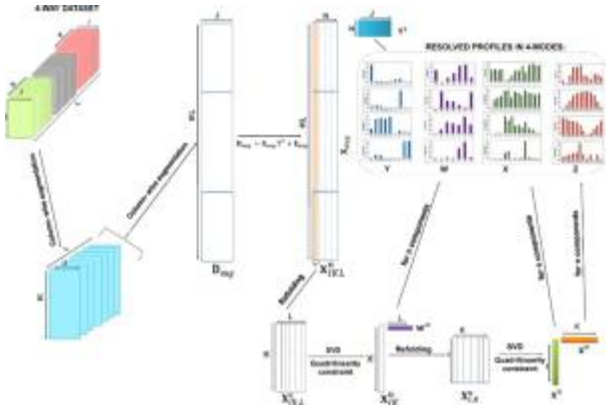
Modified MCR-ALS: Malik and Tauler [11] extendedMultivariate Curve Resolution-Alternating Least Squares (MCR-ALS) for the analysisof four-way datasets. The constraints imposed are non-negativity and quadrilinearity of the models. It is applicable for huge multidimensional datasets in curve resolution and summarizing fewer number major intrinsic factors which are functions of large number of variables involved. Theycompared the performance of MCR-ALS with PARAFAC for environmental four-way datasets from the most polluted Yamuna River, a largest tributary of Ganges (Chart 3.7).

Chart 3.7: Comparison of multivariate methods for four-way environmental data

4-way Water samples		
way	Variable	#
I	Sampling sites	15
II	Water parameters	9
III	Months	12
IV	Years	7

MCR-ALS resolved pollution profiles

- pH, temperature `
- Organic pollution
- Bacteriological pollution
- Spatial and temporal sampling sites



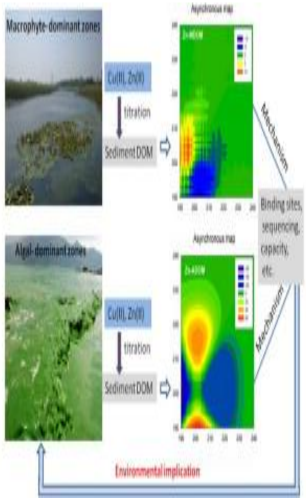
Courtesy of Ref 11

DOM in lake sediments: Xu et al. [86] observed that 2D-COS was a better approach than PARAFAC analysis for exploring HM–DOM interactions (chart 3.8).

Chart 3.8: PARAFAC for Macrophyte—dominant dissolved organic matter

Instruments	Chemometric_methods
1) Ex-Em-Fl **2)** 2D- correlation spectroscopy	☐ PARAFAC

Output	
Component	**#**
Protein	One
Humic-like	Two



Courtesy of Ref 86

Information.Bits.

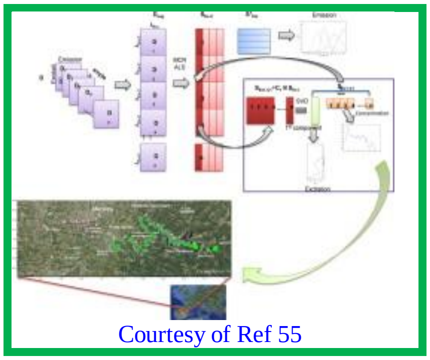
🔔 2D absorption COS showed MDOM exhibited more HM binding sites compared to ADOM

Sediment (organic materials) OM	
Macrophyte--dominant (MD) OM	**Algal-dominant (AD) OM**
193, 195, 196, 199, 201, 203, 205, 207, 208, 212, 217 nm	201, 205 nm

Inference.Bits.

🔔 2D-COS was a better approach than PARAFAC analysis for exploring HM–DOM interaction

DOM in fresh water natural systems: Zhang, Tauler et al. [55] probed into resolution of sources of DOM in natural fresh waters from Ex-Em-Fl data using MCR-ALS model. The geographical distributions of resolved contributions are mapped with MATLAB georeferenced system.



Courtesy of Ref 55

Pesticides in natural stream : Santa-Cruz and García-Reiriz [118]studied alkaline hydrolysis of pesticides with spiked water samples of natural stream with hyphenated second order instrument and advanced chemometric tools (chart 3.9).

Chart 3.9: 3way-modeling of kinetic spectroscopic data of pesticides in natural stream water

Pesticides	Unexpected compounds
carbaryl, naphthol and propoxur	tiabendazole and carbendazim

System complexity

- One of the analytes is the reaction product of another analyte

Consequence

- Linear dependency between concentration profiles

samples

Spiked water samples of a natural stream

Experiment

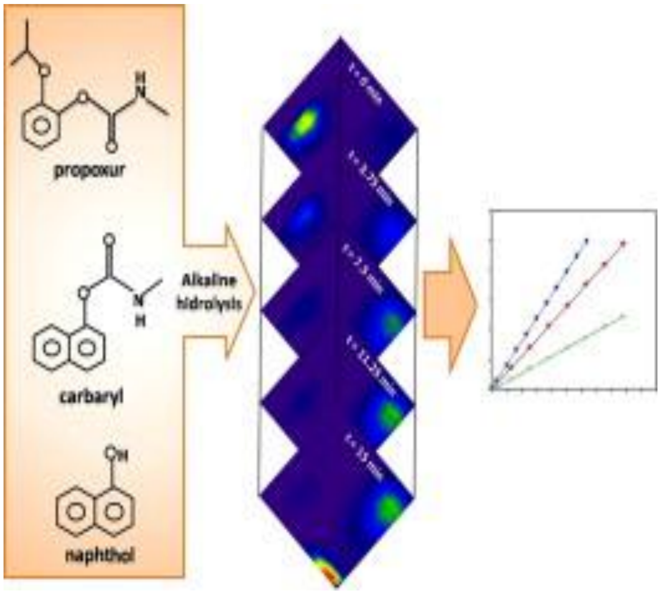
Kinetic spectroscopy	3-waydata

Chemometric_methods

PARAFAC	N-PLS	U-PLS/RTL

Chemometric_advantage

- ❑ Second order advantage to follow alkaline hydrolysis of pesticides
- ❑ U-PLS/RTL and MCR-ALS adapts to model third-order data



Courtesy of Ref 118

Information.Bits

- + U-PLS/RTL and MCR-ALS, the latter conveniently adapted to model third-order data
- + MCR-ALS, the latter conveniently adapted to model third-order data PARAFAC and MCR-ALS were the algorithms that

PARAFAC.factors

- 🕮 Two fluoroquinolones in tap water

U-PLS/RTL

- 🕮 Unfolded
- 🕮 Partial least squares with residual trilinearization

Pesticides in water: Fuentes et al. [96] analyzedimidacloprid in water samples with second order advantage of calibrating other pesticides and unexpected fluorescent compounds (chart 3.10). Imidacloprid is a neonicotinoid pesticides widely employed in the farming industry.

Chart3.10: Chemometric estimation of pesticides in water with second order calibration advantage

Instrument
1) Excitation–emission spectrophotometer with photo-induced fluorescence (PIF-EEMs)

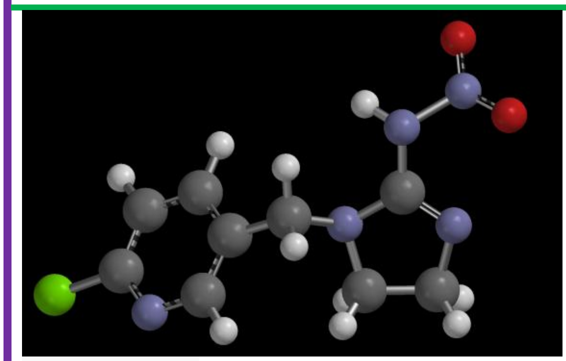
Chemometric_methods		
PARAFAC	U-PLS/RBL	RBL: residual bilinearization

Chemometric_advantage

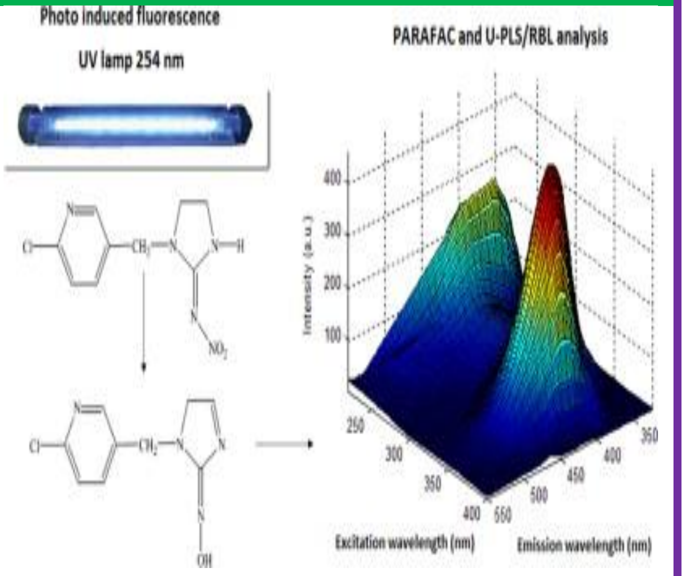
- ❑ Second order advantage of calibration and prediction in presence of unexpected photo-induced fluorescent compounds

Pesticides
Imidacloprid

Recovery by PIF-EEMs +PARAFAC or PIF-EEMs + U-PLS/RBL
101 + 10%



IUPAC name_Imidacloprid: *N*-{1-[(6-Chloro-3-pyridyl)methyl]-4,5-dihydroimidazol-2-yl}nitramide



Courtesy of Ref 96

DOM in freshwater lake: Zhang et al. [112] found photochemical behavior of DOM was related to enhanced-duration of cyanobacterial blooms in fresh water aquatic pools. The algal cells bind attached organic matter and this system is surrounded by DOM. Cyanobacterial blooms forming outer layer cause human health problems as well as ecological imbalance worldwide (chart 3.11).

Chart 3.11: 2D-correlation spectroscopic analysis of DOM

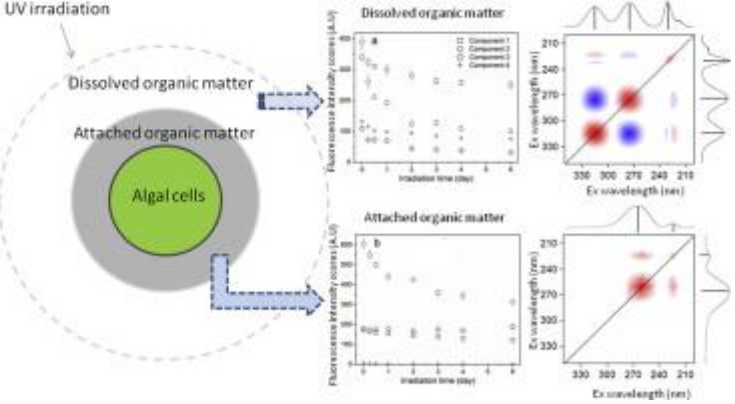
Instruments
Synchronous fluorescence (SF (2DCOS)
Chemometric_methods
PARAFAC

Information.Bits.

High photodegradation rates

DOM (k > 0.554) ; AOM (k <0.519)

PARAFAC factors for DOM and AOM	
Components	#
Tyrosine	one
Humic	one
Tryptophan-like	two



Drinking water treatment:Shutova et al. [108] reported that 867 fluorescence spectroscopic data profilesmodeled by PARAFAC for predicting DOC removal is robust compared to UV spectral studies (Chart 3.12).

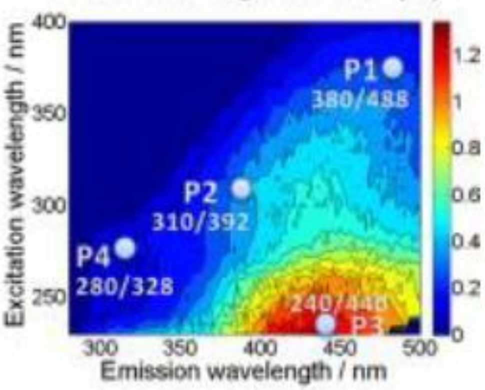
Chart 3.12: Dissolved organic carbon prediction by PARAFAC

Samples

- ❖ Drinking water treatment plants
- ❖ Various water sources
- ❖ Subtropical to temperate climate locations

Information.Bits.

🔔 4-PARAFAC factors	🕮 OM concentration 🕮 treatability



Courtesy of Ref 108

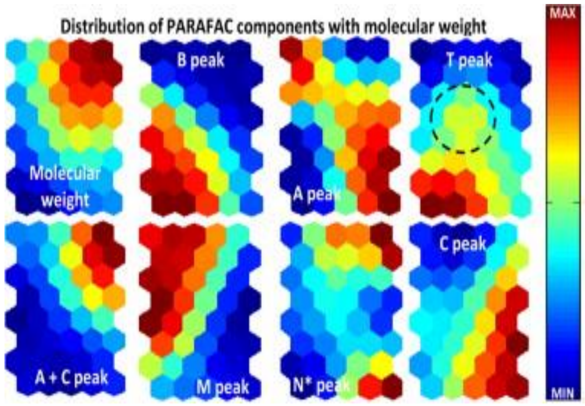
Water treatment processes: The correlations are reported between dissolved organic matter and water treatment processes by a combination of PARAFAC and SOM models for 3way fluorescence data (Chart 3.13).

Chart 3.13 : SOM and PARAFAC for DOM

Samples

- 9 Allochthonous DOM sources
  - ❖ Leaf leachates
  - ❖ Grass
  - ❖ Headwaters

Fluorescence peaks	**Corresponds To**	**Molecular weights**	
B and T	Protein/ polyphenol-like	0.5–1 kDa	highest
A + C	humic-like	>1 kDa	highest



Courtesy of Ref 88

PC#	**Correlates with**	**p <**	**%Variance explained**			**Correlates with**
PC1	Sample MW ($R^2 = 0.63$)	0.05	42% (in fluorescence properties)		PC1	○ Leachates of fresher leaves ○ Humified leaves ○ Headwaters
PC2	☞ Deciduous leachates/headwaters ☞ Coniferous leachates/headwaters.	---	27		Inference bits Structure of aged DOM arises through supramolecular assembly	
PC3	MW	0.005	13			

Tap water: Alcaráz et al. [6] utilized the second-order advantage of PARAFAC model of ex-em-fl spectra of water samples containing uncalibrated interferent substances (chart 3.14).

Chart 3.14: Second order advantage in modeling tap water pollutants with hyphenated instruments and PARAFAC/ U-PLS/RTL methods

Instruments

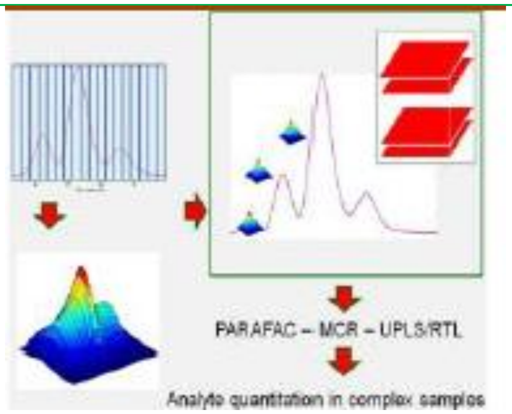
2) Fast liquid chromatographic HPLC
3) Ex-Em-fl

Chemometric_methods

PARAFAC	MCR-ALS	U-PLS/RTL

Chemometric_advantage

- ❑ Second order advantage for detection/ quantitation of /uncalibrated interferents
- ❑ U-PLS/RTL and MCR-ALS adapts to model third-order data



Courtesy of Ref 6

PARAFAC.factors

📖 Two fluoroquinolones in tap water

Component	**Rel ErrPred %**	**Analytical figures of merit**
Ofloxacin	9.9	Reduction of 70% in LODs from second- to third-order data
Ciprofloxacin	14.0%	

Information.Bits.

- \+ U-PLS/RTL and MCR-ALS, the latter conveniently adapted to model third-order data
- \+ MCR-ALS, the latter conveniently adapted to model third-order data PARAFAC and MCR-ALS were the algorithms that

MCR-ALS	\+ Better exploits second-order advantage when no peak time shifts occurred among samples

If	Temporal shifts in data
Then	Quadrilinearity is perturbed
If	Data does not adhere Quadrilinearity
Then	MCR-ALS furnished acceptable/better results

U-PLS/RTL

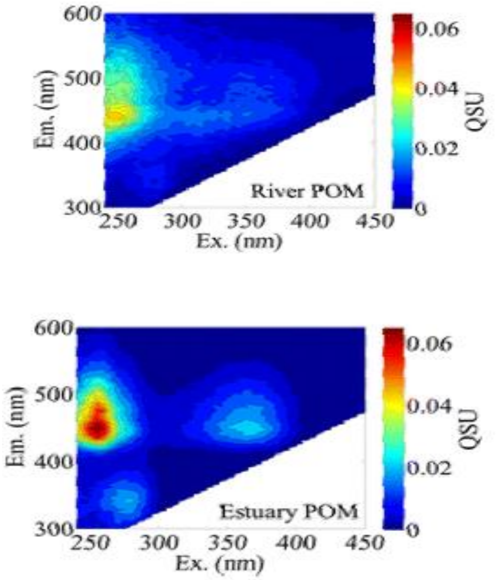
📖 Unfolded

📖 PLS withresidual trilinearization

Estuarine water quality:Osburn et al. [62] reported water quality of Neuse River Estuary (NRE) with PARAFAC model of excitation-emission fluorescence spectra of DOMs (**Chart 3.15).**

Chart 3.15:Neuse River Estuary water quality model

▲ **Samples**	❖ Base-extracted organic matter (OM) ❖ particulate _OM dissolved_OM)
▲ Location	❖ Neuse River Estuary (NRE), North Carolina,
▲ Time period	❖ before and after passage of Hurricane Irene in August 2011



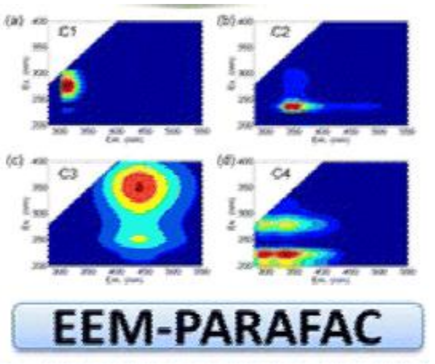
Courtesy of Ref 62

Wastewater:Li et al. [119] found that chromatographic studies are better than fluorescence spectroscopy in discriminating DOM in municipal wastewater (Chart3.16) samples.

Chart 3.16: Comparative study of HPLC/HPSEC and EEM-PARAFAC for detection of humic substances in municipal waste water samples

Limitation of EEM_PARAFAC

– EEM-PARAFAC model could not reflect the variety of DOM species with similar fluorescence but different physicochemical properties



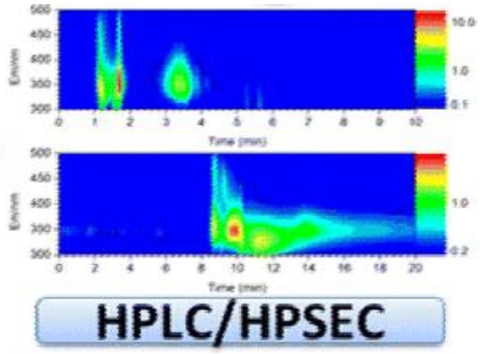
✓ The chromatography results showed that the protein-like species were variable among different municipal wastewater treatment plants, some of which are in combination with humic-like species

Instruments

1) HPLC

2) Size exclusion chromatography

3) Multi-excitation/emission fluorescence scan



Courtesy of Ref 119

WWTP: Wang et al. [107] found from PARAFAC modeling thataromatic proteins are transformed into soluble microbial byproduct-like material when microorganisms were subjected to Cu (II) stress. The phosphorous removal efficiency deteriorated in presence of excess (3 mg/L) of divalent copper. The role of EPS (extracellular polymeric substances) in P removal is studied with 3D- excitation–emission fluorescence spectroscopic data modelled with PARAFAC.

Coke wastewater (CWW) treatment plant: Ou et al. [56] made use of fluorescent characteristic variations in the study of contaminant removal efficiency in anoxic and aerobic processes(chart 3.17).

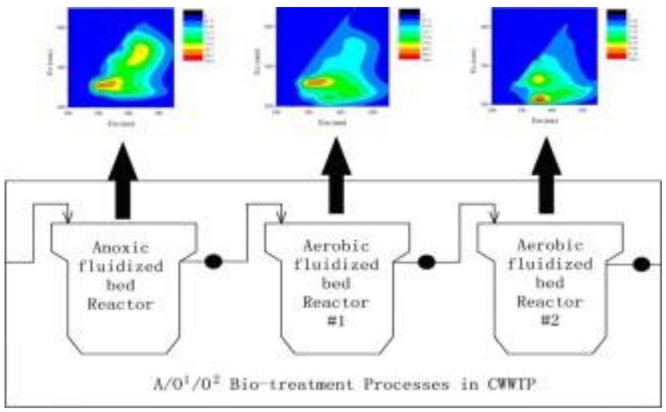
Chart 3.17: PARAFAC for coke waste water treatment process

PARAFAC # components	**Substances**
C1 and C3	2 Humic-like
C2	1 Protein-like

Correlation		**r**	**P<**		**Correlation**		**r**	**P<**
C1	COD	0.782	0.01		C1	phenols	0.796	0.01
C2	COD	0.921	0.01		C2	phenols	0.914	0.01

Inference

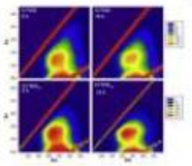
C1 and C2 might be associated with the predominating aromatic contaminants in CWW.



Courtesy of Ref 56

Waste Derived bio-based substances: Avetta et al. [37] proposed athree-component PARAFAC model for soluble bio-based (humic and fulvic) substances derived from waste.

Dissolved organic carbon (DOC) in Aquaculture industry: Hambly et al. [106] used dissolved organic carbon (DOC) analysis by EEM-PARAFAC to characterise and understand the accumulated dissolved organic matter (DOM) in mature recirculating aquaculture systems (RASs) stocked with rainbow trout and operated at steady state at four feed loadings (chart 3.18).



Courtesy of Ref 37

Chart 3.18:Dissolved organic carbon (DOC) analysis in recirculating aquaculture systems (RASs)

Objective

⊥ Future RAS management strategies

Chemometrics.Methods

ॐ PARAFAC ▶ EEM toolbox

Inf.Bits.

▶ Positive correlations
  - ▶ Fluorescence components
  - ▶ DOC

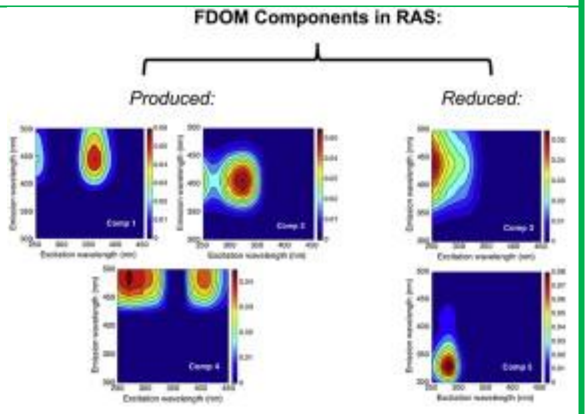
PARAFAC factors

▶ Five components

○ Three sources
  - Feed
  - Influent tap water
  - Groundwater

○ Fish related processes

○ Water treatment system



Courtesy of Ref 106

Literature.evidence

ॐ Previously identified four components in
ॐ Fresh water
ॐ Coastal marine water
ॐ Wetlands
ॐ Drinking water

Aluminum-based coagulants drinking water treatment plant

Inf.Bits.

- ▶ Positive correlations
  - ▶ Fluorescence components
  - ▶ DOC

Drinking water treatment plant: Sanchezet al. [60] reported along-Term composition of DOM in a two aluminum-based coagulants full scale drinking water treatment plant from 1000 water samples collected before and after parallel coagulation treatment basins (chart 3.19). Sanchez et al. [109]identifiedtwo humic-like and two protein like substances in raw and treated Northeast Ohio water from time series data for 32 months (chart 3.19).

Chart 3.19: PARAFAC model in drinking water treatment process

Chemometrics.Methods

🔔 PARAFAC

○ Water Samples ○ Raw ○ Treated	▶ 680
○ Duration	▶ 32 months
○ Location	▶ Northeast Ohio

Inf.Bits.

- ▶ Positive correlations
  - ▶ Fluorescence components
  - ▶ DOC

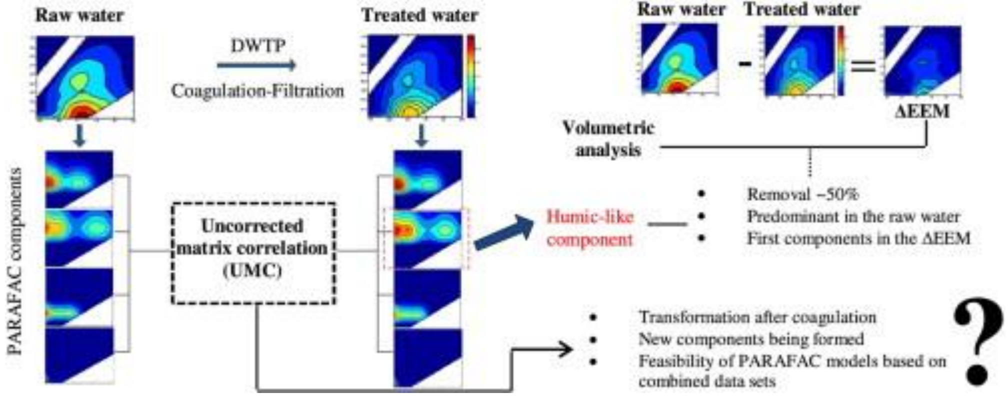
PARAFAC factors

▶ Four components	○ Two humic-like ○ Two with protein nature

+ PARAFAC on combined raw and treated water to track DOM in a DWTP

Inf.Bits.

🔔 Introduced coagulation assessment using ΔEEMs-PARAFAC and volumetric analysis



Courtesy of Ref 60,109

Aerosols: Matos et al.investigated Water-soluble and Alkaline-soluble Organic Matter in aerosols in different seasons by PARAFAC-ALS models (chart 3.20).

Chart 3.20: PARAFAC model for soluble organic matter in aerosols

Chemometrics.Methods

🔔 PARAFAC

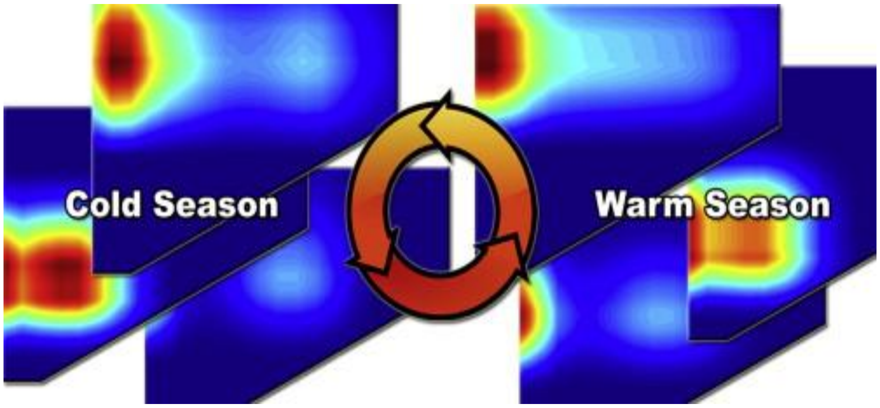
📖 Aerosol Samples

- ○ Water-/ Alkaline-soluble Organic Matter
- ○ Seasons
  - ▶ Cold
  - ▶ Warm

Inf.Bits.

▶ fluorescence map

▶ Different trends with seasonal change



PAHs in coal-tar: Mooreet et al.[17]focused the attention on PAHs in coal tar samples by monitoring spectro-fluorimeter data (chart 3.21a).

Chart 3.21a: PAHs in coal tar by Ex-Em-Fl tensorial data modeling

📖 Coal-tar Samples

▶ Seven PAHs

Chemometrics.Methods

🔔 PARAFAC

cryogenics + fiber-optic probe

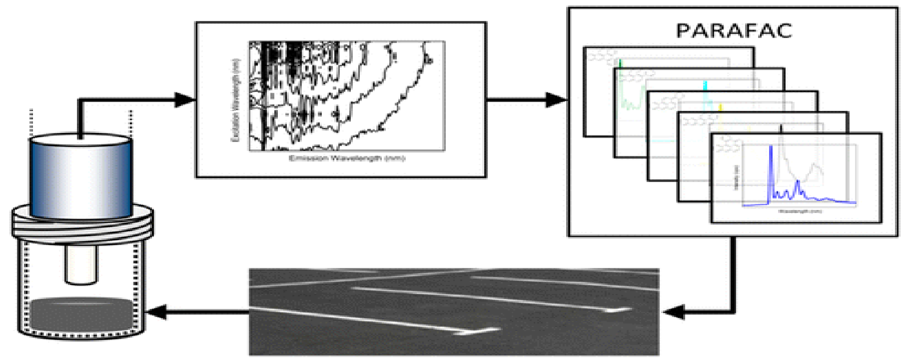
🔔 EEMs are recorded at 4.2 K

Inf.Bits.

🔔 Environmentally friendly approach for routine analysis of large number of coal-tar samples

PAHs

1) Benz[*a*]anthracene
2) Benzo[*k*]-fluoranthene
3) Benzo[*b*]fluoranthene
4) Benzo[*a*]pyrene
5) Chrysene
6) dibenz[*ah*]anthracene
7) Indeno[*123-cd*]pyrene

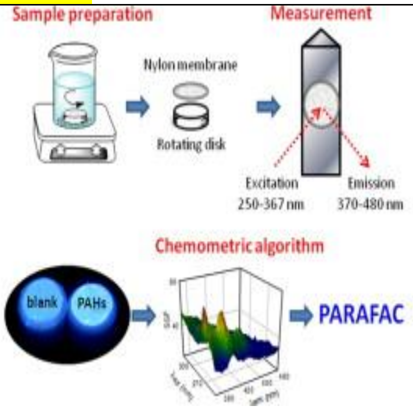


Courtesy of Ref 17

PAHs in presence of interferences: Cañas et al. [3] reported detection of 20–100 ng $L^{-1}$of PAHs (benzo[a]pyrene, dibenz[a,h]anthracene, benz[a]anthracene, and chrysene) in presence of interferences with 5–7% of relative errors of prediction (Chart 3.21b).

Chart 3.21b: PARAFAC for PAHs in heavy fuel oil analysis

1) GC × GC-TOFMS	▶ APEQ (automated peak extraction and quantification) ▶ PARAFAC



Courtesy of Ref 3

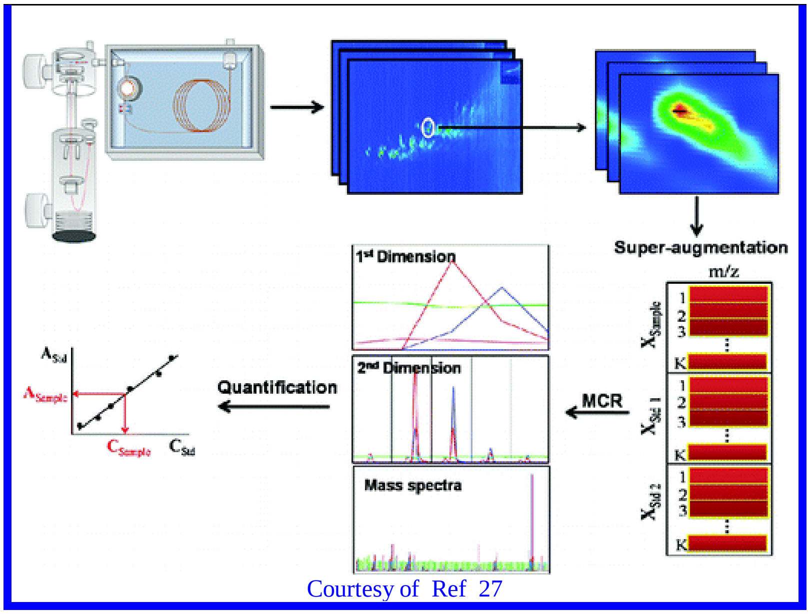
Inf.Bits.

+ Green method

PAHs in heavy Fuel Oil: Parastar,Tauleret al. [27] resolved pure component elution profiles in the two chromatographic dimensions as well as their pure mass spectra of PAHs in Heavy Fuel Oil using MCR-ALS(Chart 3.22). The relative errors in estimated samples are <6%.

Chart 3.22: PARAFAC for PAHs in heavy fuel oil analysis

2) GC × GC-TOFMS	▶ APEQ (automated peak extraction and quantification) ▶ PARAFAC

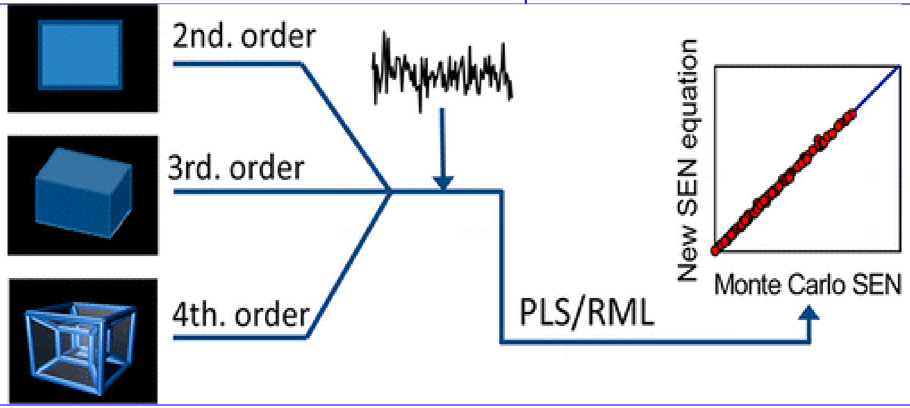


Courtesy of Ref 27

Figure of merit in m-way calibration: The expressions in the closed-form for analyte sensitivities in one-, two-, and three way data (vectors, matrices, and three-dimensional arrays, respectively) built with data for a group of samplesfor calibration are available. Allegrini and Olivieri **[**21**]** put forward a new expression for figure of merit for the set multivariate calibration algorithms (MultiCalibAlg) based on PLSR combined with residual multilinearization (chart3.23). An extensive study with Monte Carlo noise addition simulations for a second-, third-, and fourth-order data showed a decrease in average prediction error with increase in the order of data.

Chart 3.23: Figures of merit for different orders of instrumental data

Second-order advantage	**Figures of merit**
○ Number of analytes ○ Interfering agents ○ Different degrees of overlapping in component profiles ○ Different numbers of instrumental data modes per sample	⊥ Analytical sensitivity ⊥ Limit of detection ⊥ Limit of quantitation ⊥ Uncertainty in predicted concentration

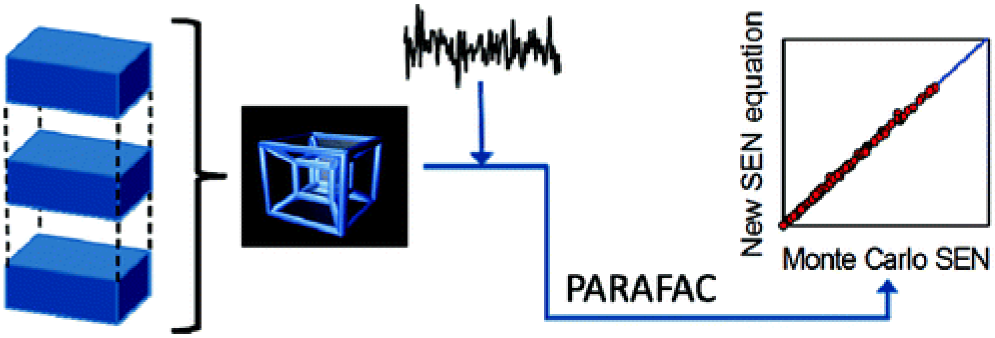


Courtesy of Ref 21

Olivieri and Faber[22] derived equations for sensitivities (Alg. 3.1)of four-way data calibration using the quadrilinear PARAFAC model.

Alg. 3.1: Sensitivity of 4-way calibration data with quadrilinear PARAFAC model

Basis	Computation of the uncertainty in Jacobian matrix of fitted PARAFAC parameters
Data	Four-way tensor
Method	Four-way simulated data + Monte Carlo noise
	Widely different overlapping situations
Application	Two experimental analytical data systems three- four-way data



Graphical algorithm [Graph_alg.] Courtesy of Ref 22

Omidikia et al. [8] discussed selectivity and uniqueness characteristics of PARAFAC model (chart 3.24). The effect of selective windows of profiles on unique resolution of three-way data sets and selectivity constraint on the unique recovery of two-way data sets are reported.

Chart 3.24 chromat-2: uniqueness and selectivity characteristics of PARAFAC

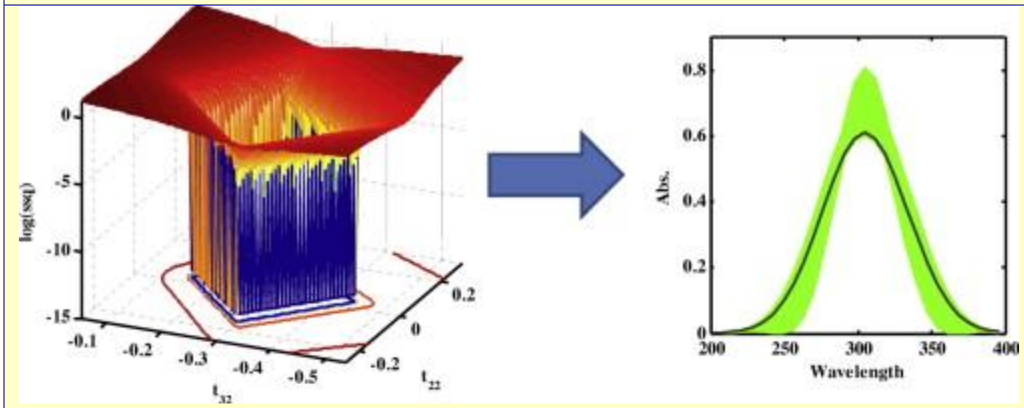
Inf.Bits.

▶ PARAFAC is a trilinear model
  - + promising exploratory tool for data analysis
  - + Unambiguous recovery of profiles is a distinguishable advantage
  - + Non-destructive methodology

– Linear dependency in of three-way data profiles
  - – Destroys trilinearity
  - – Increases ambiguity in the curve resolution.

– Linearly dependent loadings of PARAFAC
  - – Deteriorates totally or partially uniqueness

– Rank overlap ---> Rank deficiency
  - – Resolution of systems decreases



Courtesy of Ref 8

Automated peak extraction and quantification (APEQ) in chromatography

Furbo and Christensen [23] proposeda multi-stage method to extract retention times (t_R), spectra, and signal intensity from spectro-chromatographic tensorial data (Alg. 3.2).

Alg.3.2 Automated peak extraction and quantification from GC_MS with PARAFAC

Data: Full-scan chromatograms
#samples :12, In each sample: 30 PAHs

- Split chromatograms into sections.
- Detect Peaks
- Approximate elution profiles, spectra, and signal intensities
- estimate number of important peaks in each section

```
For each section
    o Model peaks and background with PARAFAC
    o Develop models with a different number of factors
End for

For each section
    o select optimal model
end for
```

Output: optimal model

+ It detected all 30 PAHs and 14 compounds not intentionally added.
+ Model is with lowest dimensionality in most cases
+ One factor describes baseline
+ extracted the tRs, mass spectra, and relative concentrations
+ The quantification was compared to integration of extracted ion chromatograms.
+ significantly better for peaks with overlap
+ It had comparable quality for peaks with little or no overlap

Nano_biological data: Akhlaghi etal.**[24]**made multiway investigation of interaction between Fluorescence Labeled DNA Strands and Unmodified Gold Nanoparticles. PARAFACis successful in application of a multiway chemometric toolbox (chart 25). The restricted Tucker3 model is better than PARAFAC to resolve multidimensional nano-biological data.

Chart 25: Tucker-3 vs PARAFAC for nano biochemical data

- PARAFAC
- Tucker3
  - Unrestricted
  - Restricted
    - Limited rotational freedom
    - Unique results
    - Better interpretability results
    - Chemically more meaningful than PARAFAC

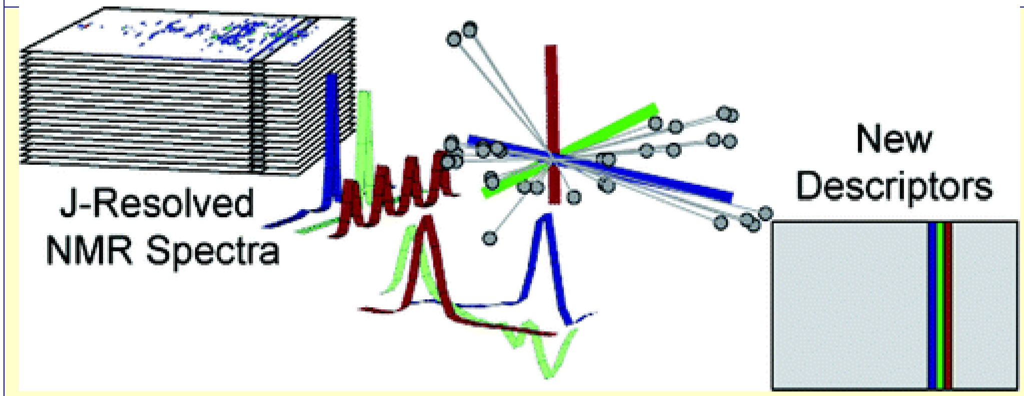
Metabolic profiling: Yilmaz et al. [26] reported mapping of correlated concentration variances of known and unknown secondary metabolites in extracts of natural products by J-resolved NMR spectra (chart 26).

Chart 26: 2D-NMR spectral modeling of metabolites with PARAFAC model

#2D-NMR spectra : 96	Compound : saffron samples extracted with (D4) methanol	Preprocessing Spectra: Alignement Splitting into 0.04 ppm wide (narrow) windows

Features. PARAFAC

- ❑ PARAFAC is a mathematical decomposition method of 3-way data tensors
  - Fits three-way experimental data to a model whose parameters
  - Concentrations / individual component spectra along the chemical shift axis
  - Spectral profiles along J-coupling axis
- ❑ PARAFAC is an advanced preprocess tool
  - Keeps J-coupling information of 2D-NMR spectra



Courtesy of Ref: 26

Pesticides estimation: Rubio et al. [97]reportedsimultaneous determination of two carbamate pesticides (carbaryl and carbendazim) and of the degradation product of carbaryl (1-naphthol) in iceberg lettuce by PARAFAC model (chart 3.27). The different dilutions of the extract from iceberg lettuce in the standard addition method forms 4th way of data tensor.

Chart3.27: 4-way data analysis with PARAFAC

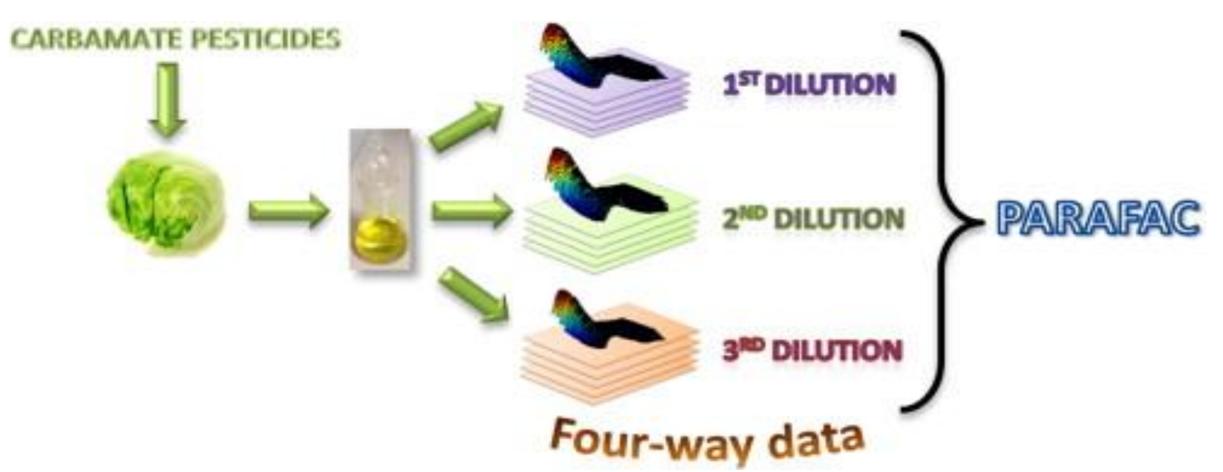
Instruments

Excitation–emission fluorescence 3-way data

Expt.Methods

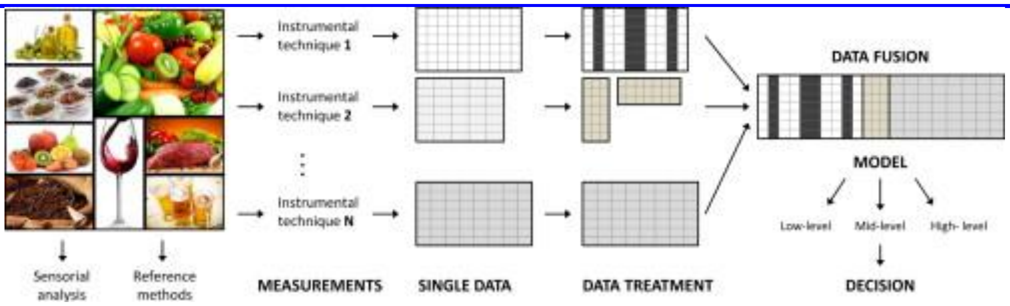
Standard addition method
Different dilutions (with 3-Way data for each dilution) of the extract from iceberg lettuce 4th-Way

Algorithm for chemometric analysis of 4-way data	
Step: 0	**Data:** 4-way data tensor
Step: 1	Four-way model ---> Identification of two fluorescent matrix constituents ---> Matrix contribution in each dilution of the extract
Step: 2	Subtraction of this contribution was from the previous signals --> subsequent three-way data analysis on tensors corresponding to each dilution



Courtesy of Ref: 97

Dietetometrics

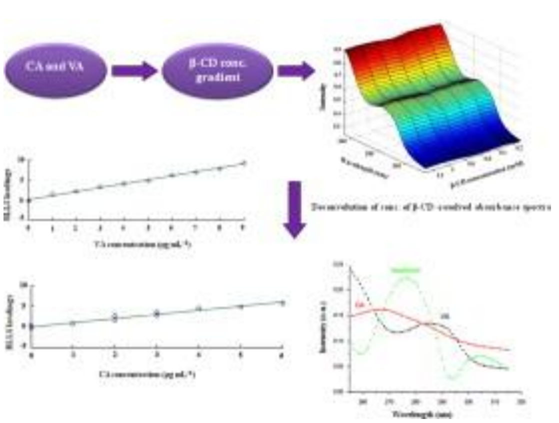


Courtesy of Ref:2

Simultaneous estimation of caffeic (CA) and vanillic (VA) acids: These are good model compounds for fruit juices samples. The severe concentration profile overlapping between CA and VA in β-cyclodextrin (CD) concentration dimension limits the application of PARAFAC model. But, differences in the spectral absorbance changes of the β-CD complexes signals of the investigated analytes, opened a new approach for second-order data generation rendering resolution of the model compounds possible. The analysis showed that BLLS/RBL functions better than PARAFAC in this instance. This approach of Khani et al. [98] is a promising tool in real samples with advantages like accuracy, sufficient spectral resolution and concentration prediction even in the presence of unknown interferents (chart 3.28). The comparison with HPLC-photodiode array detection endorses this approach.

Chart 3.28: Complexes of in β-cyclodextrin with model compounds of fruit juice samples

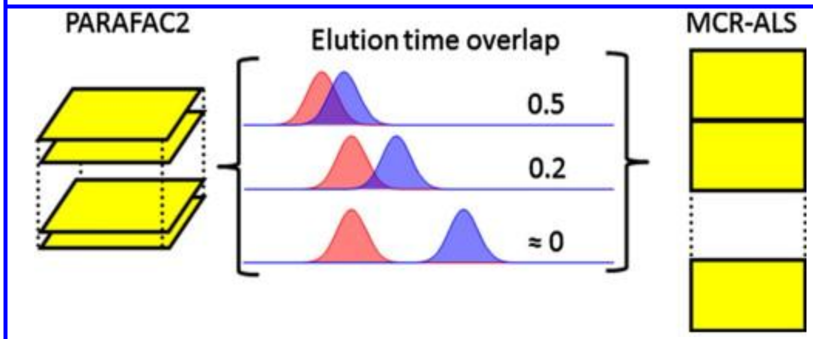
Instruments	**DATA**
Excitation–emission fluorescence	---> 3-way data
Methods	Calibration
PARAFAC	Bilinear least squares/residual bilinearization (BLLS/RBL)
Task Deconvolution of trilinear data --> spectral and concentration profiles of CA and VA as a function of β-CD concentrations	



Courtesy of Ref: 98

Fruits/ fruits juices contamination with Pesticides: Bortolato and Olivieri [4] estimated benzimidazolic and carbamate pesticides in fruits/juice and polycyclic aromatic hydrocarbons in the presence of potential interferents in water samples by hyphenated instrumental (LC-DAD,LC-fluorescence) techniques (Chart 3.29).

Chart 3.29: Pesticides in fruit juices with PARAFAC2



Data : [Simulated; Experimental]

Courtesy from Ref 4

Caution. Bits.

If Interferents are present

Then Overall results favor MCR-ALS over PARAFAC2

Elution time profiles show changes from sample to sample ⟶

Loss of trilinearity in data tensor ⟶

Perturbation/failure of factor methods

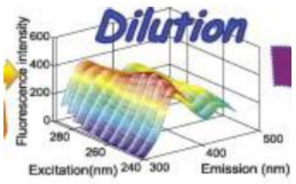
based on internal implicit tri-linearity

Lime tree flowers : Rubio et al. [7] estimated carbaryl, carbendazim and 1-naphthol in dried lime tree flowers using experimental design for calibration samples, standard addition method, hyphenated instruments and advanced chemometric technique, PARAFAC (chart 3.30).

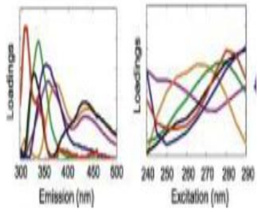
Chart 3.30: Chemometric estimation of carbaryl, carbendazim etc. in lime tree flowers

- + Adherence to trilinearity of data tensor guarantees uniqueness of PARAFAC solution
- + Factors of decomposition match up with analytes

Expt.Methods	
ॐ Standard addition method	▶ Analyte estimation
ॐ D-optimal design	▶ Calibration standard



Courtesy of Ref 7



Adulteration of artificial food colorants: Masoum et al. [91] made a quality assessment of adulterated saffron samples with PARAFAC analysis of visible spectral data at different pHs(chart 3.31). Saffron is costly ingredient with multifaceted applications in culinary spice, dyes, cooking, and also medical purposes. The scarcity as well as price is a reason for rampant adulteration and thus detection is of prime concern to avoid health hazards.

Chart 3.31: PARAFAC in detection of Food adulteration

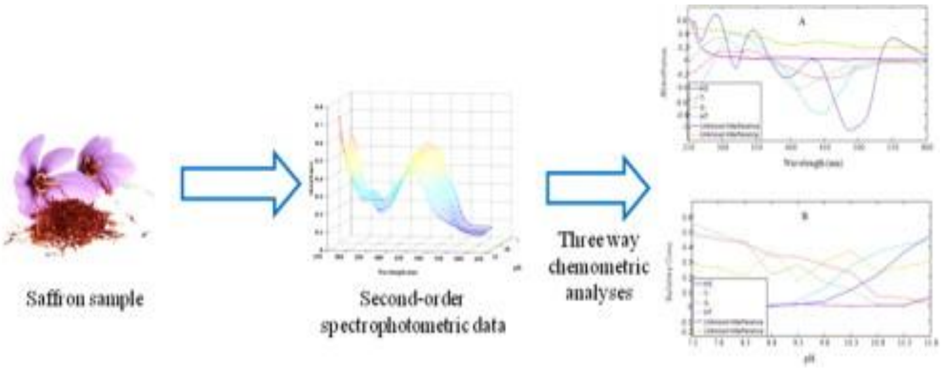
- **Analyte**
  - ⊖ Saffron
- **Adulterants**
  - 🗁 Tartrazin
  - 🗁 Sunset yellow

Instrument

- 🔔 Uv-Vis spectrophotometer
- 🔔 pH meter

Data

3way tensor



Courtesy of Ref 91

Chemometrics.Methods

☎ PARAFAC

– Rank deficiency in concentration mode → impaired system

Remedy

subtract first pH spectrum from each spectrum at each pH → Three-way variation array V

Output

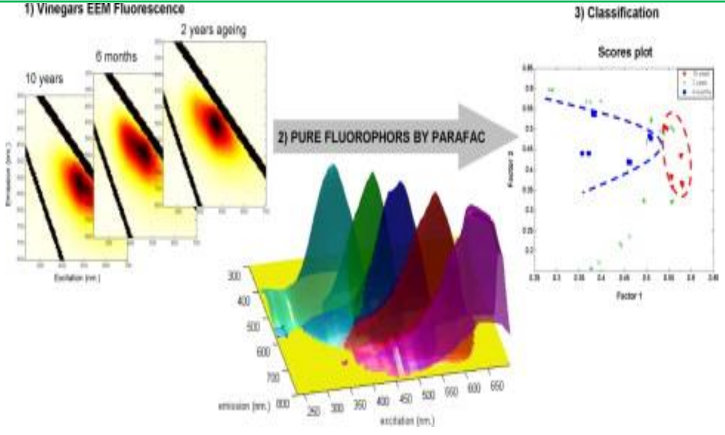
- ○ Extraction of extent reaction profile
- ○ Mixture reaction spectral profiles,
- ○ Relative concentrations of analytes

Sherry vinegar: Callejón et al. [99] employed a band of chemometric models to discriminate Sherry vinegars accordingly to their ageing (chart 3.32).

Chart 3.32: Age wise discrimination of Sherry vinegars

Instruments		
Excitation–emission fluorescence spectrometer		
Chemomet_Methods Robust classification models		
PARAFAC	**SVM**	**PLS-DA**

Quality of $$$ vinegar		Ageing time in oak barrels
Vinagre de Jerez	minimum	6 months
Reserva	at least	2 years
Gran Reserva	at least	10 years



Courtesy of Ref 99

southwestern Spain
Vinagre de Jerez
Reserva
Gran Reserva

Objectives

- Characterization of Sherry vinegar
- Classification/ discrimination of Sherry vinegars based on ageing

Goal achieved

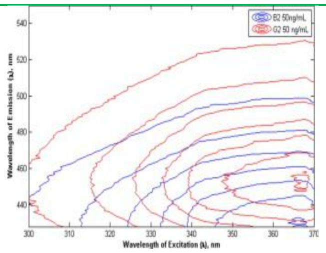
○ Detection of fraudulent samples

○ Assessment of the quality of vinegars

Luna et al. [12] reported limits of detection (LOD) and limits of quantification (LOQ) for aflatoxins B2 and G2 in peanuts using second order standard addition method and PARAFAC modeling (chart 3.33). Both naturally contaminatedand spiked complex matrix of peanuts samples showed promising results at reduced cost.

Chart 3.33: Chemometric estimation of carbaryl, carbendazim etc. in lime tree flowers

Instruments	Chemomet_Methods
excitation–emission fluorescence spectrometer	PARAFAC,
Expt.Methods	Second order Standard addition method



Courtesy of Ref 12

Compound	µg kg−1	
	LOD	LOQ
Aflatoxins B2	0.05	0.16
Aflatoxins G2	0.04	0.12

Discrimination of white wines according grape variety: Azcarate et al. [64] found U-PLS-DA and SPA–LDA have better discriminating power in assessing Argentinean white wine quality from fruits from different varieties (chart 3.34).

Chart 3.34: SPA_LDA for white wine discrimination based origin of grape furit

Instruments
4) Ex-Em-Fl spectrophotometer

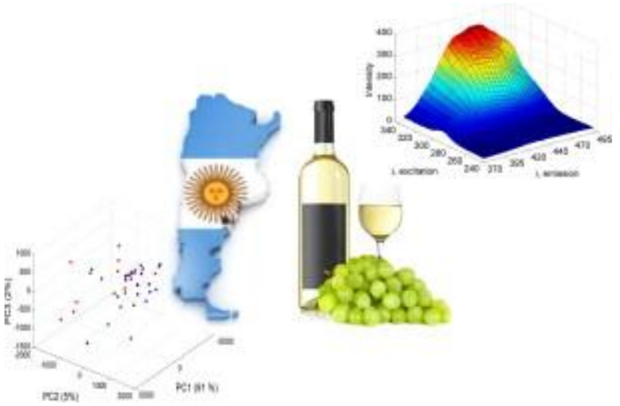
Chemometric_methods		
PARAFAC	PCA	DA
SIMCA,	N and U-PLS	SPA–LDA

	#Samples
Training	21
Test	20
	Kennard–Stone algorithm

	%Accuracy
U-PLS-DA	76
SPA–LDA	80

Statistics for model

- Mean of accuracy,
- Sensitivity
- Specificity



Courtesy of Ref64

Fluoroquinolones in urine samples: Vosough et al. [92] estimated Fluoroquinolones in urine samples by PARAFAC and U-PLS/RBL methods with acceptable accuracy (chart 3.35).

Chart 3.35: PARAFAC and U-PLS/RBL procedures for Clinical diagnostic analysis

Instruments
excitation–emission fluorescence spectrometer
Chemomet_Methods

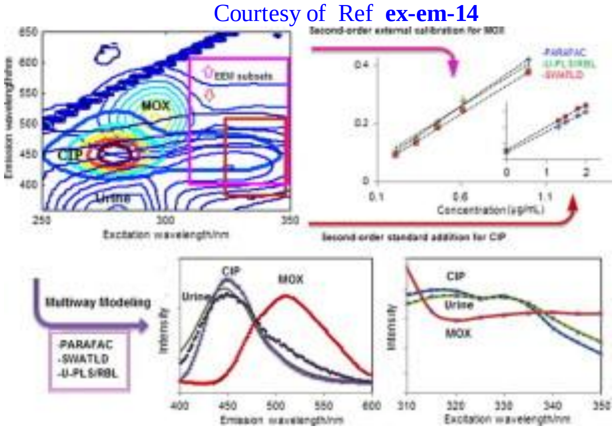
Advantages of chemometric analysis

- + Elimination of preliminary sample preparation
- + Simple experimental procedure
- + Desirable analytical performance

PARAFAC	unfolded PLS + residual bilinearization procedure (RBL)
self-weighted alternating trilinear decomposition (SWATLD)	
Utility of second-order advantage”	

Statistics
- RMSEP
- Recovery
- Elliptica joint confidence region (EJCR) plots

Courtesy of Ref **ex-em-14**



Courtesy from Ref 92

Biochemistry

The spectroscopic, kinetic and molecular dynamic studies of enzyme structure and their activities was the core of bio-physical and bio-chemical research. The advances in analysis of 3-way/4-way data from hyphenated instruments opened new vistas in age old enzyme activity in vitro as well as in vivo.

Enzyme activity: Baum et al. [10] [FT-IR-1] studied spectral profiles evolution in enzyme catalyzed reactions for preparation of pectin lyase, glucose oxidase, and a cellulose by PARAFAC using FTIR (chart 3.36).

Chart 3.36: PARAFAC for FTIR data

Instrument
FT-IR

Chemomet_Methods
PARAFAC

Experimental Procedure
Spectral evolution for reaction

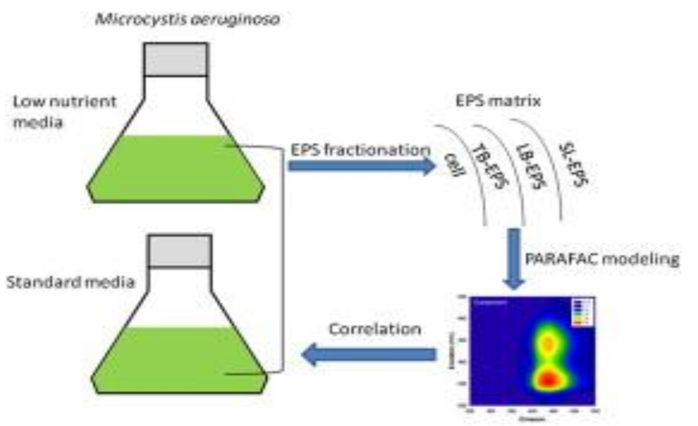
Courtesy of Ref 10

Polymers in cyanobacterium: Xu et al. [110] employed fractionation procedure and PARAFAC to explore into extracellular polymeric substances of cyanobacterium Microcystis aeruginosa (chart 3.37).

Chart 3.37: Protein like substances in cyanobacterium Microcystis aeruginosa by 3way-modeling

Instruments	Chemomet_Methods
▶ Excitation–emission fluorescence spectrometer	▶ PARAFAC

Component	Ex	Em	substances
#1	220	340	Protein-like
#2	280	340	
#3	200 220 270	296	
#4	250 340	438	Humic-like

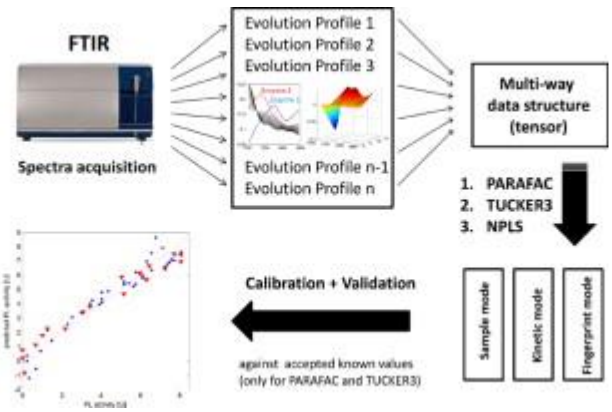


Courtesy of Ref 110

Enzyme activity in plants: Baum et al. [9] assessed simultaneous enzyme activities of pectin lyase and pectin methyl esterase using multi-way chemometric methods (chart 3.38).

Chart 3.38: Enzyme activity of pectin lyase and pectin methyl esterase

Instrument	
FT-IR	
Chemomet_Methods	
▶ PARAFAC	▶ N-PLS
▶ TUCKER3	



Courtesy of Ref 9

Medical diagnosis

Magneto encephalograms in Alzheimer's disease: Acar, Bro et al. [120] performed PARAFAC analysis of resting state magnetoencephalograms of patients with Alzheimer's disease (chart 3.39).

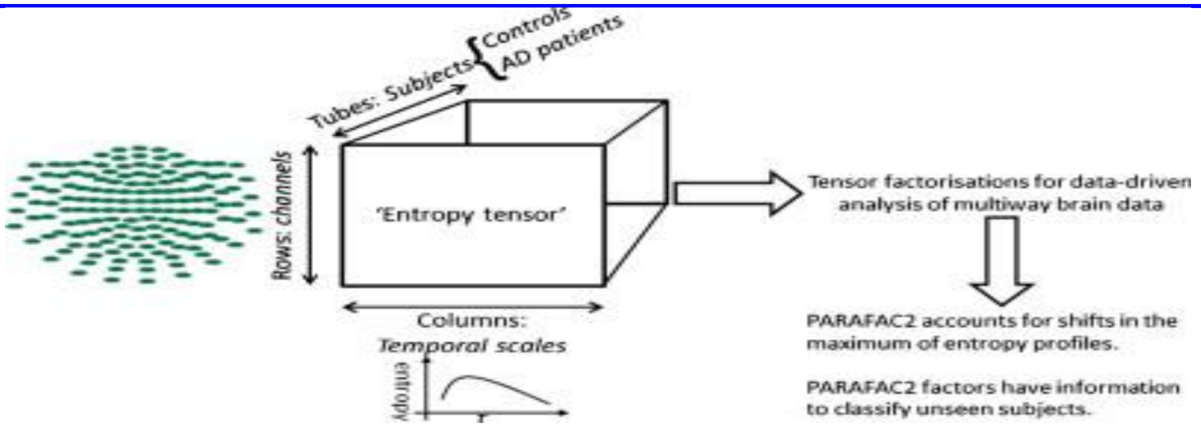
Chart 3.39: PARAFAC analysis of Magneto encephalograms

Instrument	
Magnetoencephalogram	
Chemomet_Methods	
ARAFAC	PARAFAC2

Subjects	#
Alzheimer patients	36
control subjects	26

Brain signal features

▶ Entropy tensor between temporal scales and MEG channels



Courtesy of Ref 120

Pharmaceutical formulation: Hegazy et al.[93] reported estimation of Linezolid by chemometric procedure without separation of its degradation products (chart 3.40).

Chart 3.40: pKa values of Linezolid and its degradation products

Instrument	
Spectrophotometer	HPLC

Chemomet_Methods	
PARAFAC	PARAFAC2

	RMSEP
PCR,	0.058,
PLS,	0.026,
PARAFAC	0.101
N-PLS	0.026

Experimental design (Multilevel multifactor)	
	Conc range μg mL^{-1},
LIN	12–18
ALK	2.4–3.6
OXD	1.2–1.8
Zero order Response	220-320 nm

Inf.Bits.

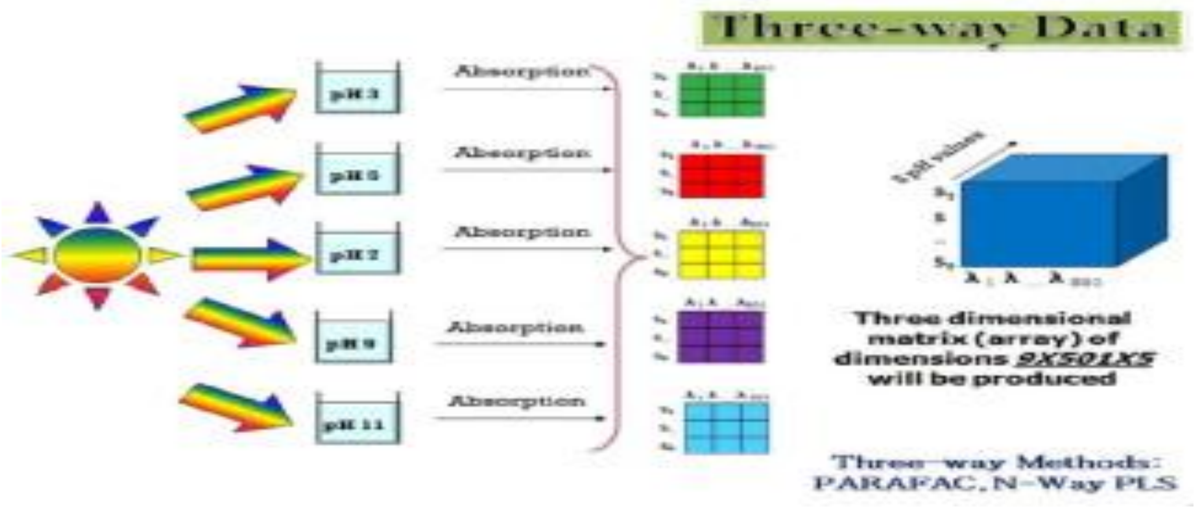
♎ The results obtained were statistically compared to that of a reported HPLC method, and there was no significant difference between the proposed methods and the reported method regarding both accuracy and precision

pKa of Linezolid and its products

Instrument	
Spectrophotometer	pH
Chemomet_Methods	PARAFAC
Task	pKa

Compound	pK_a
LIN,	5.70
ALK	8.90
OXD	6.15

OX : oxidative (OXD) degradation products of **Linezolid**
ALK: Product in presence of alkaline solution



Courtesy of Ref 93

Chemical Kinetics by chemometric models of 5-way data: Qing et al. [1] made a rigorous kinetic investigation of hydrolysis of naptalam (NAP) (chart 3.41).

Chart 3.41: 5-way data in calibration and kinetic studies

Instrument	
HPLC-DAD	pH

5-way tensor

4-way instrumental

1) Elution time **2)** Spectra
3) kinetic time **4)** pH

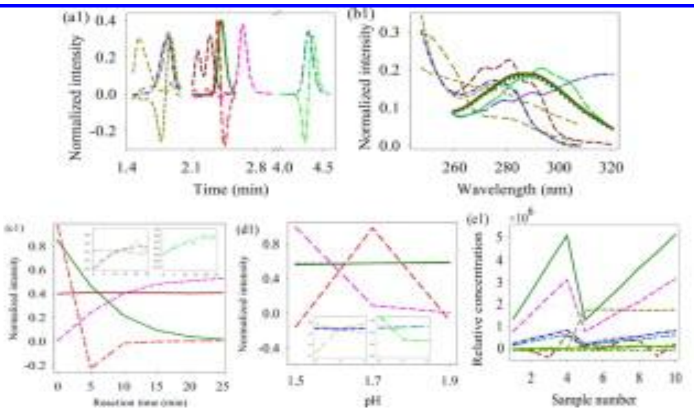
5) Response

Chemomet_Methods	
Factor	PARAFAC
Calibration	4th Order AQQLD
Basis of alternating quinquelinear decomposition	pseudo-fully stretched matrix forms of the quinquelinear model

Inf.Bits.

AQQLD

- Faster convergence rate
- Insensitive to the excess component number adopted in model
- Resolution by Chemomet_method --> Elution time, spectral, kinetic time and pH profiles
    - Good agreement with experimental observations



Courtesy of Ref 1

3-way (chromatography) data: Parastar and Akvan [5] compared chemometric algorithms for analysis of chromatographic 3-way data (chart3.42) of course exploiting second order advantage.

Chart 3.42: Analysis of 3-way chromatographic data with chemometric methods

Instrument

1) HPLC-DAD

Chemometric_method

1) MCR-COW-PARAFAC
2) PARAFAC,
3) COW-PARAFAC
4) MCR-ALS
5) MCR-COW-MCR

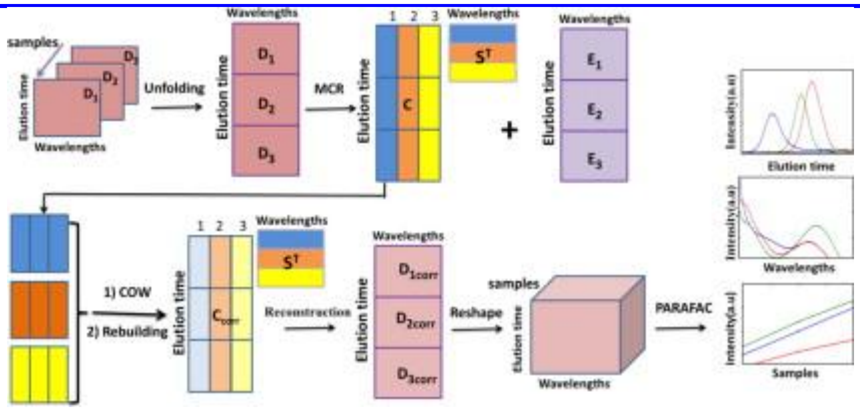
Statistics

Lack of fit (LOF)
Relative error (RE)
Spectral correlation coefficients

Inf.Bits.

MCR-COW

+ Correct elution time shifts are obtained for completely overlapped chromatographic data by coeluted interferences
+ PARAFAC analysis of aligned chromatographic data
+ Unique decomposition of overlapped chromatographic peaks even in presence of interferences
+ MCR-COW-PARAFAC compared with PARAFAC, COW-PARAFAC, multivariate curve resolution-alternating least squares (MCR-ALS), and MCR-COW-MCR
+ MCR-COW-PARAFAC is a better model in terms of statistics



Courtesy of Ref 5

Outer sphere and inner sphere complexes of Eu^{3+}: Ishida et al. [84] studied the complexes of Eu^{3+} with kaolinite with Time-resolved laser fluorescence spectral (TRLFS) data with PARAFAC model (chart3.43). The adsorption of metal ion is on ligand and gibbsite, a reference material.

Chart 3.43: Complexes of Eu3+ with kaolinite

Instrument

1) Time-resolved laser fluorescence spectrometer

Inf.Bits.

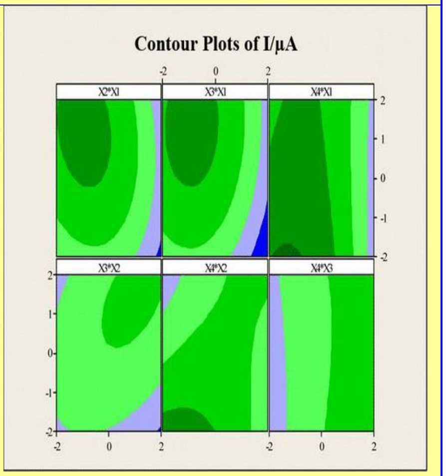
▶ Outer-sphere complex has rapid fluorescence decay compared to Eu3+ aquo ion
    ▶ **Reason :** Energy transfer to the surface.
▶ Inner-sphere complex became dominant at relatively high pH, high salt concentration

Chemometric_method Factors — Complex A — Outer-sphere B and C — Inner-sphere C — Inner-sphere &polynuclear	and low Eu3+ concentration At relatively high Eu3+ concentrations, poly-nuclear complex of Eu^{3+} formed

Spectro-electrochemistry

Spectro-voltametry: Khoobi et al. [19]reported analytical voltammetric determination of dopamine (DA) in the presence of epinephrine (EP) with second order advantages (chart3.44).

Chart 3.44: Voltametric determination of dopamine (DA) in presence of interferent epinephrine

Instruments			
1) Voltammetry	Electrode: gold nanoparticles chemically modified carbon paste electrode (AuNPs/CPE).		 Contour Plots of I/µA Courtesy of Ref 19
3-way Data generation	Various pulse heights in differential pulse voltammetry (DPV) technique		
Expt design	Central composite rotatable design		
Models	Response surface methodology	Alternating least-squares	

González-Diéguez et al. **[25]** tested dopamine oxidation at different concentrations by measuring spectroelectrochemical responses in the UV/visible range employing screen-printed electrodes (chart3.45)spectrophotometricand electrochemical methods measure the analyte indistinguishingly.

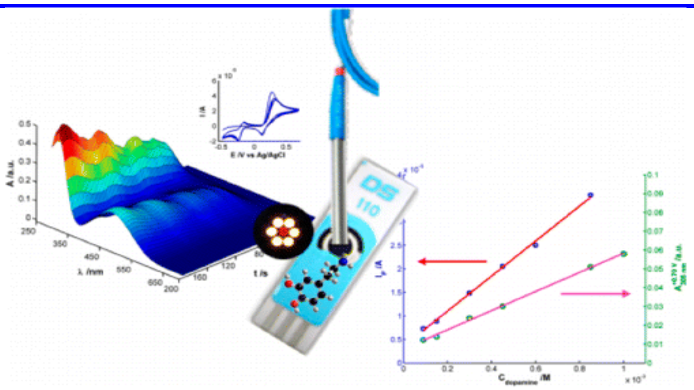
Chart 3.45: Estimation of dopamine with spectroelectrochemistry

If	[dopamine] < 10^{-3} M
Then	higher generation of dopaminochrome and its derivatives& absorption band centered at 305 nm
If	[dopamine] > 10^{-3} M
Then	dopaminoquinone is stable &maximum of absorbance, 395

Inf.Bits.

Spectroelectrochemistry

\+ Autovalidated technique



Courtesy of Ref 25

4. Function Base (fnBase) with numerical input (NI)

Many of the 3D surfaces found in chemical experiments range from simple linear functions in two variables to combination of complicated functions in several variables. The latter ones include polynomials up to third order, square root, trigonometric and exponential functions. In general the algebraic sum or product of these individual functions is considered as the response. In order to interpret the 3D surfaces and contour diagrams in any interdisciplinary task including basic chemistry, a systematic understanding of some of the mathematical functions in two variables is a prerequisite. A discussion of the functions, 3D surfaces, contour diagrams and those reduced to equivalent univariate experiments follow. Some of them represent typical mathematical functions used to validate the numerical methods or to find the limitations of the algorithms.

Three/higher dimensional response surface: Although 3D diagrams are described as fancy, they throw light on maxima, minima and saddle points. Movement along 3D-surface is to probe into smoothness of surface, breaks, troughs, valleys, flat profile, multiple local/global extremes to understand/ control/ predict phenomenon through approximate models. Hitherto, surfing in open oceans is an expert skill, but yet tense at least for spectators. The recent revolution in science is creation of ocean waves in artificial pond and thus controllable. This relieves the tension even for amateurs to surf phase wise to become expert surfers to do a great job even in rough seas. This analogy is also inspiration for scientific pursuit from best to sets of best of bests.

Contour/Iso-response diagrams: Contour diagrams are described as 2D representation of 3D surface in variable axes at iso-response values. Each iso-contour represents the lines of equal responses. It gives quantitative information of the variation of the response with the simultaneous variation of the two variables, x and y.

Plots of Resp with x and with y: These are like conventional two dimensional plots. Each curve (say Resp with x) is the change of response with x at constant y. Thus any curve represents a univariate experiment at the specified value of y. The quality and texture of the surfaces depend not only on the number of points and interpolation technique employed. So, a systematic study of the factors affecting the response surfaces follows.

- Linear Models : The most popular and widely researched and applied linear model is with one error free explanatory, one random response variables and linear (normally distributed) parameters. The advances were to tackle hetero-sedastic normal noise in y, outliers in x or y or both x and y, different scales of magnitudes of x and/or y, data structure (binary, floating point, symbolic etc.). Muti-linear models generally have more than one x- variable (i.e. second order X tensor), rarely multi-response (i.e. second order Y tensor) or both, but the parameters still are linear. Non-parametric regression does not have the constraints about the statistical distribution of parameters. A multi-coloured canvas in MATLAB is spread for surfaces/contours of various types of linear and non-linear functions.

Continuous Functions

Model with Linear variables and linear parameters: The equation of Response (Resp) for a linearfunction in x (factor 1) and y (factor 2) is in Eqn. 4.1.The coefficients can be represented in a tensor notation as lin^T, a column vector.

<table>
<tr><th colspan="3">Linearmodel</th></tr>
<tr><th>Algebraic form</th><th>Matrix notation</th><th></th></tr>
<tr>
<td>Resp= $a0 + a1 * x + a2 * y$
Eqn. 4.1
<table><tr><td>[x, y]</td><td>Linear terms or (factor 1, factor 2)</td></tr><tr><td>a0</td><td>Intercept</td></tr><tr><td>[a1, a2]</td><td>Coefficients of linear terms</td></tr></table></td>
<td>$x=\begin{bmatrix}x\\y\end{bmatrix}$; $lin=\begin{bmatrix}a1\\a2\end{bmatrix}$;
$lin^T * x = \begin{bmatrix}a1 & a2\end{bmatrix} * \begin{bmatrix}x\\y\end{bmatrix}$
$= a1 * x + a2 * y$
$fn(x) = z = \mathrm{Resp}$
$= a0 + lin^T * x$</td>
<td>Gradient
$g = \begin{bmatrix}\frac{\partial fn}{\partial x}\\ \frac{\partial fn}{\partial y}\end{bmatrix} = \begin{bmatrix}a1\\a2\end{bmatrix}$
Hessian
$H = \begin{bmatrix}\frac{\partial g}{\partial x}\\ \frac{\partial g}{\partial y}\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$</td>
</tr>
</table>

Case 1: The simplest case is fn(x) = 0, i.e. parameter vector $\left([a0, a1, a2]^T\right)$ is $[0,0,0]^T$. The plot of RESP vs x and y is a horizontal plane passing through origin and the contour diagram is a set of points with zero magnitude. A perusal of RESP vs y is a straight lines passing through zero. Similarly Resp versus x is also a set of dots with zero magnitude parallel to horizontal axis. The surface has zero value at all grid points. The figure 4.1 is for fn(x) = 1;

Case 2: Resp = a1*x

$$= 0 + a1 * x + 0 * a2 * y = [0] + [x\ y] * [a1\ 0]^T$$

Contour diagram is a set of lines parallel to Resp axis. The response on each of the line equal and indicates the value of Resp at the corresponding coordinates of x and y. The gradient is towards positive side of axis. The 3D surface is a plane inclined at an angle of $\tan^{-1}$ (a1) that is 45° when a1 =1. Plots of Resp vs y are a set of parallel lines towards y axis. When a1 is of negative sign (Fig. 4.1), the 3D surface is inclined in opposite and gradient is towards -ve side.

Case 3 : Resp= [0] + [0 a2] * [xy]T

= 0+0 *a1 *x +a2 * y = a2 * y

The contour in x, y space is a set of parallel lines with respect to x axis and the response surface is an inclined plane. The figure mimics univariate experiments at constant values of x and y, respectively. The straight lines are along the diagonal. The figure offers pictorial explanation that the response is independent of x (straight lines, parallel to Resp axis) and increases with y(linear behaviour).

When a2 is negative the inclination of figure is in opposite direction. This model represents experiments wherein factor 1 has no influence while that of factor 2 produces a proportional change in response. In such cases univariate experiments are adequate and there is no need for multivariate statistical experimental design. However, in order to arrive at the best calibration model the distribution of the points should be chosen based on D-optimal design.

Case 4: Resp = [a0] + [a1 a2] * $[xy]^T$

= a0 + a1 * x + * a2* y

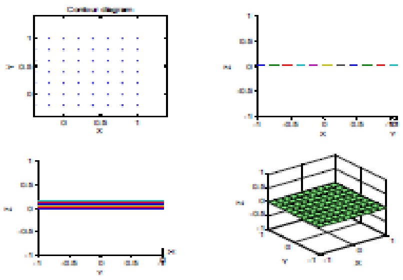
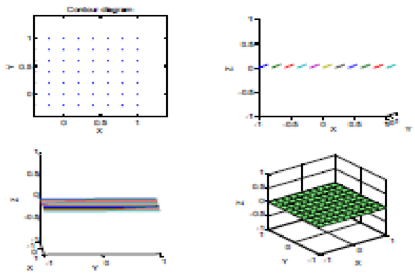
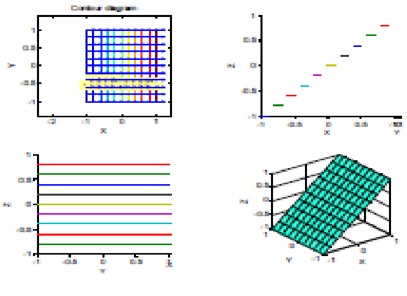
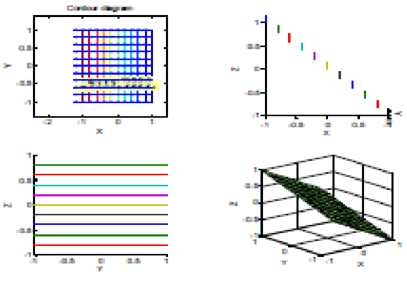
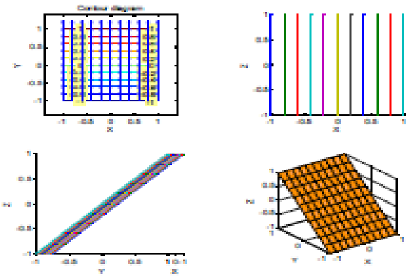
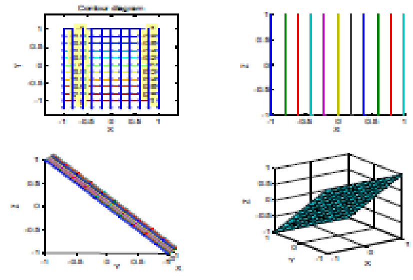
The contour diagram is a set of parallel lines, but is inclined towards diagonal. The extent of inclination depends on the correlation of x with y. The 3D surface is still a plane but is directed towards x or y space. Response vs y is at several values of x are a set of parallel straight lines. The only difference in the Figure is that x contributes to Resp in the former case. For a model with parameter vector [0 -1 -1], the figures are similar but differ in the direction of inclination. When the sign of the coefficient of x is opposite to that of y, the sign of the response depends upon the relative magnitudes of the parameters and the range of x and y. A perusal of gradient (g) for models these (vide supra) reveal that it is a constant (a) and Hessian is a zero vector. There is no minimum or maximum except at infinity and that is the reason why the line or plane continuously increases or decreases. It appears that a discussion of cases 1 to 4 is trivial. But they are instrumental to develop an in-depth knowledge of gradient, Hessian of higher order polynomials in multivariate space.

The intercept a0 becomes zero for mean centred data. The earlier practice of analysing mean entered data in statistics is as result of reduction of number of parameters by one. However, in many chemical problems chemometricians discourage this approach as the information about the origin is lost.

```
function om_xyz(range, eqn)
      disp(range)
      eval(range)
      [r,c] = size(x);
      one = ones(r,c);
      disp(eqn)
      eval(eqn)
      xyz_plot(z,x,y)
```

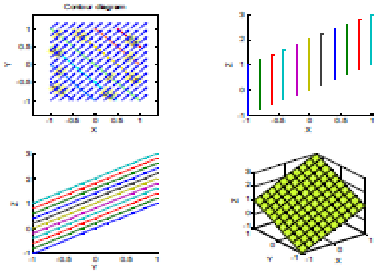
```
%
% xyz_plot.m
%
function xyz_plot(z,x,y,titlez)
if nargin <4, titlez= 'Contour diagram';end
    ztitle = 'Z';
 %
 figure
%
     subplot(221)
     [DX,DY] = gradient(z,.2,.2);
     [c,h]=contour(x,y,z);
     h = clabel(c,h);
     set(h,'BackgroundColor',[1 1 .6])
     axis equal,axis ('square'),
     xlabel('X'),ylabel('Y'),zlabel(ztitle),
     zz =0.4;
     v = axis; axis([v(1)-zz,v(2)+zz,v(3)-zz,v(4)+zz])
       hold on
      quiver(x,y,DX,DY),
      title(titlez),hold off


     subplot(222)
     zz =1;
     v = axis; axis([v(1)-zz,v(2)+zz,v(3)-zz,v(4)+zz]), surfl(x,y,z)    ,
plot3(x,y,z)

     axis equal,axis ('square'), xlabel('X'),ylabel('Y'),zlabel(ztitle),view([0,0])
```

```
     subplot(223)
     zz =1;
     v = axis; axis([v(1)-zz,v(2)+zz,v(3)-zz,v(4)+zz]), surfl(x,y,z) ,
plot3(x,y,z)
     axis equal,axis ('square'),
xlabel('X'),ylabel('Y'),zlabel(ztitle),view([90,0])
     subplot(224)
     surfl(x,y,z)

     axis equal,axis ('square'), xlabel('X'),ylabel('Y'),zlabel(ztitle)
```

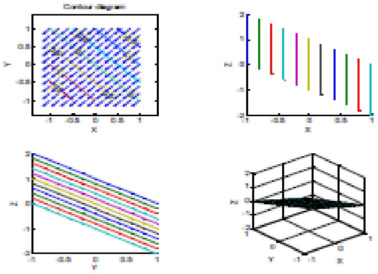
Fig. 4.1: Linear Model	**z = a0 + a(1) * x + a(2) * y**
 **Z = 0**	 **Z = 1**
 **Z = x**	 **Z = -x**
	

Z = y

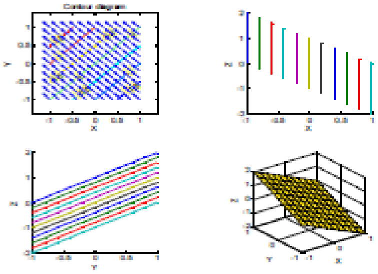
Z = -y



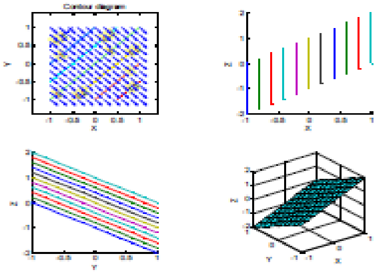
Z = 1+ x + y



Z = - x - y



Z = - x + y



Z = x - y

```
 %
%   xyz_lin.m   R S Rao   14-7-15;  5/9/94
%
function  lit_lin
    n = 99;
while n >0
    clear,clc
 a=[ 0      0      0      %1
     1      0      0      %2
     0      1      0      %3
     0     -1      0      %4
     0      0      1      %5
     0      0     -1      %6
     1      1      1      %7
     0     -1     -1      %8
     0     -1      1      %9
     0      1     -1      %10
                             ];

     eq = ' z = a0 + a(1) * x + a(2) * y  ';
     xyz_dis
         lt= -1; inc = 0.2; ut=1;
         rangexy= 'lt= -1; inc = 0.2; ut=1;[x,y] = meshgrid(lt:inc:ut,lt:inc:ut);';

        a(1) =a(n,2);a(2) = a(n,3);a0= a(n,1);
a(n,:)
        disp(eq)
        disp([ a0 a(1)  a(2)   n])
```

```
         [x,y] = meshgrid(lt:inc:ut,lt:inc:ut);
         z = a0 + a(1) * x + a(2) * y ;
         eval(eq)
           xyz_plot(z,x,y)
           next(0)
end
```

Non-linear models:

Binary cross product terms: The simplest non-linear model is obtained by the product of two linear variables viz. Resp = a3 * x *y which is hyperbolic in nature. The contours are curved in all four directions and contour diagram (Fig. 4. 2) passes through zero at $[0\ 0\ ]^T$. A non-linearity in a model is introduced whenever a non-linear function or variable is added to any of the linear models. The 3D surface non-linear depending upon the rotation about Resp axis and titling. In univariate experiments,thefigure may be misleading as one may infer that the variation of Resp with y is linear. In fact it is an artefact of reduction of equation Resp = a3 * x * y to Resp = a1 * y. The figurecan be similarly explained based on the equation Resp = a3 *x at constant but different numerical values of y. A negative value of the coefficient results in 3D surface opening upwards.

Cross productmodel		
Algebraic form	Matrix notation	
Resp= $a0 + a12 * x1 * x2$	$xcp = [x * y]$; $cp = [a12]$; $cp^T * x = [a12] * [x * y]$ $fn(x) = a0 + cp^T * xcp$	

[x*y]	Cross product of linear terms --> nonlinear
a0	Intercept
[a12]	Coefficient of cross product

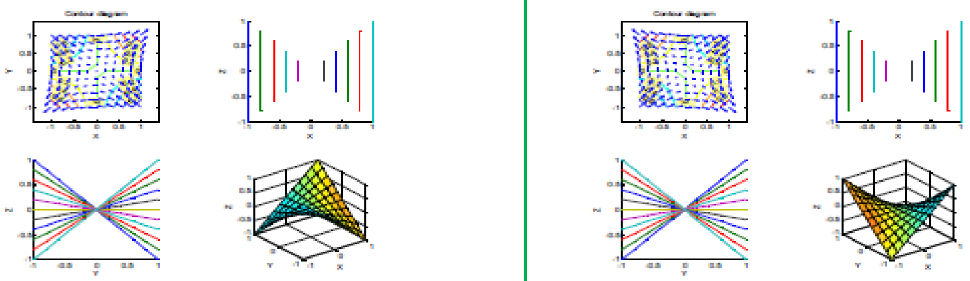


Fig. 4.2: Cross product Model **z = a(1) *x.*y + a(2);**

z = x.*y **z = - x.*y**

```
%
%      xyz_xy.m 4/10/94
%
function xyz_xy
clean
```

```
llt = -1; ult=1; inc= 0.2;
[x,y] = meshgrid(llt:inc:ult,llt:inc:ult);

           disp(['   x*y  a0 # '])
            b = [     1    0   1
                     -1     0   2
                    -0.1   0   3    ]
            nn = input(' Give   n < 1 to 3> : ');
if nn == 0, return, end
a(1) = b(nn,1); a(2) = b(nn,2);

         eq = 'z = a(1) *x.*y + a(2);'  ;
         eval(eq);
         disp(eq)
         xyz_plot(z,x,y);
```

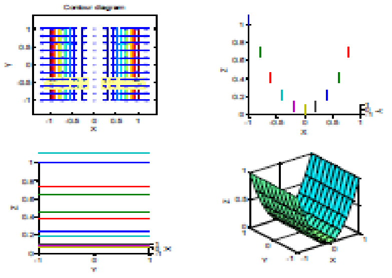
Quadratic models:

Case 1: The quadratic empirical models in one (x or y) or two variables are symmetric. The contour diagram (Fig.4.3) for Resp = a3 $*x^2$ is a set of unequidistant parallel lines and the response surface. The figure isa bowl opening upwards. Similarly the plots of response vs.y are also parallel lines with respect to y axis, when coefficient of x is negative the 3D surface opens downwards the contour diagram appears to be same but the magnitude of the response is opposite in sign.

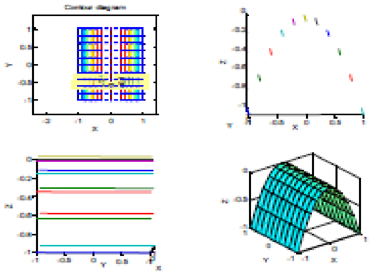
<table>
<tr><th colspan="2">Quadratic model</th><th></th></tr>
<tr><th>Algebraic form</th><th>Matrix notation</th><th></th></tr>
<tr><td>Resp= $a0 + a11 * x^2 + a22 * y^2$</td><td>$xquad = \begin{bmatrix} x^2 \\ y^2 \end{bmatrix} ; quad = \begin{bmatrix} a11 \\ a22 \end{bmatrix}$
$fn(x) = a0 + xquad^T * quad$</td><td><table>
<tr><td>$\begin{bmatrix} x^2 \\ y^2 \end{bmatrix}$</td><td>Quadratic terms --> nonlinear</td></tr>
<tr><td>a0</td><td>Intercept</td></tr>
<tr><td>$\begin{bmatrix} a11 \\ a22 \end{bmatrix}$</td><td>Coefficient of quadratic terms</td></tr>
</table></td></tr>
</table>

Fig. 4.3: Quadratic Model

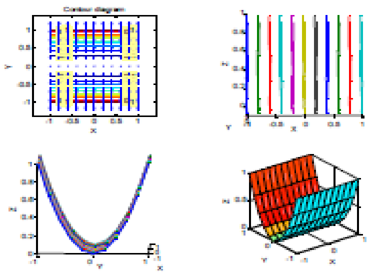
`z = a0 + a(1) * x.^2 + a(2) * y.^2;`



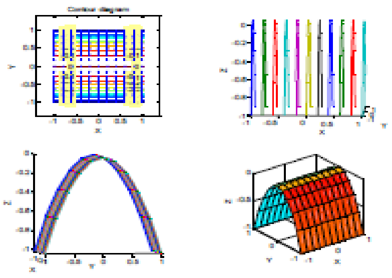
$z = x^2$



$z = - x^2$

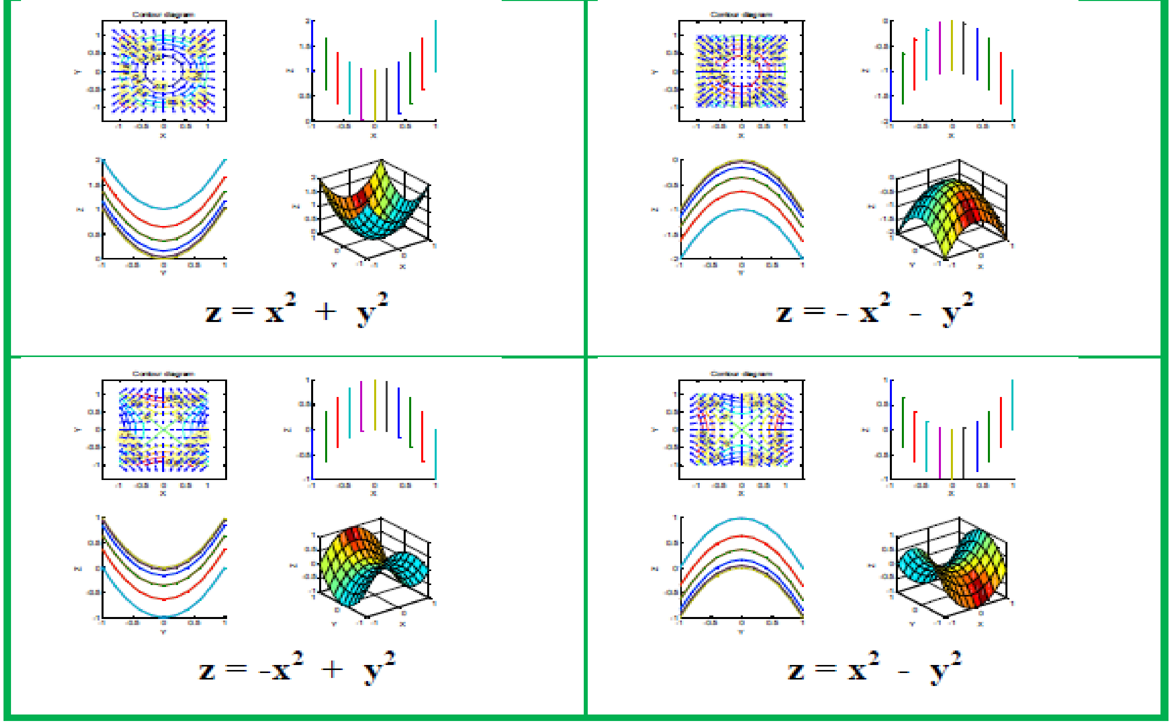


$z = y^2$



$z = - y^2$

Fig. 4.4: Quadratic models with negative signs



$z = x^2 + y^2$ $z = - x^2 - y^2$

$z = -x^2 + y^2$ $z = x^2 - y^2$

```
%   xyz_quad.m   9-7-15; 5/9/94
%
function xyz_quad
 clean
         coe = [
     0     1     0     1
     0    -1     0     2
     0     0     1     3
     0     0    -1     4
     0     1     1     5
     0    -1    -1     6
     0    -1     1     7
     0     1    -1     8
                       ]

   om_3d2d(coe);

function  om_3d2d(coe);
  n1 = -1; n2 = .2; n3= 1; n4 = -1; n5= .2; n6=1 ;
 [x,y] = meshgrid(n1:n2:n3,n4:n5:n6);
 [row,col] = size(coe);

for i = 1: row
           a(1) =coe(i,2);a(2) = coe(i,3);a0= coe(i,1);
           eq = ' z = a0 + a(1) * x.^2 + a(2) * y.^2;  ';
           disp(eq);
           disp([ a0 a(1) a(2) coe(i,4)])
           eval(eq);
           xyz_plot(z,x,y);
%
           next(0)
end
```

Case 2: The contour diagram for z= y^2 is similar to z = a3 * x^2, but parallel to x axis. The response vs. y is a set of overlapping parabolic lines and those with respect to x are parallel lines to response axis. It is as result of a fact that response does not depend upon x.

Case 3: The utmost important quadratic model in two variables is Resp = a0 + a1 +x^2 +a2 *y^2. The sign/ magnitudes of coefficients of quadratic termsrange of x and y result in simple quadratic surfaces to those with saddle points. It is also possible that some time any of the surfaces so far discussed (vide supra) may result. The significance of this model is discussed under canonical transformation of the full quadratic model.

The contour diagram forz = [x^2+ y^2] is a set a concentric circles and the response surface is a bowl opening upwards. The projection, rotation about Resp axis and tilt will give rise to apparently different shapes of 3D surface. If the range of the variable studied is limited, the profile appears to be linear or nonlinear with either increasing or decreasing slope. Extrapolation of the result based on any of the rigorous modelling strategies predicts sometimes absurd results. For example an increasing trend will be predicted as decreasing and vice versa. When both coefficients a3, a4 are negative the response surface opens downwards and response vs. y are a set of non-overlapping parabolic curves. Interesting cases arise when the coefficients a3 and a4 are of opposite sign and their magnitudes differ significantly. The heuristics developed in this and in other laboratories are incorporated in the knowledge base of canonical analysis program. For example when the parameter vector is $[0\ 0\ 0\ -1\ 1]^T$ zero crossing occurs as is evident in the knot in the response surface.

Full second order model: The literature on application of full second order model (fig. 4.5)finds a place in every sub-discipline of chemistry. The reasons for not using higher order polynomials (except in rare instances) are

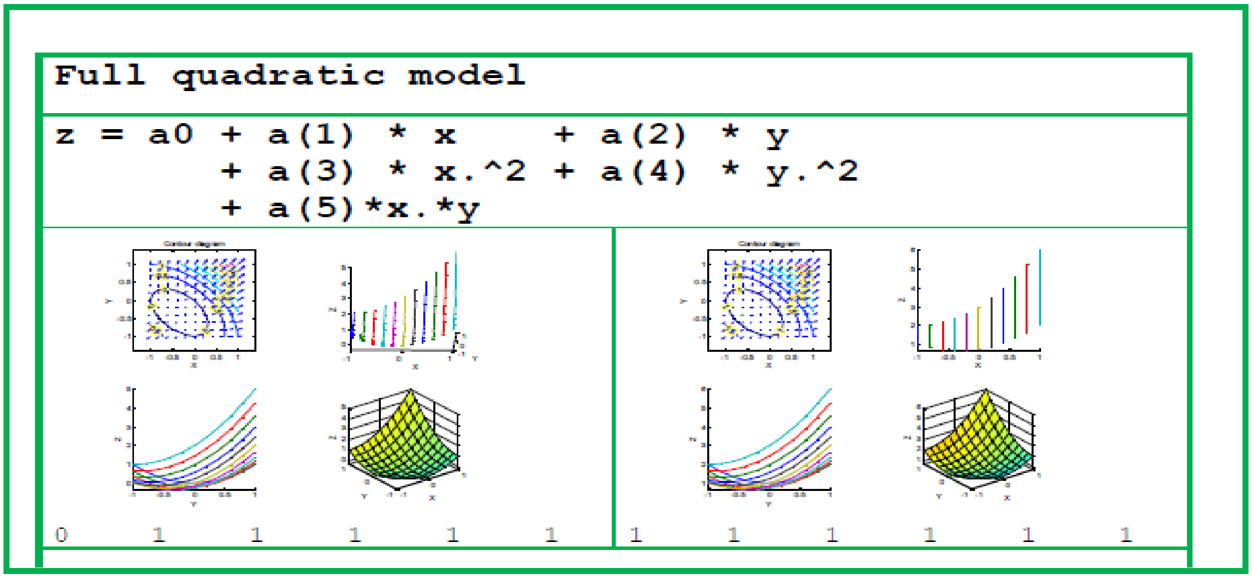
- The higher order polynomials are sensitive even to small errors in data
- The number of regression coefficients (basis functions) increase rapidly with order of the polynomial.

Full quadratic model	
Algebraic form	**Matrix notation**
$\text{Resp} = a0 + a1 * x + a2 * y + a12 * x * y + a11 * x^2 + a22 * y^2$	$\text{Re}\,sp = \begin{bmatrix} 1 & x & y & x*y & x^2 & y^2 \end{bmatrix} * \begin{bmatrix} a0 \\ a1 \\ a2 \\ a12 \\ a11 \\ a22 \end{bmatrix}$

Fig 4.5 Full quadratic model

```
%       xyz_fullquad.m    4/10/94
%
function xyz_fullquad
 clc, n = 99;
%
while n ~= 0
        clc
%    a0    x     y     xx    yy    xy Number
     a = [
```

```
     0     1     1     1     1     1     1
     1     1     1     1     1     1     2
     0     0     0     0     0     0     3
     1     0     0     0     0     0     4
     0     1     0     0     0     0     5
     0    -1     0     0     0     0     6
     0     0     1     0     0     0     7
     0     0    -1     0     0     0     8
     0     0     0     1     0     0     9
     0     0     0    -1     0     0    10
     0     0     0     0     1     0    11
     0     0     0     0    -1     0    12
     0     0     0     0     0     1    13
     0     0     0     0     0    -1    14
                              ];
                          [r,c] = size(a);

      eq = 'z = a0 + a(1) * x + a(2) * y + a(3) *x.^2 + a(4) * y.^2 + a(5)*x.*y ';
      rangexy = ' lt= -1; inc = 0.2; ut=1;[x,y] = meshgrid(lt:inc:ut,lt:inc:ut); ';
%
      xyz_dis
      disp(rangexy)
      eval(rangexy)
a(n,:)
%
      [r,c] = size(x);
      one = ones(r,c);
       a0   = a(n,1) ;
       a(1) = a(n,2);
       a(2) = a(n,3);
       a(3) = a(n,4);
       a(4) = a(n,5);
       a(5) = a(n,6);
%
      disp(eq )
      eval(eq )
%
      xyz_plot(z,x,y)
      next(0)
end% while
```



Full quadratic model

z = a0 + a(1) * x + a(2) * y
+ a(3) * x.^2 + a(4) * y.^2
+ a(5)*x.*y

0 1 1 1 1 1 | 1 1 1 1 1 1

$\frac{\partial \mathrm{Re}\,sp}{\partial x1} = a1 + a12*y + a11*2*x = 0$ $\frac{\partial \mathrm{Re}\,sp}{\partial x2} = a2 + a12*x + a22*2*y = 0$	$\begin{bmatrix} 2*a11 & a12 \\ a12 & 2*a22 \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} = -\begin{bmatrix} a1 \\ a2 \end{bmatrix}$
$\begin{bmatrix} \frac{\partial \mathrm{Re}\,sp}{\partial x} \\ \frac{\partial \mathrm{Re}\,sp}{\partial y} \end{bmatrix} = \begin{bmatrix} a1 & 2*a11 & a12 \\ a2 & a21 & 2*a22 \end{bmatrix} * \begin{bmatrix} 1 \\ x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$g = \quad b^T \quad + 2*c*x$ $H = \quad 2*c$

Canonical analysis

A perusal of chemometric literature reveals that most of the data sets in the experimental design to obtain optimum operating conditions are analysed by a full second order empirical model. Further the magnitudes of the explanatory variables are coded (-1 to 1). This has the advantage that the design matrix is orthogonal and thus MLR results in unbiased estimates of regression coefficients. The coded variables are easier for calculations without resort to computers and for general comprehension. A close examination of the estimates of the parameters with coded and uncoded X variables are different, although the corresponding response surfaces are similar. In order to automate the interpretation of full quadratic models with different magnitudes of coefficients, canonical analysis is useful.

At the stationary point, the partial derivative of Resp with respect to x and y are individually equal to zero. In canonical analysis, the stationary point (maximum) of the estimated response surface of full quadratic model is calculated by equating the partial derivatives to zero. The origin of factor space is then translated (shifted) to the stationary point. This operation removes first terms. The new factor axes are then rotated such that they overlap with the principal axes second order model. The interaction terms are absorbed during rotation. The full quadratic model now reduces to canonical form.

Canonical analysis of full quadratic model in algebraic and matrix notations	
Resp = $b11 * x\_can^2 + \mathrm{b22} * \mathrm{y}\_can^2$	Resp = $\begin{bmatrix} x\_can^2 & y\_can^2 \end{bmatrix} * \begin{bmatrix} b11 \\ \mathrm{b22} \end{bmatrix}$

Shifting of factor space to stationary point

The coefficients of linear terms are reduced to zero by translation of the factor space to location of stationary point. It is effected by adapting a coding scheme where Cx1 = s(1); Cx2 = s(2);in the coded and uncoded form have the same magnitudes. This S matrix looks as if unchanged

$$x\_trans = x - s(1)*one$$
$$y\_trans = y - s(2)*one$$

Rotation of factor space axis

The result of rotation of the new factor axis can be obtained as the solution of Eigen value problem. It is established that there exists a set of rotated factor axes such that the half diagonal terms of resulting matrix are equal to zero. In other words the origin of the translated and rotated coordinate system is at the stationary point and model does not contain linear and interaction terms.

The salient features of response surface like the presence of a ridge, saddle region, flattened region or normal parabolic profiles can be predicted from the signs and relative magnitudes of b11 and b22. The m file can_kb.m implements the KB in the form of if – then –else rules and is made an integral part of the number crunching program can.m (appendix A6). It is a tiny knowledge based system and appears to be first of its kind and is continuation of our earlier efforts in complex equilibria, neural nets, experimental design and modelling. The above equations can be written in the matrix notation as described in Alg. 4.1.

Alg. 4.1: Algorithm for canonical analysis		
Function	$fn(x) = a0 + b^T * x + x^T * c * x$ $g = b^T + 2*c*x$ $H = 2*c$	S = [2*a11 a12 a12 2*a22];
Gradient	$g = b^T + 2*c*S = 0$	f = [a1 a2]';
Solution	$S = (-b^T)*(c)^{-1}$	s = - pinv(S) * f;
Response	$z\_start = a0 + b^T * S + S^T * c * S$	
	Coordinate system is at the stationary point and model does not contain linear and interaction terms	

Interpretation of results of canonical analysis

Canonical analysis reduces all non-degenerate two factor second order polynomial models.Thus the signs and magnitudes of the coefficients of the two quadratic terms reveal all the essential features of the response surface.The results (appendix-A6), anexcerpt of exhaustive test cases, in table4.1 testifies functioning of the program.

Table 4.1: Shapes of 3D- surfaces based on signs and magnitudes of Canonical regression coefficients

+	+	≈	$z = 1.00 * x^2 + 1.00 * y^2$	Parabolic bowl opening upwards	a
-	-	≈	$z = 1.00 * x^2 + 0.90 * y^2$	Parabolic	b
-	-	<	$z = -1.0 * x^2 - 1.0 * y^2$	Flattened Parabolic opening downwards	c
-	-	<<	$z = -0.9 * x^2 - 0.02 * y^2$	Ridge	d
+	-	≈	$z = -0.7 * x^2 - 1.00 * y^2$	Saddle region	e

The knowledge base, Eigen value analysis and output along with m-files are given in appendix-6.

Confidence intervals

The uncertainty region of the response surface is calculated by the equation similar to that for a one factor least squares. The only difference is that the design matrix X contains two columns corresponding to x, y

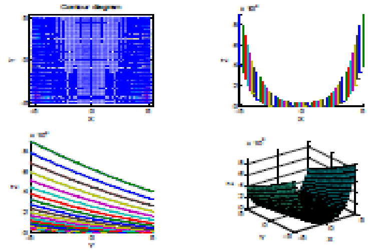
$$s^2 = s_r^2 * \left[1 + x0 * \left(x' * x \right)^{-1} * x0' \} \right]$$

Rosenbrock function

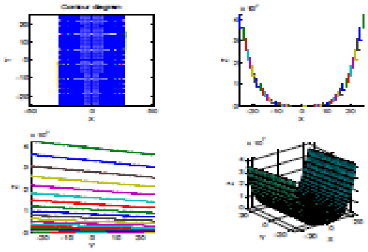
It is one of the challenging functions in optimisation as a result of deep valley (Fig. 4.6, Fig 4.7) and many of gradientalgorithms fail. The direct search methods including simplex procedure have been used. The 3D-surface for the function (Fig. 4.7) is with several optima but mostly looks like flat one with an inclination towards the diagonal of one of the factor axes.

Fig 4.6 Rosenbrock function

$$f(x,y) = (a - x)^2 + b(y - x^2)^2$$

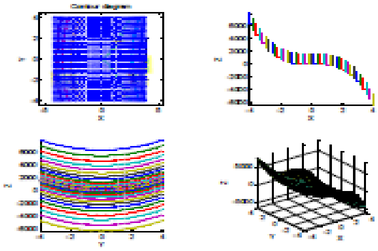


```
[x,y] = meshgrid(-5:0.2:5,-5.:0.2:5);
z = 100*(-x.^2 + y).^2 + (one - x).^2 ;
```

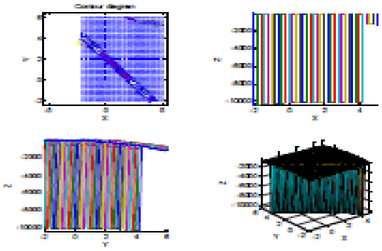


```
[x,y] = meshgrid(-25:1:25,-25.:1:25);
 z = 100.* (x.^ 2 -y).^ 2 +(one-x).^2 ;
```

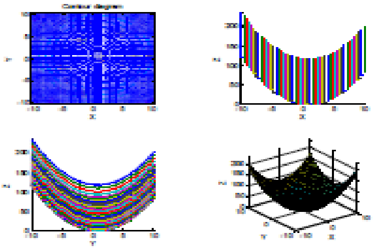
Fig. 4.7 Non-linear functions



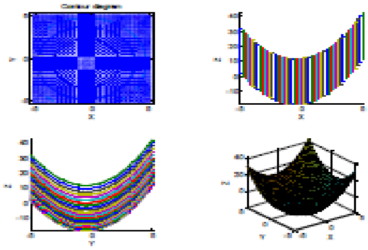
```
[x,y] = meshgrid(-4:0.2:4,-4.:0.2:4);
z = -(100.* (x.^3 - y.^2) + (one-x).^ 2 );
```



```
[x,y] = meshgrid(-2:0.2:5,-2.:0.2:6);
  z = -[(x + 5 * y).^ 2 + ((x+y - 2.04*one ).^2).^(-
1) + ((x + y -one*2.19).^2).^(-1)];
```



```
[x,y] = meshgrid(-10:0.4:10,-10.:0.4:10);
z = -[one-(x - one *.9).^ 2 - (y - .8*one).^2] ;
```



```
[x,y] = meshgrid(-5:0.2:5,-5.:0.2:5);
z = (x.^2 + y -11.) + (x + y.^2-7) ;
```

```
%
%  xyz_rosenbrock.m   5/9/94
%
function  xyz_rosenbrock
clean
%
%     xs = [1.05875 1.1848]
%
```

```
rangexy = '[x,y] = meshgrid(-5:0.2:5,-5.:0.2:5); ';
      eqn = 'z = 100*(-x.^2 + y).^2 + (one - x).^2 ;';
      om_xyz(rangexy, eqn)


%   simplex/review
%   range -25 to 25
%   xs = [-0.548 0.251]    x0 = [-1 2]
     range = '[x,y] = meshgrid(-25:1:25,-25.:1:25); ';
     eqn = ' z = 100.* (x.^ 2 -y).^ 2 +(one-x).^2 ;';
     om_xyz(range, eqn)
%
%
rangexy= '[x,y] = meshgrid(-4:0.2:4,-4.:0.2:4);';
      eqn = 'z = -(100.* (x.^3 - y.^2) + (one-x).^ 2 );';
      om_xyz(rangexy, eqn)

%
%     xs = [1.56 3.17]

range = '[x,y] = meshgrid(-2:0.2:5,-2.:0.2:6); ';
      eqn= '  z = -[(x + 5 * y).^ 2 + ((x+y - 2.04*one ).^2).^(-1) + ((x + y -
one*2.19).^2).^(-1)]; ';
      om_xyz(range, eqn)

%
%
%
range  ='[x,y] = meshgrid(-10:0.4:10,-10.:0.4:10); ';
     eqn = 'z = -[one-(x - one *.9).^ 2 - (y - .8*one).^2] ;';
     om_xyz(range, eqn)
%
%     simplex/review/

range = '[x,y] = meshgrid(-5:0.2:5,-5.:0.2:5);';
     eqn = 'z = (x.^2 + y -11.) + (x + y.^2-7) ;';
     om_xyz(range, eqn)
%
%     om_xyz.m
%
function om_xyz(range, eqn)
      disp(range)
      eval(range)
      [r,c] = size(x);
      one = ones(r,c);
      disp(eqn)
      eval(eqn)
      xyz_plot(z,x,y)
      next(0)
```

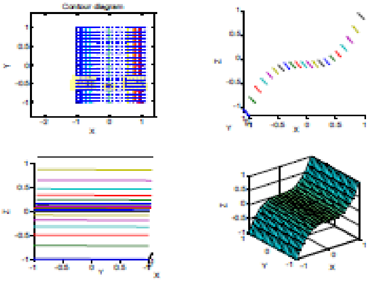
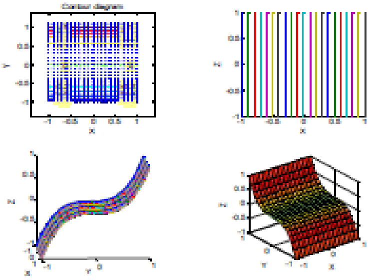
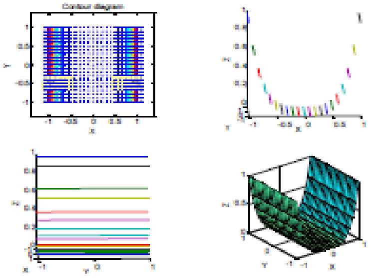
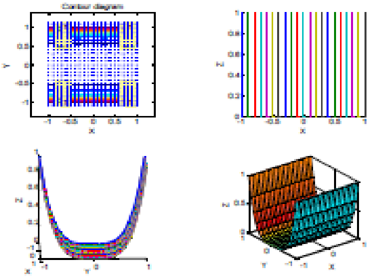
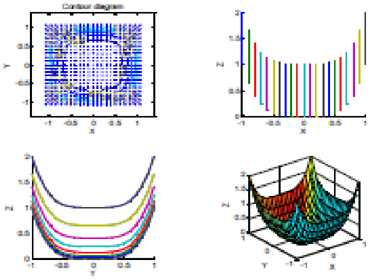
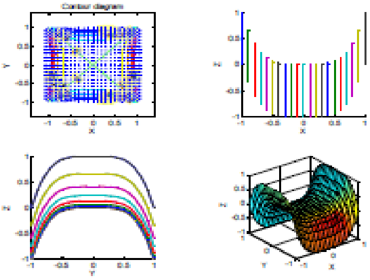
🕮 Polynomial function

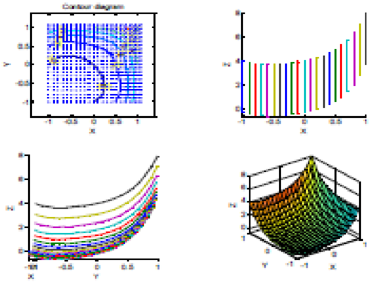
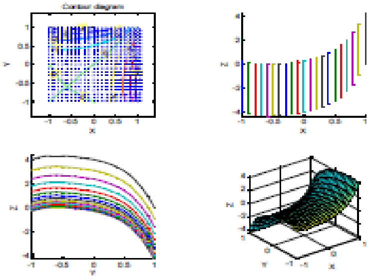
The third order and fourth order two variable polynomial profiles are incorporated in fig. 4.8

Fig. 4.8: Matlab program and multi-dimensional graphics for third order and fourth order polynomials

```
 %
%     xyz_pol34.m  R S Rao 5/9/94
%
```

```
clean

nn = input(' Give   n < 1 to 8> : ');
%
%
           lt = 1;inc=0.1; [x,y] = meshgrid(-lt: inc:lt ,-lt :inc:lt );
           disp('lt = 1 ,inc = 0.1, [x,y] = meshgrid(-lt :.05:lt ,-lt :.05:lt );')
            par =[
%    x     y     x2    y2    x3    y3    x4    y4 #
     0     0     0     0     1     0     0     0     1
     0     0     0     0     0     1     0     0     2
     0     0     0     0     0     0     1     0     3
     0     0     0     0     0     0     0     1     4
     0     0     0     0     1     1     0     0     5
     0     0     0     0     0     0     1     1     6
     0     0     0     0     0     0     1    -1     7
     1     1     1     1     1     1     1     1     8
     1    -1     1    -1     1    -1     1    -1     9
        ]
%
%
            Fn1 = ' z =  a(1) * x    + a(2) * y +   a(3) *x.^2 + a(4) * y.^2 +
....';
            Fn2 = '      a(5) * x.^3 + a(6) * y.^3 + a(7)*x.^4 + a(8) * y.^4
;';
            disp(Fn1), disp(Fn2)
            a = par(nn,:)
%
            z =  a(1) * x + a(2) * y +        ...
a(3)  *x.^2 + a(4) * y.^2 + ....
a(5)  *x.^3 + a(6) * y.^3 +  ....
                 a(7)  *x.^4 + a(8) * y.^4 ;
%
            xyz_plot(z,x,y)
```

Polynodirmial Functions	z = a(1) * x + a(2) * y + a(3) *x.^2 + a(4) * y.^2 + ... a(5) * x.^3 + a(6) * y.^3 + a(7)*x.^4 + a(8) * y.^4 ;
 z = x.^3	 z = y.^3
 z = x.^4	 z = y.^4
 z = x.^4 + y.^4	 z = x.^4 - y.^4

 **z = a(1) * x + a(2) * y + a(3) *x.^2 + a(4) * y.^2 + ... a(5) * x.^3 + a(6) * y.^3 + a(7)*x.^4 + a(8) * y.^4**	 **z = a(1) * x - a(2) * y + a(3) *x.^2 - a(4) * y.^2 + ... a(5) * x.^3 - a(6) * y.^3 + a(7)*x.^4 - a(8) * y.^4**

Exponential function

The exponential functions in two or multi-dimensional space are very important in error analysis, curve resolution, modelling by artificial neural networks and understanding the profiles of UV-VIS spectra/chromatograms.

The response surface (Fig. 4. 9) of

$z = a1*\exp(x) + a2*\exp(y)$

is a widened bowl (lamp). When either a1 or a2 is of negative sign the surface is twisted. On the other hand when both a1 and a2 are negative the response surface is mostly a flat one with a twist at middle. The function

$z = -0.1 * \exp(x^2 - y^2)$

is exponential of quadratic in two dimensions and response surface is spectacular as it looks like a flat surface with a hole. An in-depth study of multi-dimensional exponential functions is a thrust area of investigation even in twenty first century.

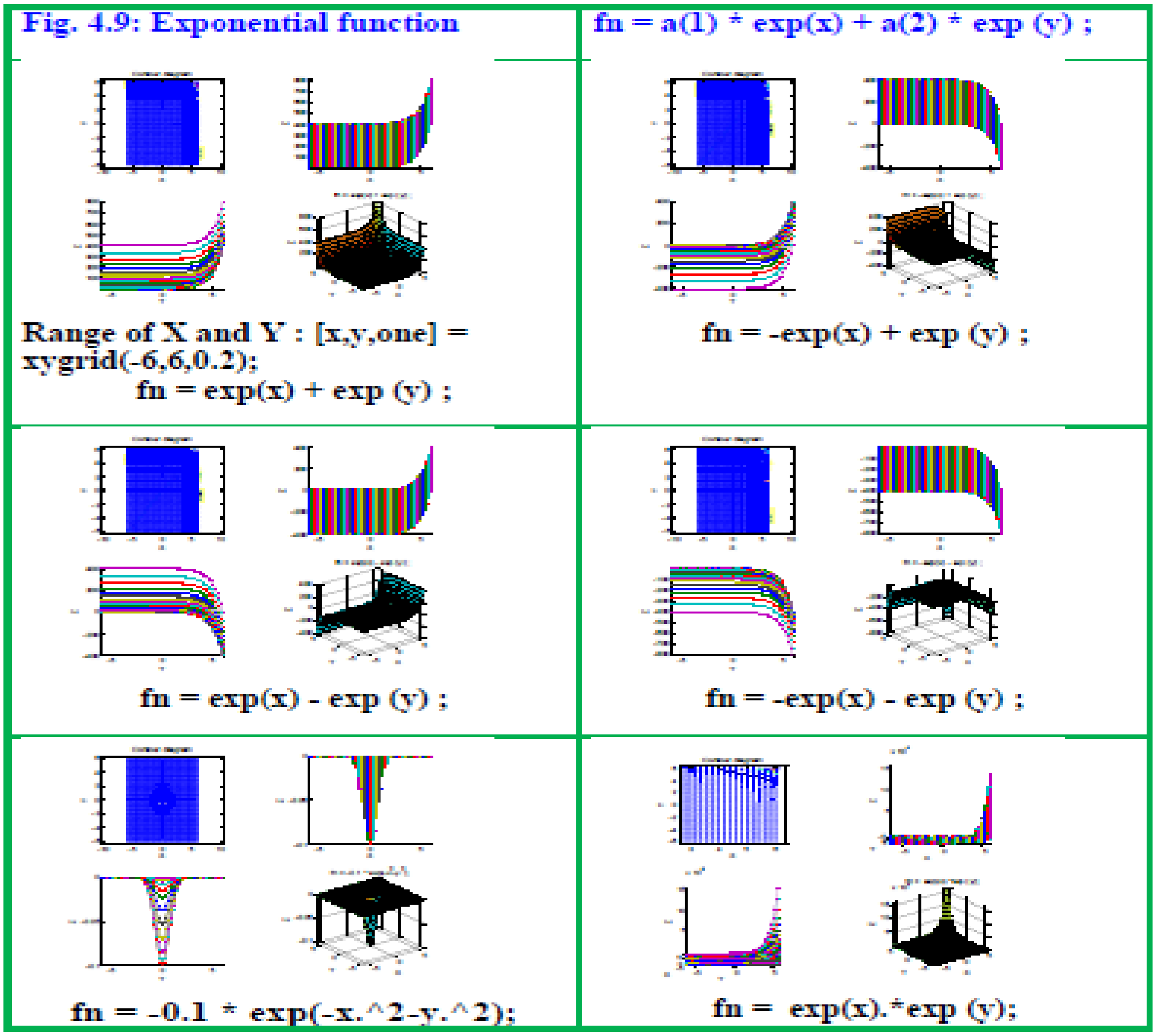


Fig. 4.9: Exponential function

fn = a(1) * exp(x) + a(2) * exp (y) ;

Range of X and Y : [x,y,one] = xygrid(-6,6,0.2);
fn = exp(x) + exp (y) ;

fn = -exp(x) + exp (y) ;

fn = exp(x) - exp (y) ;

fn = -exp(x) - exp (y) ;

fn = -0.1 * exp(-x.^2-y.^2);

fn = exp(x).*exp (y);

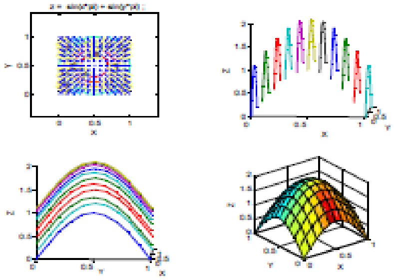
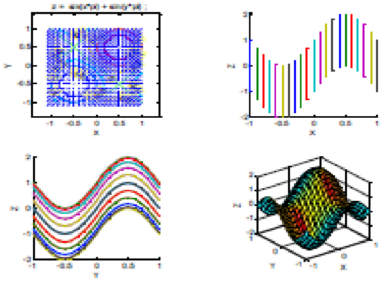
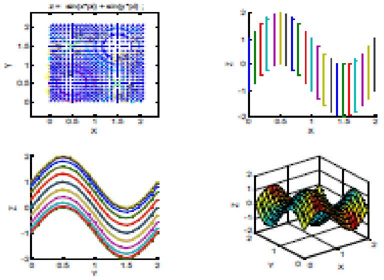
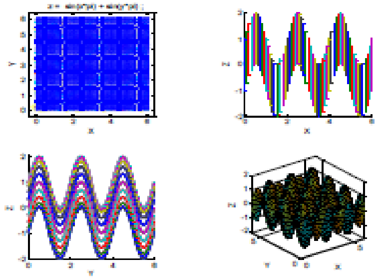
Sigmoid function

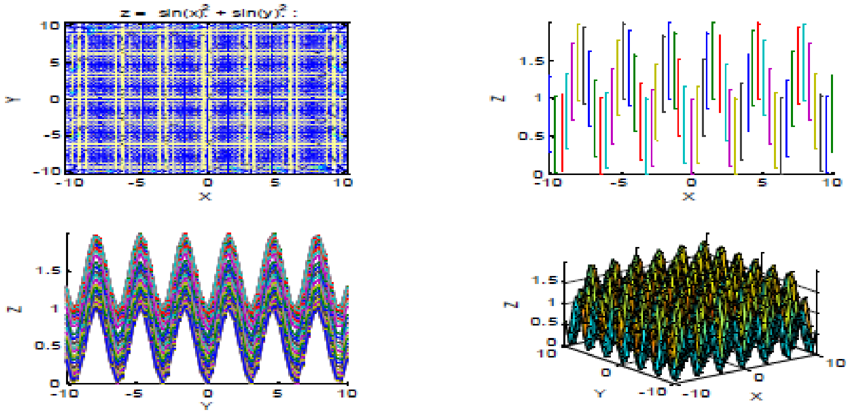
Sigmoid, tanh, radial basis functions are extensively discussed with graphics output in our earlier papers on neural networks [68,69].

Trigonometric functions

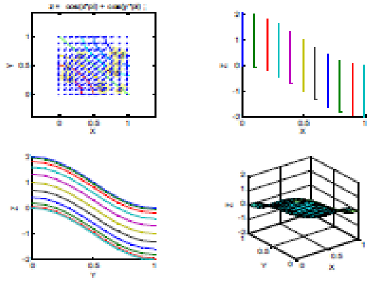
The number of optima in the response of this category of functions depend upon the range of x and y. The response surface and contour diagram for Resp = sin (x)+sin (y) in the range 0 to pi (Fig. 4. 10) are similar to Resp = x^2 +y^2 as far as number of optima are concerned. With increase in range from 0-п to 0-2п two distinct optima with Resp = +2 and -2 appear. The response vs.y amply demonstrates this and the set of non-overlapping lines correspond to univariate experiments. As range is further increased, multiple optima occur and it is what is usually observed in real life problems. In fact after looking at the experimental contours, one tries to model the phenomena. This is the inspiration behind hybrid models with polynomial terms and trigonometric/ exponential functions.Here, shape reproduction is prime goal and not mapping first principles of that disciplineinto mathematical frame.

Fig. 4.10: Trigonometric functions

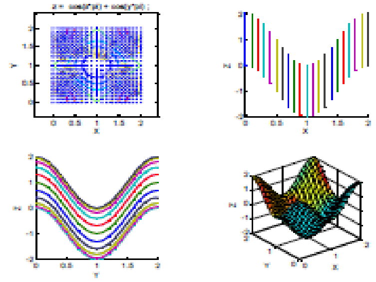
 **[x,y] = meshgrid(0:0.1:1,0.:0.1:1);** **z = sin(x*pi) + sin(y*pi) ;**	 **[x,y] = meshgrid(-1:0.1:1,-1.:0.1:1);** **z = sin(x*pi) + sin(y*pi) ;**
 **[x,y] = meshgrid(0:0.1:2,0.:0.1:2);** **z = sin(x*pi) + sin(y*pi) ;**	 **[x,y] = meshgrid(0:0.1:6,0.:0.1:6);** **z = sin(x*pi) + sin(y*pi) ;**



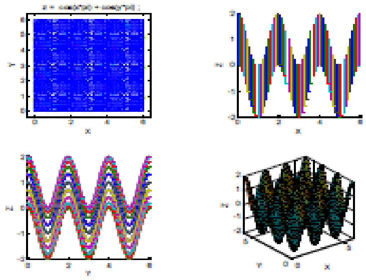
```
[x,y]  = meshgrid(-10:0.4:10,-10.:0.4:10);
 z =   sin(x).^2 + sin(y).^2  ;
```



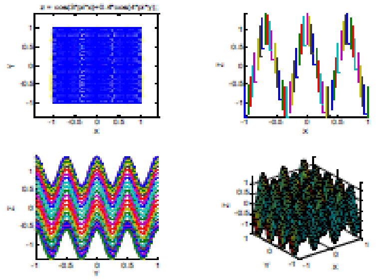
```
[x,y] = meshgrid(0:0.1:1,0.:0.1:1);
 z =  cos(x*pi) + cos(y*pi) ;
```



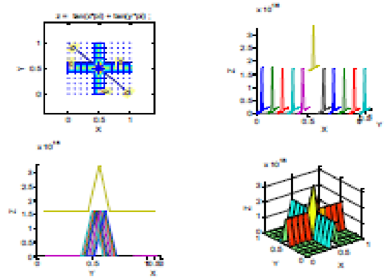
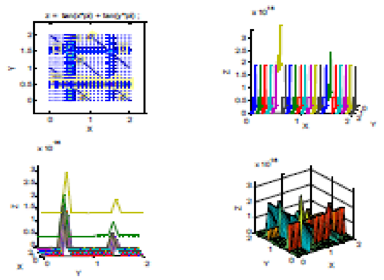
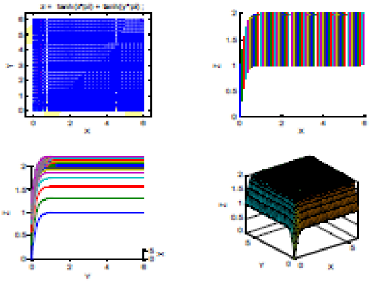
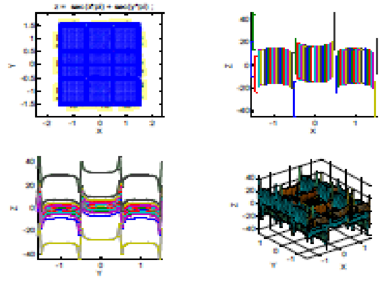
```
[x,y] = meshgrid(0:0.1:2,0.:0.1:2);
 z =  cos(x*pi) + cos(y*pi) ;
```

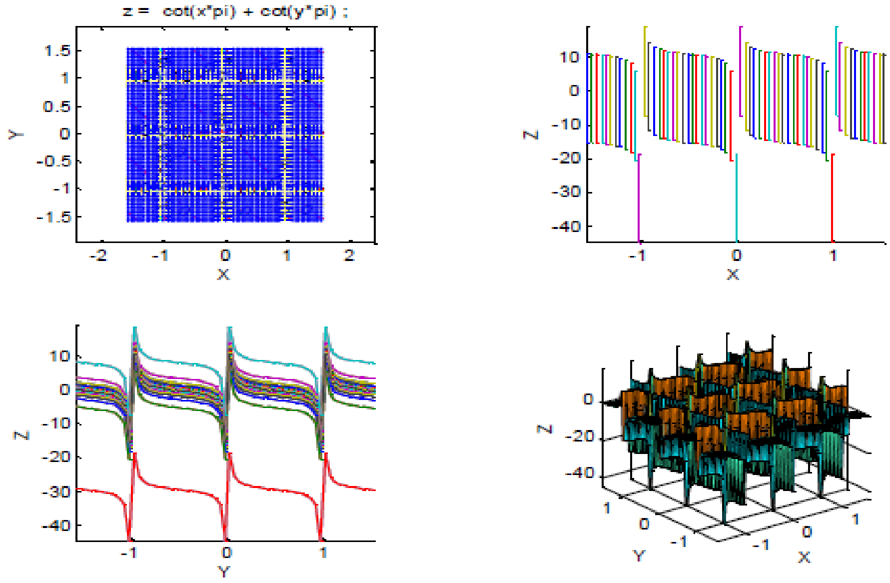


```
[x,y] = meshgrid(0:0.1:6,0.:0.1:6);
```



```
[x,y] = meshgrid(-1:0.04:1,-
```

z = cos(x*pi) + cos(y*pi) ;	**1.:0.04:1);** **z = cos(3*pi*x)+0.4*cos(4*pi*y);**
	
[x,y] = meshgrid(0:0.1:1,0.:0.1:1); **z = tan(x*pi) + tan(y*pi) ;**	**[x,y] = meshgrid(0:0.1:2,0.:0.1:2);** **z = tan(x*pi) + tan(y*pi) ;**
	
[x,y] = meshgrid(0:0.1:6,0.:0.1:6); **z = tanh(x*pi) + tanh(y*pi) ;**	**[x,y] = meshgrid(-pi/2+0.01:0.05:pi/2-0.01,-pi/2:0.05:pi/2);** **z = sec(x*pi) + sec(y*pi) ;**



[x,y] = meshgrid(-pi/2+0.01:0.05:pi/2-0.01,-pi/2:0.05:pi/2);
z = cot(x*pi) + cot(y*pi) ;

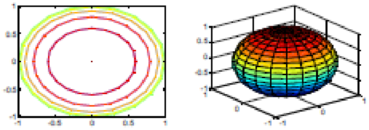
```
%
%      xyz_tri.m
%
function th_tri(nn)
     clean,n = 99;


while n >0
if nargin <1
      nn = input('give n < 1 to 17> [if n=0, exit]: ');
end
if nn == 0, return,end
%
if nn ==1
range = '[x,y] = meshgrid(0:0.1:1,0.:0.1:1);';
          st = 'z =  sin(x*pi) + sin(y*pi) ;';
end
%
if nn ==2
range = '[x,y] = meshgrid(-1:0.1:1,-1.:0.1:1);';
          st = 'z =  sin(x*pi) + sin(y*pi) ;';
end
%
if nn == 3
range = '[x,y] = meshgrid(0:0.1:2,0.:0.1:2);';
          st = 'z =  sin(x*pi) + sin(y*pi) ; ';
end
%
if nn == 4
range = '[x,y] = meshgrid(0:0.1:6,0.:0.1:6);';
          st = ' z =  sin(x*pi) + sin(y*pi) ;';
end
%
if nn == 6
range = '[x,y] = meshgrid(-10:0.4:10,-10.:0.4:10);';
          st = ' z =  sin(x).^2 + sin(y).^2 ;';
end
%
if nn == 7
range = '[x,y] = meshgrid(0:0.1:1,0.:0.1:1);';
          st = ' z =  cos(x*pi) + cos(y*pi) ;';
end
%
if nn == 8
range = '[x,y] = meshgrid(0:0.1:2,0.:0.1:2);';
          st = ' z =  cos(x*pi) + cos(y*pi) ;';
end
if nn == 9
range = '[x,y] = meshgrid(0:0.1:6,0.:0.1:6);';
          st = 'z =  cos(x*pi) + cos(y*pi) ;';
end
%
if nn == 10
range = '[x,y] = meshgrid(-1:0.04:1,-1.:0.04:1);';
            st = 'z = cos(3*pi*x)+0.4*cos(4*pi*y);';
end
%
if nn == 11
range = '[x,y] = meshgrid(0:0.1:1,0.:0.1:1);';
          st = ' z =  tan(x*pi) + tan(y*pi) ;';
end
%
if nn == 12
```

```
range = '[x,y] = meshgrid(0:0.1:2,0.:0.1:2);';
          st = 'z =  tan(x*pi) + tan(y*pi) ;';
end
%
if nn == 13
range = '[x,y] = meshgrid(0:0.1:6,0.:0.1:6);';
          st = ' z =  tan(x*pi) + tan(y*pi) ;';
end
%
if nn == 14
range = '[x,y] = meshgrid(0:0.1:6,0.:0.1:6);';
          st = ' z =  tanh(x*pi) + tanh(y*pi) ;';
end
%
if nn == 13
range = ' [x,y] = meshgrid(-1:0.01:1,-1.:0.01:1);';
          st = ' z =  cosh(x*pi) + cosh(y*pi) ;';
end
%
if nn == 15
range = ' [x,y] = meshgrid(-5:0.01:5,-5.:0.01:5);';
          st = ' z =  sinh(x*pi) + sinh(y*pi) ;';
end
%
if nn == 16
          range = '[x,y] = meshgrid(-pi/2+0.01:0.05:pi/2-0.01,-pi/2:0.05:pi/2);';
          st = ' z =  sec(x*pi) + sec(y*pi) ;';
end
%
if nn == 17
           range = '[x,y] = meshgrid(-pi/2+0.01:0.05:pi/2-0.01,-pi/2:0.05:pi/2);';
           st = ' z =  cot(x*pi) + cot(y*pi) ;';
end
%
       disp(range),disp(st)
       eval(range),eval(st)
       xyz_plot(z,x,y,st)
end% end while
```

Combination of functions from different paradigms

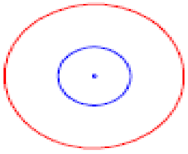
The combinations of deterministic, stochastic, fuzzy and chaotic domains result in combinatorial number of functions with interesting properties. The Hamilton's geometric algebra is unique and has potential applications in interdisciplinary research. The response surfaces for some of the typical non-linear functions are given in fig. 4.11. As visual appreciation beyond 3D is difficult and impossible beyond fifth dimension, multivariate mathematical modelling is used.

Fig. 4.11: Hybrid functions

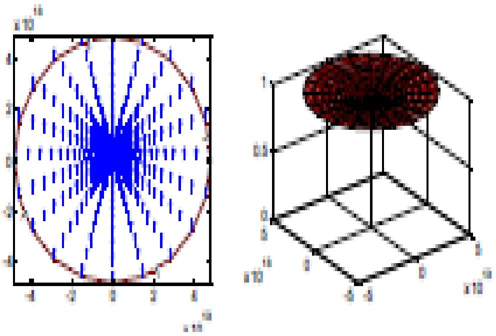


Sphere

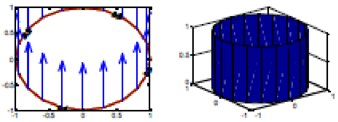
```
clean
[x,y,z] = sphere;
axis square
subplot(222),surf(x,y,z)
subplot(221),contour(x,y,z)
```



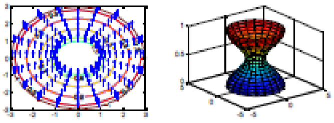
`Concentric circles`



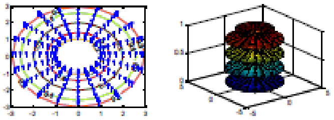
Disk with a hole



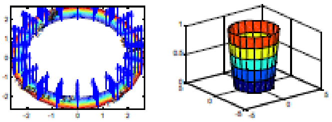
Cylinder



Basket

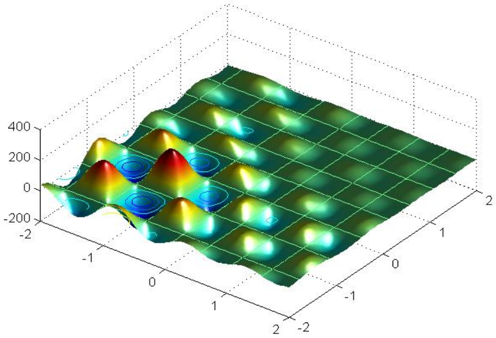


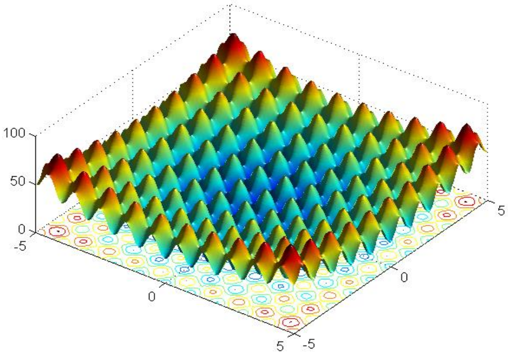
Stack of discs



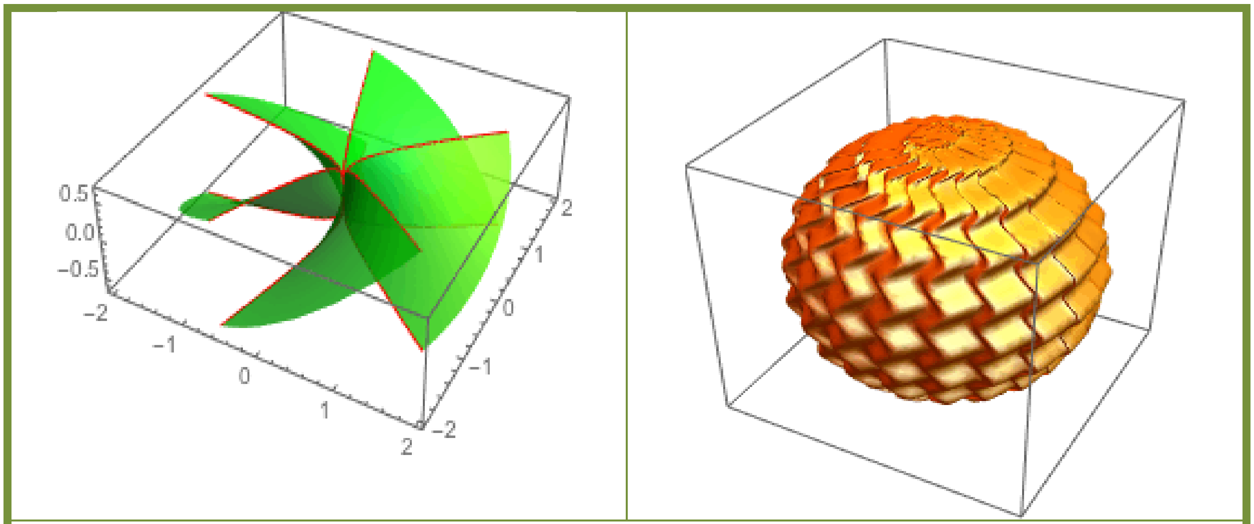
Glass

Multiple optima





(Courtesy of Matlab)



(Courtesy of Mathematica)

5. Unfolding (decomposition) of mthorder tensorial Data into lower order

A 3-way-tensor can be decomposed into 2way-matrices and PCA is applied. It has the limitation that the results vitiated as compared to multi-way PCA of unfolded tensor. The complete description and mathematical implications [113] will be published separately

6. MethodBasein Chemometrics

We introduced method bases in optimization, complex equilibria and chemical kinetics under the head data-, knowledge and method bases in chemical sciences [120-123]. The necessary conditions, algorithms in matrix algebra, failure conditions and remedial measures for factor analysis, one of popular dimension reduction techniquesare introduced inAppendix A4.The recent developments in methods, algorithms of PCA, PLSC, factor analysis and their applications in chemical sciences will be reviewed separately.

Modeling

It is a well-known fact that the complex physico-chemical processes cannot be modelled from the first principles. As such many simplifying assumptions are made in developing models. It is a misconception that models are useless simply because they cannot explain the observed phenomena very closely or fail in prediction. Respecting the proposition that all models are wrong but some are Useful, the activity of modelling continues to get better insight into the physic-chemical processes.

The variation of response or its function with the magnitudes of influencing physico-chemical factors can be represented precisely using the models. The classification of models is based on several heads like theoretical and empirical models depending upon the existence of physic chemical theory. The latter may be soft and hard models based on the type of statistical techniques employed.

Theoretical Models: The variation of absorption of a complex with its concentration, change of log k with substituent parameters etc. are expressed by Beer-Lambert's law and Hammett equation, respectively. These being linear, the slope and intercept of the regression lines bear chemical significance. The slope of the Beer-Lambert's law is extinction coefficient.

Empirical Models: It is well known that the primary data (response), percentage yield etc depends upon influential factors, like Concentration of chromogenic reagent, temperature and ambient conditions. The parameters like equilibrium/rate constant, Kovot index and dielectric constant vary with physic-chemical factors like mole fraction, pH, ratio of ingredients etc. In most of the cases, it is not possible to derive theoretical models for the variations from the first principles. Therefore, the variation is explained using empirical models only. Here, the regression coefficients are useful only to reproduce the data reasonably but do not have any physic-chemical significance.Recently, Kowalski proposed hybrid models called chemnets. Here, the experimental data is fitted into theoretical models. If the trends in the residuals indicate the inadequacy of the theoretical model or violation of some of the assumptions, then the residuals are fitted into an empirical model. The availability of empirical models in abundance will result in modification of existing theoretical models/cropping up of new ones.

7. Software

Kriging interpolation algorithm is one of the popular numerical methods used to develop 3D-response surfaces/contour diagrams to represent Resp=f(x, y) (chart 7.1). MATLAB and SURFER are widely employed to develop 3D surfaces and contour maps from data or in a user-chosen equation.

Chart 7.1: Method – base for 3D surfaces

Kriging
B-splines
Cubic splines
Bi-cubic splines
Finite Element method
Neural network
Bi-variable polynomial interpolation
Blending function
Tensor product interpolation
Moving least squares

MATLAB: It is a high tech computational and visualization software tool implementing data display in multi-dimensional tables, 2D-, 3D- surfaces, traditional algebraic, matrix, tensor formats, structures (like relational data bases), object oriented format and cells. MATLAB is now one of the coveted high level computational paradigms serving as a language, white box mode tool boxes and also black-box inbuilt functions. MATLAB recognizes whether the variable is a scalar/vector/matrix and performs algebraic/trigonometric/exponential/logarithmic operations following the rules of matrix algebra. Further, extended rules are invoked in order deal with element-wise algebraic/Boolean manipulations rending the programming nearer to the human conception.

Tensor algebra and display is only part of its scope. It has symbolic algebra toolbox and modules can be developed in OOP and also in relational DB. Thus, TensorLab is more apt as it handles scalar to m-way tensors. Thus, I preferred TensorLab in order to bring algebraic manipulations as well as matrix, extended-matrix and tensorial operations with paper and pencil as well as using MATLAB software and TENSLAB, its evolved form.

In chemometric computations involve procedures dependent on chemical process, statistical methods and mathematical algorithms. During the development of number crunching and heuristic programs for complex equilibria, chemical kinetics and quantitation in different languages and packages – FORTRAN, BASIC, TURBOPROLOG, MATLAB and dBase III+, it was recognized that many modules can be viewed as objects. However, Object Orientated Programs (OOP) can be developed in any programming language. Here, many modules are separately developed in OOP and tested as separate objects. The concept of OOP is achieved due to built – in features of MATLAB.

PCA (Principal component analysis)

The linear/ non-linear regressions belong to hard-modeling category.The objective of PCA was to surmount the difficulty of analysing data with statistically correlated variables. But, if the errors in x or y are high or if there are outliers, again the difficulty arises. If the errors are high, more number of PCs than actual number results. This is intelligently used by taking into consideration of PCs explaining the variances up to a pre-fixed limit (say 99%, 95%, 90% .... in different fields and tasks). But, if outliers are present, the PCA collapses and it paved way in proposing robust_PC analysis. A list of multi-variate analysis software is given in Chart 7.2.

Chart 7.2: Typical multi-variate software web sites

Free software	
Parallel factor analysis, N-way partial least-squares and other multi-way methods	http://www.models.kvl.dk/algorithms

Multivariate curve resolution	http://www.ub.es/gesq/mcr/mcr.htm
Multivariate curve resolution	http://personal.ecu.edu/gemperlinep
Several first-order and multi-way methods	www.chemometry.com
Generalized rank annihilation and direct trilinear decomposition	http://www.cpac.washington.edu

Chart 7.2b: Commercial software		
Software Company Website		
First-order partial least-squares GRAMS IQ	Thermo Scientific	www.thermoscientific.com
First-order partial least-squares PLS Toolbox	Eigenvector.com	www.eigenvector.com
First-order partial least-squares PLS Toolbox	The MathWorks	www.mathworks.com
First-order and N-way partial least-squares, and multivariate curve resolution	UNSCRAMBLER Camo	www.camo.com
First-order partial least-squares and multivariate curve resolution	PIROUETTE Infometrix Software	www.infometrix.com
First-order partial least-squares EZINFO	Umetrics	www.umetrics.com
PARAFAC and other multi-way methods 3 Way Pack	The Three-mode Company	http://three-mode.leidenuiv.nl

XYZTOB:A software package, XYZTOB (Three Dimensional Tool Box) in the form of a tool box is developed in MATLAB (Version 3.5f). It runs on any IBM micro-computer environment. XYZTOB is modular and easily upgradable, as source files are available to the user. All computations are performed in matrix notation. In the data mode with several x and y variables, classical univariate scatter diagrams with simple statistics like variance, covariance is given. This is useful to choose the interaction terms, detect collinearity between explanatory variables and approximately to predict order of polynomial. The object oriented representation of functions is used to display as well as to execute by MATLAB is developed. They are under rigorous testing to include in HOT_XYZ at undergraduate level and also to non-majors of tensorial computations.

Acknowledgements

RSR expresses gratitude to the teachers Sri Bhavani, Kesava Rao in CVR GMH school, Sri Kedareswara Rao, Khasim peeran, Radha Krishna Murthy, Raghavachari, Bhavannarayana, Subba Raju, Nakka Rama Rao,Subrahmanyam , Rama Krishna, Vijaya Kumar of SRR Govt College. The nice gesture of colleges Prof K V Suryanarayana, Prof J Lakshmi Narayana, Dr. L. Venkateswara Rao, and Prof, K.Narasimha Murti, classmate at under graduation and M Suryanarayana, class fellow in high school and BSc. through over four and half decades deserves appreciation. I was academically influenced by error propagation effects by Cabani, critical analysis of research results of Prof. Harry Irving, Prof. W A E McBryde, H. Siegel and Zuberhuler. The computer series in chemical education was a source of inspiration to publish in CAI, ICAI, HAI and ICE.We introduced Martell model, SCOGS, Gran-plots in complex equilibria course for M. Sc., physical chemistry from 1978 onwards.

I was a classical chemist, four decades back, performing experiments with off-the shelf pure chemical compounds (pharmaceutical-drugs (anti-tubercular, antihypertensive, anticancer), amino acids, hydrazides, hydrazones etc.) using classical pH meters, spectrophotometers under ambient conditions of one atmosphere and below 50^0C in 10 to 100 dm^3 volumes. The purpose was to arrive at chemically significant parameters (equilibrium constants, rate constants, extinction coefficients and so on) to probe into binding characteristics, rate of formation, solute-solvent interactions in aqueous and aquo-organic mixtures (of varying dielectric constant, coordinating behavior, solvating capacity) with a goal of as accurate as possible estimation of analytes, control equilibrium and rate of chemical interactions to advantage. With time and nature’s course, we plunged into writing in-house FORTRAN code for relieving the drudgery of number crunching in the initial stage. Subsequently, we started using the then state-of-art-software developed by expert chemists. The saga took a new direction in developing expert systems, databases, pedagogical MATLAB modules for Chemometrics. We had the opportunity of calculating high end computations with supercomputers (Italy) and started calculating electron density, ESP, quantum chemical derived chemically significant parameters for small (20 to 30 atoms) molecules in gas phasefollowed in aqueous and water miscible solvents. The application of evolutionary algorithms to chemical problems including SXR, quantitation from multi-sensor data broadened the perspective of inter-/intra-/cross cutting-edge disciplines. At this juncture of exploding literature reports, one should spare a moment to arrive at unbiased futuristic trend of human endeavors in science, just not be plugged with blur vision of truth as false and false as truth.

The project on complex equilibria did not receives funds. The unsuccessful fund attraction did not hamper upward trend in research outcome. The application of chemometric methods in complex equilibria resulted in many publications in international journals. The suggestion of applying Chemometrics in chromatography was not considered with a plea that instrument vendor supplies software. The use of PLSR, PCR and NNs in NIR spectrophotometer was ignored by a national instrument manufacturer saying that the program can be directly given to them without exhaustivetesting on many synthetic/real life samples and simulation studies. The intelligent database for chemometric technique with KB was not understood by their workgroup. The application of NNs in nuclear based research was thought out-of-focus of the program although extensive literature was available for nature inspired algorithms in reactor research.

In 1992, extensive simulations of proton-ligand complexes (of overlapping and non-overlapping) with the effect of random noise of Gaussian and other distributions were carried out. This was in continuation of our studies on effects of pessimistic errors in concentrations of metal, ligand, acid, alkali etc. on betas of ML_lH_h type complexes. Our proposal of safe area diagrams in our PL studies and SITECON software were to caution the best use of approximate methods before using the then advanced algorithms like MARQUARDT, Eigen analysis etc. Our recent interest in applying nature inspired algorithms to the complex equilibria, SXR, kinetics and MVC are in the back-drop of the difficulties of optimization, solution methods practiced in the last century.

REFERENCES

[1] Xiang-Dong Qing, Hai-Long Wu, Xi-Hua Zhang, Y. Li, Hui-Wen Gu, Ru-Qin Yu, *Anal. Chim.Acta.,***2015,** 861, 12-24.
A novel fourth-order calibration method based on alternating quinquelinear decomposition algorithm for processing high performance liquid chromatography–diode array detection– kinetic-pH data of naptalam hydrolysis

[2] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, *Anal. Chim.Acta.,***2015,** (In Press).
Data fusion methodologies for food and beverage authentication and quality assessment – A review

[3] A. Cañas, P. Richter, G. M. Escandar, *Anal. Chim.Acta.,***2014,** 852, 105-111.
Chemometrics-assisted excitation–emission fluorescence spectroscopy on nylon-attached rotating disks. Simultaneous determination of polycyclic aromatic hydrocarbons in the presence of interferences

[4] S. A. Bortolato, A. C. Olivieri, *Anal. Chim.Acta.,***2014,** 842, 11-19.
Chemometric processing of second-order liquid chromatographic data with UV–vis and fluorescence detection. A comparison of multivariate curve resolution and parallel factor analysis 2

[5] H. Parastar, N. Akvan, *Anal. Chim.Acta.,***2014,** 816, 18-27.
Multivariate curve resolution based chromatographic peak alignment combined with parallel factor analysis to exploit second-order advantage in complex chromatographic measurements

[6] M. R. Alcaráz, G. G. Siano, M. J. Culzoni, A. M. de la Peña, H. C. Goicoechea, *Anal. Chim.Acta.,***2014,** 809, 37-46.

Modeling four and three-way fast high-performance liquid chromatography with fluorescence detection data for quantitation of fluoroquinolones in water samples

[7] L. Rubio, M.C. Ortiz, L.A. Sarabia, *Anal. Chim.Acta.*,**2014,** 820, 9-22.
Identification and quantification of carbamate pesticides in dried lime tree flowers by means of excitation-emission molecular fluorescence and parallel factor analysis when quenching effect exists

[8] N. Omidikia, H. Abdollahi, M. Kompany-Zareh, *Anal. Chim.Acta.*,**2013,** 782, 12-20.
On uniqueness and selectivity in three-component parallel factor analysis

[9] A. Baum, P. W. Hansen, A. S. Meyer, J. D. Mikkelsen, *Anal. Chim.Acta.*,**2013,** 790, 14-23.
Simultaneous measurement of two enzyme activities using infrared spectroscopy: A comparative evaluation of PARAFAC, TUCKER and N-PLS modeling

[10] A. Baum, A. S. Meyer, J. L. Garcia, M. Egebo, P. W. Hansen, J. D. Mikkelsen, *Anal. Chim.Acta.*,**2013,** 778, 1-8.
Enzyme activity measurement via spectral evolution profiling and PARAFAC

[11] A. Malik, R. Tauler, *Anal. Chim.Acta.*,**2013,** 794, 20-28.
Extension and application of multivariate curve resolution-alternating least squares to four-way quadrilinear data-obtained in the investigation of pollution patterns on Yamuna River, India—A case study

[12] A.S. Luna, R.A. Luiz, I.C.A. Lima, P.H. Março, P. Valderrama, R. Boqué, J. Ferré,*Anal. Chim.Acta.*,**2013,** 778, 9-14.
Simultaneous determination of aflatoxins B2 and G2 in peanuts using spectrofluorescence coupled with parallel factor analysis

[13] C.W. Cuss, C. Guéguen *Anal. Chim.Acta.*,**2012,** 733, 98-102.
Determination of relative molecular weights of fluorescent components in dissolved organic matter using asymmetrical flow field-flow fractionation and parallel factor analysis

[14] F. D. Cid, R. I. Antón, R. Pardo, M. Vega, E. Caviedes-Vidal, *Anal. Chim.Acta.*,**2011,** 705, 1–2, 243-252.
Modelling spatial and temporal variations in the water quality of an artificial water reservoir in the semiarid Midwest of Argentina

[15] Hadi Parastar and Roma Tauler,*Anal. Chem.2014,86,286−297.*
Multivariate Curve Resolution of Hyphenated and MultidimensionalChromatographic Measurements: A New Insight to Address CurrentChromatographic Challenges

[16] B.R. Kowalski,*Anal. Chem.***1993,**65,835A.

[17] A. F. T. Moore, H. C. Goicoechea, F. Barbosa, Jr., and A. D. Campiglia, *Anal. Chem.*, **2015,** 87,(10),5232–5239.
Parallel Factor Analysis of 4.2 K Excitation–Emission Matrices for the Direct Determination of Dibenzopyrene Isomers in Coal-Tar Samples with a Cryogenic Fiber-Optic Probe Coupled to a Commercial Spectrofluorimeter

[18] F. Laborda, E. Bolea, and J. J. Lamana, *Anal. Chem.*, **2014,**86,2270−2278.
Single Particle Inductively Coupled Plasma Mass Spectrometry: A PowerfulTool for Nanoanalysis

[19] A. Khoobi, S. M. Ghoreishi, M. Behpour, and S. Masoum, *Anal. Chem., **2014,***86, (18), 8967–8973.

Three-Dimensional Voltammetry: A Chemometrical Analysis of Electrochemical Data for Determination of Dopamine in the Presence of Unexpected Interference by a Biosensor Based on Gold Nanoparticles

[20] S. Nie, E. Benito-Peña,, H. Zhang, Y. Wu, and D. R. Walt,*Anal. Chem.*, **2013,**85,9272−9280.
Multiplexed Salivary Protein Profiling for Patients with Respiratory Diseases UsingFiber-Optic Bundles and Fluorescent Antibody-BasedMicroarrays

[21] F. Allegrini and A. C. Olivieri, *Anal. Chem.*, **2012,** 84, (24), 10823–10830.
Analytical Figures of Merit for Partial Least-Squares Coupled to Residual Multilinearization

[22] A. C. Olivieri and K. Faber, *Anal. Chem.*, **2012,** 84, (1), 186–193.
New Developments for the Sensitivity Estimation in Four-WayCalibration with the Quadrilinear Parallel Factor Model

[23] S. Furbo and J. H. Christensen, *Anal. Chem.*, **2012,** 84, (5), 2211–2218.
Automated Peak Extraction and Quantification in Chromatography with Multichannel Detectors

[24] Y. Akhlaghi, M. Kompany-Zareh, and M. R. Hormozi-Nezhad, *Anal. Chem.*, **2012,**84,(15),6603–6610.
Multiway Investigation of Interaction between Fluorescence Labeled DNA Strands and Unmodified Gold Nanoparticles

[25] N. González-Diéguez, A. Colina, J. López-Palacios, and A. Heras, *Anal. Chem.*, ***2012,***84, (21), 9146–9153.
Spectroelectrochemistry at Screen-Printed Electrodes: Determination of Dopamine

[26] A. Yilmaz, N. T. Nyberg, and J. W. Jaroszewski, *Anal. Chem.*, **2011,** 83 (21), 8278–8285.
Metabolic Profiling Based on Two-Dimensional J-Resolved 1H NMR Data and Parallel Factor Analysis

[27] H. Parastar, J. R. Radović, M. Jalali-Heravi, S. Diez, J. M. Bayona, and R. Tauler, *Anal. Chem.*, **2011,** 83, (24), 9289–9297.
Resolution and Quantification of Complex Mixtures of Polycyclic Aromatic Hydrocarbons in Heavy Fuel Oil Sample by Means of GC × GC-TOFMS Combined to Multivariate Curve Resolution

[28] A.C. Olivieri, *Anal. Chem.*, **2008,** 80, 5713–5720.
Analytical advantages of multivariate data processing. One, two, three, infinity?

[29] A. G. Reiriz, P. C. Damiani, A. C. Olivieri, F. C. Canada,A. Munoz de la Pen, *Anal. Chem.*, **2008,**80,7248–7256.
Nonlinear Four-Way Kinetic-Excitation-EmissionFluorescence Data Processed by a Variant of Parallel Factor Analysis and by a Neural Network Model Achieving the Second-Order Advantage: Malonaldehyde Determination in Olive Oil Samples

[30] R.E. Mohler, K. M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, *Anal. Chem.*,**2006,** 78, 2700–2709.
Comprehensive two-dimensionalgas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells,

[31] S. O. Hagan, W. B. Dunn, M. Brown, J. D. Knowles, and D. B. Kell,*Anal. Chem.*,**2005,**77,290-303.
Closed-Loop, Multiobjective Optimization of Analytical Instrumentation: Gas Chromatography/Time-of-Flight Mass Spectrometry of theMetabolomes of Human Serum and of Yeast Fermentations

[32] J. M. Rooney, and E. A. H. Ha,*Anal. Chem.*,**2004,**76,23,6861-6870.
Surface Plasmon Resonance: Theoretical Evolutionary DesignOptimization for a Model Analyte Sensitive Absorbing-Layer System

[33] J. M. Henshaw, L.W. Burgess, K. S. Bookish and B.R. Kowalski*Anal. Chem.*,**1994,** 66, 3328.

[34] A.P.Wade, P.M. Shiundu and P.D. Wentzell,*Anal.Chim.Acta*,**1990,** 237361.

[35] A. Mohd, J.R.Dean and W.R.Tomlison,*Analyst*,**1992,** 1171743.

[36] G. Nageswara Rao, S. V. V. Satyanarayana Rao, A. Ravindra Babu, R. Sambasiva Rao,*Asian J.Chem.*,**1992,** 4, 99.
Computer Oriented Mathematical Models for chemical systems

[37] P. Avetta, S. Berto, A. B. Prevot, M. Minella, E. Montoneri, D. Persico, D. Vione, M.C. Gonzalez, D.O. Mártire, L. Carlos, A. Arques, *Chemical Engineering Journal,* **2015,** 274,247-255. Photoinduced transformation of waste-derived soluble bio-based substances

[38] C. Li, S. C. Yang, Q. S. Guo, Kai-Yan Zheng, Yong-Feng Shi, Xue-Feng Xiao, Guang-Qiang Long, *Chemomet. Intel. Lab. Syst.*, **2014,** 136, 115–120.
Determining the geographical origin of the medicinal plant Marsdenia tenacissima with multi-element analysis anddata mining techniques

[39] A. G. García-Reiriz, *Chemomet. Intel. Lab. Syst.*, **2014,** 137, 120–127.
Parallel factor analysis and multivariate curve resolution as data fusiontools to supervise a stream

[40] Zhaomin L, X. Yan, Q. Jiang *Chemomet. Intel. Lab. Syst.*, **2014,**137, 128–139.
Batch process monitoring based on just-in-time learning andmultiple-subspace principal component analysis

[41] C. Ranzan, A. Strohm, L. Ranzana, L. F. Trierweiler,B. Hitzmann, J. O. Trierweiler, *Chemomet. Intel. Lab. Syst.*, **2014,** 132,133–140.
Wheat flour characterization using NIR and spectral filter based on Ant ColonyOptimization

[42] N. Akvan, H. Parastar, *Chemomet. Intel. Lab. Syst*, **2014,**137,146–154.
Second-order calibration for simultaneous determination ofpharmaceuticals in water samples by solid-phase extraction and fasthigh-performance liquid chromatography with diode array detector

[43] Y. Li, H. L. Wu, X. H. Zhang, Y. Chen, H. W. Gu, Qi Zuo, Y. Zhang, Shan-Shan Guo, Xin-Yi Liu, Ru-Qin Yu, *Chemomet. Intel. Lab. Syst.*, **2014,** 136, 15–23.
Estimating the chemical rank of three-way fluorescence data by vectorsubspace projection with Monte Carlo simulation

[44] P. L. Pisano, María F. Silva, A. C. Olivieri, *Chemomet. Intel. Lab. Syst.*, **2014,**132,1–7.
Exploration of liquid chromatographic-diode array data for Argentineanwines by extended multivariate curve resolution

[45] B. Song, Y. Ma, H. Shi, *Chemomet. Intel. Lab. Syst.*, **2014,** 135, 17–30.
Multimode process monitoring using improved dynamic neighbourhood preserving embedding

[46] M. Silvestri, A. Elia, D. Bertelli, E. Salvatore, C. Durante, M. Li Vigni, A.Marchetti, M. Cocchi,*Chemomet. Intel. Lab. Syst.*, **2014,** 137, 181–189.
A mid-level data fusion strategy for the Varietal Classification ofLambrusco PDO wines

[47] J. J. Feng, G. Hao, C. Y. Min, L. Xie, L. Liu, T. Lv, H. P. Xie, *Chemomet. Intel. Lab. Syst.*, **2014,** 135, 201–207.
Metabonomics of liver toxicity from Traditional Chinese Medicine Huang-Yao-Zistudied by mass spectrum-based orthogonal projectionmethod

[48] D. Brie, R. Klotz, S. Miron, S. Moussaoui, C. Mustin,P. Bécuwe, S. Grandemange, *Chemomet. Intel. Lab. Syst.*,**2014,**137,21–32.
Joint analysis of flow cytometry data and fluorescence spectra as a non-negative array factorization problem

[49] S. S. Ouertani, G. Mazerolles, J. Boccard, S. Rudaz, M. Hanafi, *Chemomet. Intel. Lab. Syst.*, **2014,** 133,25–32.
Multi-way PLS for discrimination: Compact form equivalent to thetri-linear PLS2 procedure and its monotony convergence

[50] A. Eslami, E.M. Qannari, A. Kohler, S. Bougearda, *Chemomet. Intel. Lab. Syst.*, **2014,**133, 63–69.
Multivariate analysis of multiblock and multigroup data

[51] S. Elcoroaristizabal, A. de Juan, J. A. García, N. Durana, L. Alonso, *Chemomet. Intel. Lab. Syst.*, **2014,**132,63–74.
Comparison of second-order multivariate methods for screening and determinationof PAHs by total fluorescence spectroscopy

[52] J. Orzel, M. Daszykowski, *Chemomet. Intel. Lab. Syst.*, **2014,** 13774–81.
A rapid validation of the antioxidant capacity of food commodities basedon their fluorescence excitation emission spectra as applicable to coffee and peppermint extracts

[53] Bo Lu, I. Castillo, L. Chiang, T. F. Edgar, *Chemomet. Intel. Lab. Syst.*, **2014,** 135, 90–109.
Industrial PLS model variable selection using moving window variable importance in projection

[54] C. Paulick, M. N.Wright, R. Verleger, K. Keller, *Chemomet. Intel. Lab. Syst.*,**2014,** 137, 97–109.
Decomposition of 3-way arrays: A comparison of differentPARAFAC algorithms

[55] X. Zhang, R. Marcé, J. Armengol, R. Tauler, *Chemosphere*, **2014,** 111, 120-128.
Distribution of dissolved organic matter in freshwaters using excitation emission fluorescence and Multivariate Curve Resolution

[56] Hua-Se Ou, Chao-Hai Wei, Ce-Hui Mo, Hai-Zhen Wu, Y. Ren, Chun-H. Feng, *Chemosphere*, **2014,** 113, 158-164.
Novel insights into anoxic/aerobic1/aerobic2 biological fluidized-bed system for coke wastewater treatment by fluorescence excitation–emission matrix spectra coupled with parallel factor analysis

[57] C.W. Cuss, C. Guéguen,*Chemosphere*, **2013,** 92, 11, 1483-1489.
Distinguishing dissolved organic matter at its origin: Size and optical properties of leaf-litter leachates

[58] S. Paul, P. W. S. Heng, L. W. Chan, *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **2014,** 451, 125-135.
Elucidating chlorin e6–sucrose ester interaction using coarse-grain modeling and fluorescence spectroscopic technique

[59] R. Sambasiva Rao and G Nageswara Rao, ***2005,*** *Himalya Publisher, New Delhi (India),*
Computer applications in Chemistry

[60] N. P. Sanchez, A. T. Skeriotis and C. M. Miller, *Environ. Sci. Technol.*, **2014,** 48, (3), 1582–1591.
A PARAFAC-Based Long-Term Assessment of DOM in a Multi-Coagulant Drinking Water Treatment Scheme

[61] S. K. L. Ishii and T. H. Boyer, *Environ. Sci. Technol.*, **2012,** 46, (4),2006–2017.
Behavior of Reoccurring PARAFAC Components in Fluorescent Dissolved Organic Matter in Natural and Engineered Systems: A Critical Review

[62] C. L. Osburn, L. T. Handsel, M. P. Mikan, H. W. Paerl, and M. T. Montgomery, *Environ. Sci. Technol.*,**2012,** 46, (16), 8628–8636.
Fluorescence Tracking of Dissolved and Particulate Organic Matter Quality in a River-Dominated Estuary

[63] I. Suryanarayana, A. Braibanti, R. Sambasiva Rao, V. A. Ramam, D. Sudarsan and G. Nageswara Rao, *Fisheries Research*, **2008,**92,115-139.
Neural Networks in Fisheries Research

[64] S. M. Azcarate, A. de A. Gomes, M. R. Alcaraz, M. C. U. de Araújo, J. M. Camiña, H. C. Goicoechea, *Food Chemistry*, **2015,** 184, 214-219.
Modeling excitation–emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety

[65] V. Ananta Ramam, V.V. Panakala Rao, K Rama Krishna and R. Sambasiva Rao, *Ind J Chem*, **2006,**45A,100-105.
Model chemistries of hydrazides II. Electronic structure of five membered aromatic hydrazides

[66] I. Suryanarayana,V. Ananta Ramam, K. M. M.Krishna Prasad and R. Sambasiva Rao, *Ind J Chem.*,**2008,** 47A, 199-206.
Model chemistries of hydrazides I. Electronic parameters of aliphatic hydrazides with AM1and PM3 Hamiltonians

[67] K. RamaKrishna, Ch. V. Kameswara Rao, V. Ananta Ramam, R. Sambasiva Rao & M. V. Subba Rao, *Indian Journal of Chemistry*, **2012,** 51A, 571-579.
Model chemistries of hydrazides III: SEMO computations of isonicotinic acid hydrazide, its valence isomers and their isopropyl derivatives

[68] K RamaKrishna, V.Anantha Ramam, R. Sambasiva Rao,*J. Applicable Chem.*, **2014,** 3,5, 1807-1893**.**
Mathematical Neural Network (MaNN) Models, Part V: Radial basis function (RBF) neural networks (NNs) in Chemometrics, Envirometrics and Medicinometrics 3D-fns (ChEM)

[69] K RamaKrishna, Ch. V. Kameswara Rao, V. Anantha Ramam, R. Sambasiva Rao, *J. Applicable Chem.*, **2014,**3,6,2209-2311.
Mathematical Neural Network (MaNN) Models, Part VI: Single-layer perceptron [SLP] and Multi-layer perceptron [MLP] Neural networks in ChEM- Lab

[70] K. RamaKrishna, Ch. V. Kameswara Rao and R. Sambasiva Rao,*J. Applicable Chem.*, **2013,**2,5, 1007-1034.
E-man Part 4: Tutorial on prospects of charged system search (CSS) algorithm in chemical sciences

[71] K. RamaKrishna, G. Ramkumar and R. Sambasiva Rao,*J. Applicable Chem.*, **2013,**2, 6, 1413-1458.
Evolution of Mimics of Algorithms of Nature (E-man) Part 5#: Tutorial on Big_Bang–Big_Crunch algorithm

[72] K. RamaKrishna, Ch. V. Kameswara Rao, R. Sambasiva Rao, *J. Applicable Chem.*, **2013,**2,4,698-713.
E-man Part 3: Tutorial on gravitational algorithm in Structure activity relationships (SXR)

[73] K. Viswanath, R. Sambasiva Rao, Ch. V. Kameswara Rao, K. Rama Krishna,B. Rama Krishna and G. E. G. Santhosh, *J. Applicable Chem.*, **2012,**1, *1*, 109-124.
SVD-Part II: Application of neural networks for classification of bauxite,

[74] R. Sambasiva Rao, *J. Chem.Sci.*,**1993,**14,23-44.

The role of Simplex method in chemical research.Speciometrics -- distribution diagrams

[75] R. Bro, El M. Qannari, Henk A. L. Kiers, T. Naes and M. B. Frost *J. Chemomet.*,**2008,**22, 36–45.
Multi-way models for sensory profiling data

[76] J. A. Westerhuis, T. Kourti and J. F. Macgregor, *J. Chemomet.*,**1999,** 13,397–413.
Comparing Alternative Approaches ForMultivariate Statistical Analysis Of Batch Process Data

[77] H. L. Wu, M. Shibukawa and K. Oguma, *J. Chemomet.*,**1998,** 12,1–26.
An Alternating Trilinear Decomposition AlgorithmWith Application To Calibration Of Hplc–Dad For SimultaneousDetermination Of Overlapped Chlorinated AromaticHydrocarbons

[78] C. Weiths, *J. Chemomet.*,**1993,**79,403.

[79] T. Naes and H. Martens, *J. Chemomet.*,**1989,**2,155.

[80] A. Braibanti, R. Sambasiva Rao, G. Nageswara Rao and N. Satyanarayna,*J. Ind Council. Chem.*, **1997,** 12,27-40.
Data, Knowledge and Method bases in Chemistry II. Data Structures in Chemistry

[81] S.V.V. Satyanarayana Rao, J.S.V.M. Lingeswara Rao, A. Ravindra Babu, D. Murali Krishna and R. Sambasiva Rao *J. Ind. Chem. Soc.*,**1996,** 73, 9-19.
Chemometric Investigation of Complex Equilibria in Solution phase: V Correlation of Formation Constants of Mn(II) or Zn(II) Complexes of AADH/FAH with Co-solvent Characteristics

[82] I. Montoliu, François-Pierre J. Martin, S. Collino, S. Rezzi and S. Kochhar, *J. Proteome Res.*, ***2009,***8, (5), 2397–2406.
Multivariate Modeling Strategy for Intercompartmental Analysis of Tissue and Plasma 1H NMR Spectrotypes

[83] A. R. Babu, R. Sambasiva Rao, *J.Chem. & Eng. Data.*,**1992,**37,526-531.
Chemometric Investigation of Complex equilibria in solution phase III. Chemical models for the complexation of Ni(II) with AADH & FAH in aquo-DMF & aquo-DOX media-Correlation with solvent parameters.

[84] K. Ishida, T. Saito, N. Aoyagi, T. Kimura, R. Nagaishi, S. Nagasaki, S. Tanaka, *Journal of Colloid and Interface Science*, **2012,** 374, 1, 258-266.
Surface speciation of Eu3+ adsorbed on kaolinite by time-resolved laser fluorescence spectroscopy (TRLFS) and parallel factor analysis (PARAFAC)

[85] M. R. Banaitis, H. Waldrip-Dail, M. S. Diehl, B. C. Holmes, J. F. Hunt, R. P. Lynch, T. Ohno, *Journal of Colloid and Interface Science*, **2006,** 304, 1, 271-276.
Investigating sorption-driven dissolved organic matter fractionation by multidimensional fluorescence spectroscopy and PARAFAC

[86] H. Xu, G. Yu, L. Yang, H. Jiang, *Journal of Hazardous Materials*, **2013,** 263, 2, 412-421.
Combination of two-dimensional correlation spectroscopy and parallel factor analysis to characterize the binding of heavy metals with DOM in lake sediments

[87] S. Shao, H. Liang, F. Qu, H. Yu, K. Li, G. Li, *Journal of Membrane Science*, **2014,** 464, 33-42.
Fluorescent natural organic matter fractions responsible for ultrafiltration membrane fouling: Identification by adsorption pretreatment coupled with parallel factor analysis of excitation–emission matrices

[88] W. G. Mendoza, R. G. Zika, *Progress in Oceanography*, **2014,** 120, 189-204.
On the temporal variation of DOM fluorescence on the southwest Florida continental shelf

[89] A. Braibanti, K. Sambasiva Rao, Ch. S. V. Prasad and R. Sambasiva Rao *Samyak J. Chem.*, **1997,** 1, *1*, 17-21.

Multi-dimensional graphics in chemistry I. Applications of three dimensional surfaces and contour diagrams

[90] Y. Zhang, J. Pan, G. Zhang, X. Zhou, *Sensors and Actuators B: Chemical,* **2015,** 206, 630-639.
Intercalation of herbicide propyzamide into DNA using acridine orange as a fluorescence probe

[91] S. Masoum, A. Gholami, M. Hemmesi, S. Abbasi, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2015,** 148, 389-395.
Quality assessment of the saffron samples using second-order spectrophotometric data assisted by three-way chemometric methods via quantitative analysis of synthetic colorants in adulterated saffron

[92] M. Vosough, S. N. Eshlaghi, R. Zadmard, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* **2015,** 136, Part B, 618-624.
On the performance of multiway methods for simultaneous quantification of two fluoroquinolones in urine samples by fluorescence spectroscopy and second-order calibration strategies

[93] M. A. El-Monem Hegazy, M. S. Eissa, O. I. Abd El-Sattar, M. A. El-Kawy, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* **2014,**128,231-242.
Two and three way spectrophotometric-assisted multivariate determination of linezolid in the presence of its alkaline and oxidative degradation products and application to pharmaceutical formulation

[94] S. Leurgans and R. T. Ross, *Stat. Sci.,***1992,**7,289.

[95] C. Weiths and H. Schmidli*Stat. Sci.,***1990,**5,175.

[96] E. Fuentes, C. Cid, M. E. Báez, *Talanta*, **2015,** 134, 8-15.
Determination of imidacloprid in water samples via photochemically induced fluorescence and second-order multivariate calibration

[97] L. Rubio, L.A. Sarabia, M.C. Ortiz, *Talanta*, **2015,** 138, 86-99.
Standard addition method based on four-way PARAFAC decomposition to solve the matrix interferences in the determination of carbamate pesticides in lettuce using excitation–emission fluorescence data

[98] R. Khani, J. B. Ghasemi, F. Shemirani, *Talanta*, **2014,** 128, 254-262.
Second-order data obtained by beta-cyclodextrin complexes: A novel approach for multicomponent analysis with three-way multivariate calibration methods

[99] R. M. Callejón, J. M. Amigo, E. Pairo, S. Garmón, J. A. Ocaña, M. L. Morales, *Talanta*, **2012,** 88, 456-462.
Classification of Sherry vinegars by combining multidimensional fluorescence,parafac and different classification approaches

[100] A. R. Babu, D. M. Krishna and R. Sambasiva Rao*Talanta*, **1993,** 40, 1873-1882.
Chemometric investigation of complex equilibria in solution phase II: Sensitivity of chemical models for the interaction of AADH and FAH with Ni(II) in aqueous medium

[101] R.H.Myers, A. I Khuri and W.H. Cart Jr,*Technomertics*, **1988,** 31137.

[102] I.E. Frank, J.H.Friedman, *Technometrics*, **1993,** 35,109.

[103] M.S. Handcock and M.L.Stein, *Technometrics,***1993,** 35,403.

[104] A. C. Olivieri, G.M. Escandar, A. M. de la Peña, *Trends in Analytical Chemistry*,**2011,** 30,607–617.
Second- and higher-order multivariate calibration methods applied to non-multi-linear data. Advantages and limitations of the different algorithms,

[105] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, R.J.Poppi, *Trends in Analytical Chemistry,* **2007,** 26, 752–765.
Second and third-order multivariate calibration: data, algorithms and applications

[106] A.C. Hambly, E. Arvin, L.F. Pedersen, P.B. Pedersen, B. Seredyńska-Sobecka, C.A. Stedmon, *Water Research,* **2015,** 83, 112-120.
Characterizing organic matter in recirculating aquaculture systems with fluorescence EEM spectroscopy

[107] Y. Wang, J. Qin, S. Zhou, X. Lin, L. Ye, C. Song, Y. Yan, *Water Research*, **2015,** 73, 252-264.
Identification of the function of extracellular polymeric substances (EPS) in denitrifying phosphorus removal sludge in the presence of copper ion

[108] Y. Shutova, A. Baker, J. Bridgeman, R. K. Henderson, *Water Research,***2014,** 54, 159-169.
Spectroscopic characterisation of dissolved organic matter changes in drinking water treatment: From PARAFAC analysis to online monitoring wavelengths

[109] N. P. Sanchez, A. T. Skeriotis, C. M. Miller, *Water Research*, **2013,** 47, 4, 1679-1690.
Assessment of dissolved organic matter fluorescence PARAFAC components before and after coagulation–filtration in a full scale water treatment plant

[110] H. Xu, H. Cai, G. Yu, H. Jiang, *Water Research*, **2013,** 47, 6, 2005-2014.
Insights into extracellular polymeric substances of cyanobacterium Microcystis aeruginosa using fractionation procedure and parallel factor analysis

[111] F. Meng, G. Huang, X. Yang, Z. Li, J. Li, J. Cao, Z. Wang, L. Sun, *Water Research,* **2013,** 47, 14, 5027-5039.
Identifying the sources and fate of anthropogenically impacted dissolved organic matter (DOM) in urbanized rivers

[112] H. Xu, H. Jiang, *Water Research*, **2013,** 47 17, 6506-6515.
UV-induced photochemical heterogeneity of dissolved and attached organic matter associated with cyanobacterial blooms in a eutrophic freshwater lake

[113] K RamaKrishna, R. Sambasiva Rao,*(Unpublished)*

[114] Rasmus Bro, Ph.D. thesis, **1998,** Univ. of Amsterdam, Denmark.
Multi-way analysis in food industry.

[115] Andreas Baum Ph.D. Thesis, May **2013**, Technical University of Denmark.
Descriptive and predictive assessment of enzyme activity and enzyme related processes in biorefinery using IR spectroscopy and Chemometrics.

[116] Giorgio Tomasi, Ph.D. Dissertation, **2006,** Supervisor Rasmus Bro,Department of Food Science, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark
Practical and computational aspects in chemometric data analysis

[117] Asmund Rinnan, Ph. D thesis, **2004,** Supervisor Rasmus Bro, Department of Food Science, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark.
Applications of PARAFAC on spectral data

[118] P. Santa-Cruz, A. García-Reiriz, Talanta, **2014**,128,450-459.
Application of third-order multivariate calibration algorithms to the determination of carbaryl, naphthol and propoxur by kinetic spectroscopic measurements

[119] Wen-Tao Li, Shi-Yu Chen, Zi-Xiao Xu, Y. Li, Chen-Dong Shuang and Ai-Min Li, *Environ. Sci. Technol.*, **2014**, *48* (5), 2603–2609.

Characterization of Dissolved Organic Matter in Municipal Wastewater Using Fluorescence PARAFAC Analysis and Chromatography Multi-Excitation/Emission Scan: A Comparative Study

[120] A Braibanti,R.SambasivaRao, G Nageswara Rao, V Ananta Ramam , SVV Satyanarayana Rao, *Ann. Chim. Rome*, **2002,** 92, 689-704.
Data, Knowledge and Method bases in Chemical Sciences, Part IV Current status in databases

[121] V Ananta Ramam, D.Subrahmanyam, S.V.V. Satyanarayna Rao, R. Sambasiva Rao,*J Teaching and Research in Chem*, **1997**, 4,20-28.
Data, Method and Knowledge bases in Chemical sciences III. A knowledge based tool for Chemical kinetics

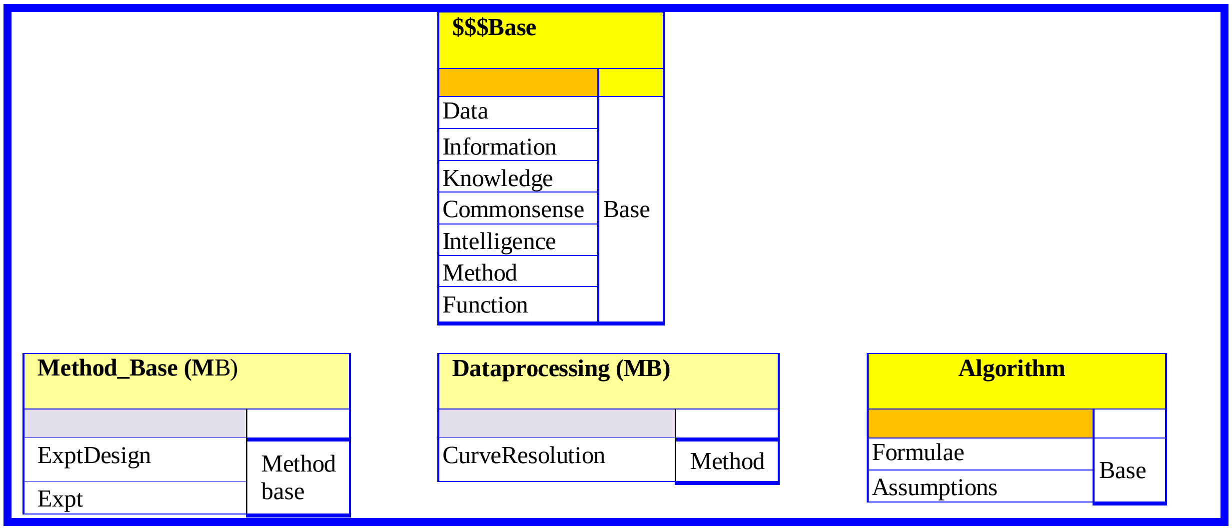
[122] A. Braibanti, R. Sambasiva Rao , G. Nageswara Rao and N. Satyanarayna, J *Ind Council. Chem*., 1997, 12, 27-40.
Data, Knowledge and Method bases in Chemistry, II. Data Structures in Chemistry

[123] A. Braibanti, R. Sambasiva Rao, G. Nageswara Rao, N Satyanarayana,*Talanta*, **1998**,47, 109-120.
Data-, Knowledge- and Method- bases in Chemistry I. BAQOR- Database for Equilibrium Constants in Aquo-organic Media,

[124] N. P. Sanchez, A.T. Skeriotis , C.M. Miller, Environ. Sci. Technol., **2014**, 48 (3), pp 1582–1591.
A PARAFAC-Based Long-Term Assessment of DOM in a Multi-Coagulant Drinking Water Treatment Scheme.

[125] Javier, Evrim Acar, Alberto Fernández, Rasmus Bro, Brain Research Bulletin, (Available online 13 May **2015**).
Multiscale entropy analysis of resting-state magnetoencephalogram with tensor factorisations in Alzheimer's disease

Appendix A1:
DataStructures(DS)



$$$Base	
Data	Base
Information	
Knowledge	
Commonsense	
Intelligence	
Method	
Function	

Method_Base (MB)	
ExptDesign	Method base
Expt	

Dataprocessing (MB)	
CurveResolution	Method

Algorithm	
Formulae	Base
Assumptions	

Instrument	
Dataprocessing	

Calibration	base
CruveFitting	
Knowledge_extraction	

Axioms	
FailureCondition	
RemedialMeasure	
EquivalentMethods	
DerivableMethod	
UnsolvedRiddles	
NumericalStability	
TheoriticalProof	

Data.pre_processing	
Scaling	Method base
Signal averaging	
…	

Data.pre_processing Tranformation	
FFT	Method base
..	
Filters	
Kalman	

Data.pre_processing	
Raleigh scattering	Method base

Data structures Numeric formats in MATLAB

Datatype
Numerical
Characters
Strings
abstracts

Data
Structures
Objects
Cells
Graphics

Int\$\$	
int8	
int16	
int32	
int64	
int8	

Floating point
double

Data. **Numerical**		
Real	Integer	Binary [0,1]
		Bipolar [-1,1]
		Multiple [1, 3 ,5,8]
	Floating point [5.123 e-340]	
Imaginary	$a+i*b$	
Quaternary	$a+i*b+j*c+k*d$	

Data. **Characters**
Digits [0:9]
Alphabets [A:Z ; a:z]
Symbols
Graphic characters

Data. **Graphics**	
unit	pixel
	Voxel
2D-image	
3D-image	

MATLAB data structures and interconversion

Data
Structure
Cell

\$\$\$2cell	**Cell2\$\$\$**
mat2cell	cell2mat
Num2cell	
Struct2cell	cell2struc

str2\$\$\$	**Cell2\$\$\$**
str2num	cell2mat
str2double	
	cell2struc

str2\$\$\$	**Cell2\$\$\$**
hex2dec	
hex2num	

	int2\$\$\$	**MatLab**
To convert		
Integer -> string	int2str	str = int2str(N)

\$\$\$2dec	**Cell2\$\$\$**
bin2dec	

\$\$\$2str	**Cell2\$\$\$**
mat2str	

Matrix -> string	mat2str	🟩	str = mat2str(A, n, 'class')	**base2dec**	🟩		int2str	🟩	
							num2str	🟩	

Appendix A2: Multichannel, hyphenated instruments of different orders

The classification of instruments based on tensorial data output per one sample of analysis forms a common base data structures in terms of dimensions, decomposability in to lower order ones and choice of multivariate algorithms for data processing for tasks viz. Curveresolution, calibration, clustering, discrimination, classification etc. Broadly, single sensor instruments of yesteryears have given birth to multi-sensor, hyphenated and higher order ones. The iterative improvements in electronics of instruments, data processing algorithms and necessity in inter disciplinary research fields changed the facet of destructive/non-destructive chemical analysis. It is no more confined to small quality control aloof laboratories, but plays a key role in all phases of life (cradle to post disposal environmental effects) cycle of each and every product in small as well as mega industrial scale.

Order instruments

The tensorial order of data per sample is called order of instrument. Chart 3.zz exemplifies a preliminary view of data structure of popular commercial instruments in vogue for chemometric research.

Order_of_ instrument	**Data tensor per**		**Advantage.chemometric**	**Instance**
	Sample	**# Samples**		
Zeroth	0	1	0	pH meter
First	1	2	1	DAD spectrophotometer
Second	2	3	2	HPLC-DAD
Third	3	4	3	Ex-EM-Fl-Time decay
Fourth	4	5	4	
mth order			Infinite	

Zero order instruments (OI: 1): The single sensor instruments like pH meter, dissolved oxygen meter produce a single datum (zero-order tensor) per sample.

OI : 0	Data 1-way Zero order tensor		
Instrument	Response	Probe	– Yesteryears – Under utilized ✋ Now ✋ First order instrument used as default
Conductometer	Conductance	Pair of Pt plates	
pH meter	pH	Glass electrode	
Ion-selecto meter	pX X : Cu, F^-, CN^-	Selectode	
Piezo-electric crystal	Volatage	Crystal	

First order instruments (OI: 1) : UV-Vis spectrophotometer, chromatographs, NMR instruments measure response at a series of so called channel values resulting in a vector of responses for each sample. A vector, first order tensor or 1-way data is produced per sample, the order of these instruments is one. The data is thus a matrix of dimensions nchannel x 2. The chemometricians found a new outcome of detecting an unknown interferent in the sample and named it as first order advantage.However, the limitation is first order instruments cannot quantify the interferents.

OI : 1	Data			Detector

	2-way First order tensor	
Instrument	X-axis	Response
UV-Vis	Wave-length	Absorbance
IR	Wave number	% transmittance
LC	Elution time	Detector response viz. absorance
NMR	Chemical shift	Intensity
MS	m/e	Intensity

GC	TOFMS
GC	NMR ^{1}H, ^{13}C, ^{15}N, ^{31}P, ^{18}O, ^{19}F
HPLC	Fluorescence
Tetra Hertz spectroscopy	IR NIR MIR FIR

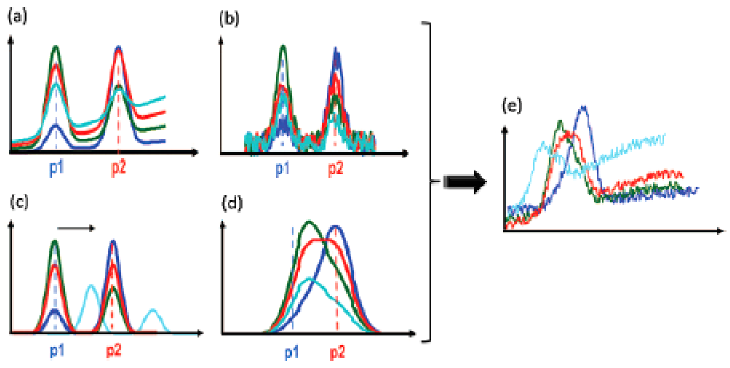
NC. Concentration in test sample
Calibration set is necessary for first order instrument

\+ Detection of interferent compounds

– Analysis of interferents which are not present in calibration set is not possible

▶ **Remedy :** II order instruments

Two-component chromatograph



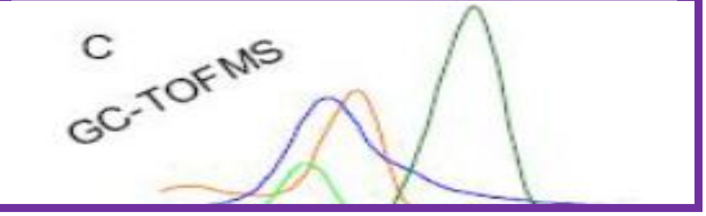
(a) baseline drift, (b) low S/N ratio, (c) RT shift, (d) coleution, (e) combination of all chromatographic challenges

Challenges in chromatography

- Baseline
- Background contributions
- Retention time shifts
- Peak shape changes
  - Coelution or
  - peak overlap (overlapped and embedded peaks).

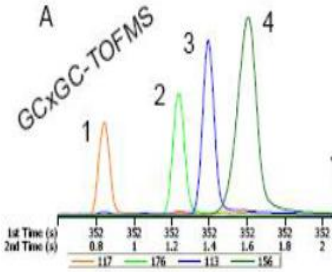
Hyphenated instruments: Herfschfeld reported around sixty viable combinationsof instruments commercially available in 1980. The viable list is long, but now the focus is around attaining lower limits of determination with high accuracy/ reproducibility from instrument point of view.The combination of pH meter, conductometer, DO etc. are routine and they produce a pool of scalar values, but obviously do not have any first order advantage. Further, combination of UV spectrum and IR spectrum also only two matrices of different sizes and can be stored in cell structure. Each profile has its first order advantage alone, but the combination does not produce second order advantage.

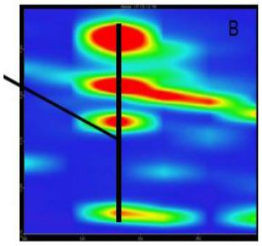
Hyphenated (Two) instruments	
GC-MS	2D-NMR J-coupling NMR
LC-MS	
LC-DAD	LC-LC
LC-NMR	LC-CE
2D-LC-Resp	CE-CE



HPLC-DAD	LC-CG	
LV-FTIR	GC-GC	

Hyphenated (Three) instruments	
LCxLC-DAD	GCxGC-FID
GC×GC-MS	GCxGC-ECD
LCxLC-MS	LCxLC-UV
GC×GC-TOFMS	
LCxLC-TOFMS	





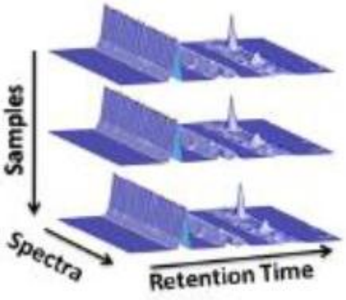
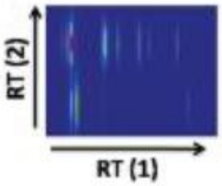
Higher(>1) order instruments

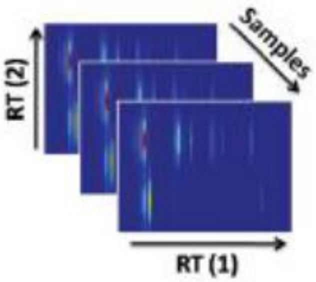
Chemical analysis entered a new era with higher (especially second and third) order instruments. Not only S/N ratio dramatically improved, but also estimation of multiple analytes in presence of uncalibrated interferents and matrices.What was not possible with classical destructive separation techniques, has become a child play in different dimensions of hyphenated instruments.

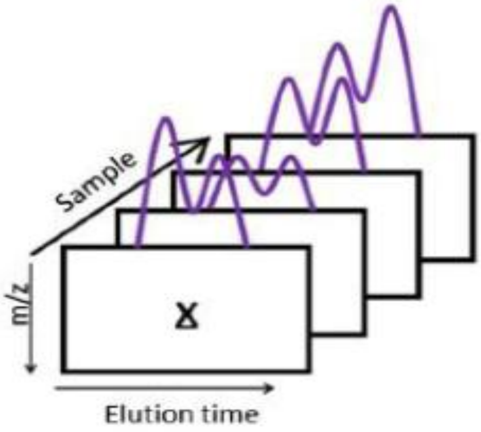
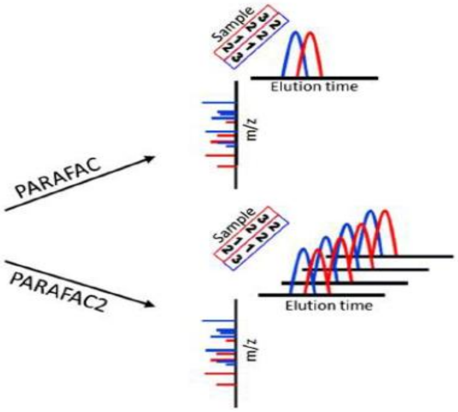
Second order instruments (OI: 2): If HPLC is hyphenated with DAD spectrophotometer, first order full spectrum is obtained at each elution time. The output is a matrix for a sample and it is an example of second order instrument.

Second order Instruments	**X**			**Data tensor per**		
			y(Response)	**Sample**	**#Sample**	**advantage**
Ex-Em-Fluorescence	Excitation-wave length	Emission-wave length	Absorbance	3	4	Second

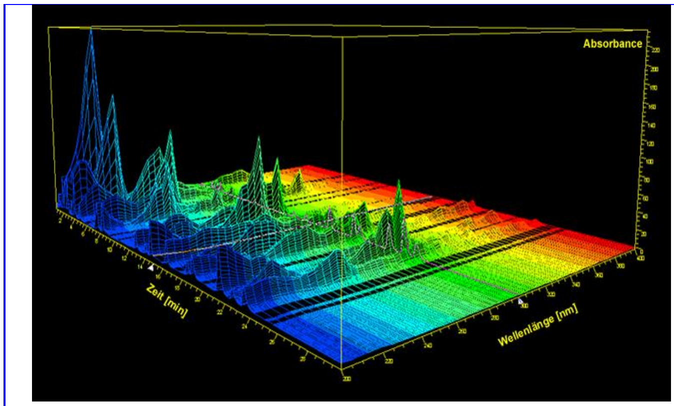
OI=2		#samples with Multi-channel
		DO = IO +NVar = 2+1 =3

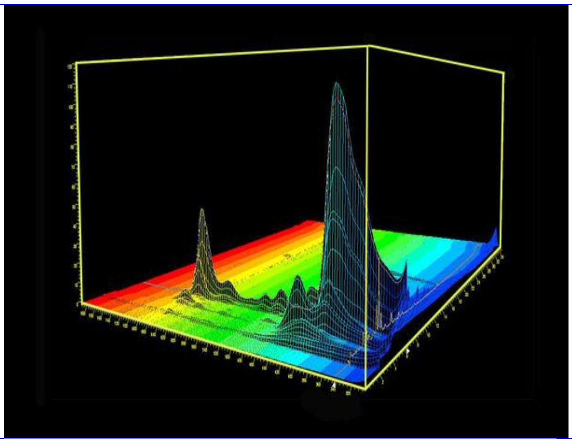
 Matrix Size: [RT, #wavelengths]		 Set of matrices or Tensor of third order Size: [RT, #wavelengths, #samples]
 Matrix		

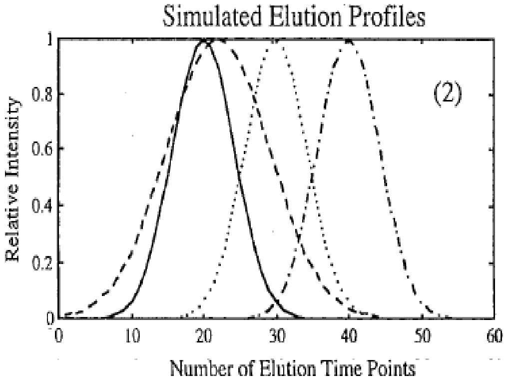
Size: [RT(1), RT(1)]		 Set of matrices or Tensor of third order Size: [RT(1), RT(1), #samples]

LC-MS	**PARAFAC model**
	

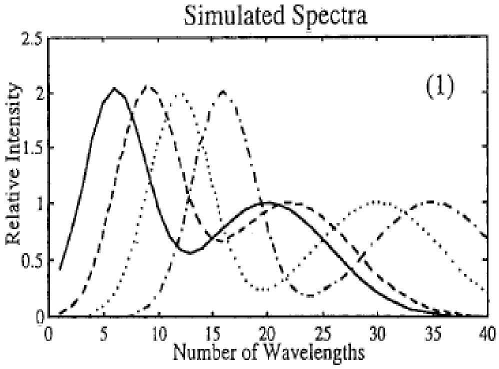
Excitation-emission-fluorescence data matrix	

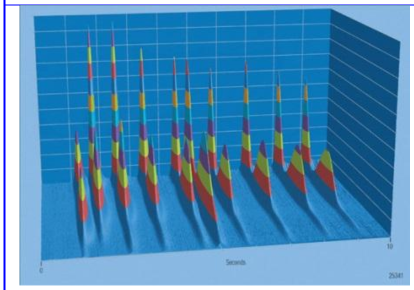




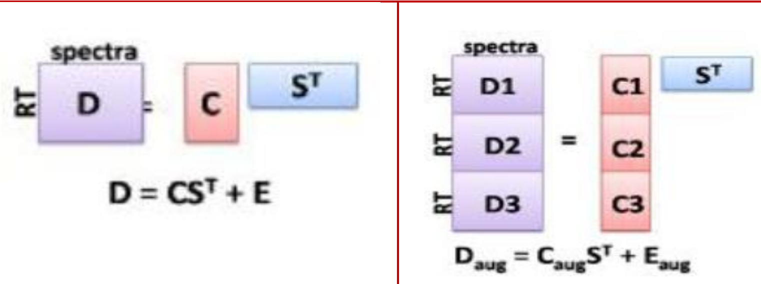


Courtesy of Ref: Fxyz-22





Assumption : If Resp = LinFn(C,S)



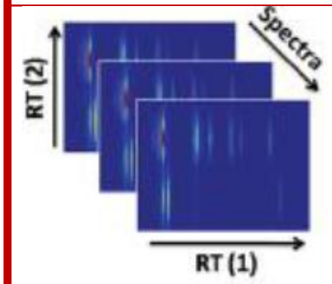
$D = CS^T + E$

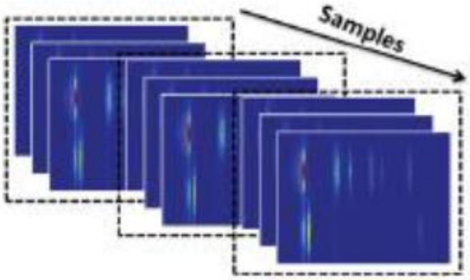
$D_{aug} = C_{aug}S^T + E_{aug}$

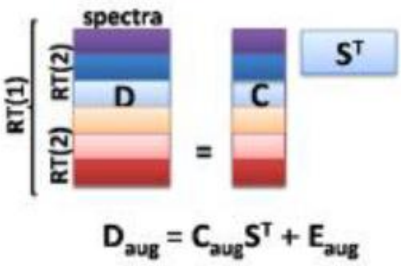
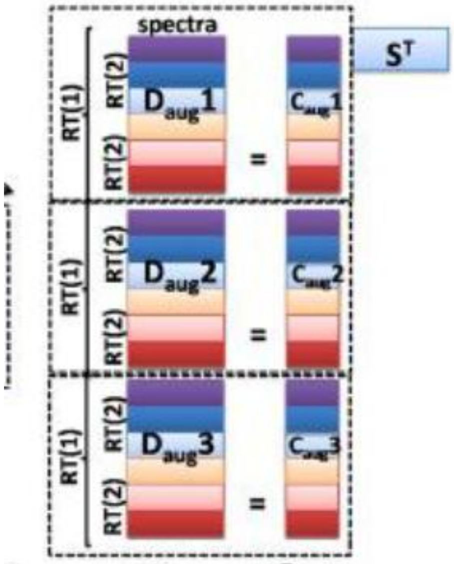
Third order instruments (OI: 3)

Third order Instruments	X				Data tensor per		
				y(Response)	Sample	#Sample	advantage
Ex-Em- Time delay Fluorescence	Excitation-wave length	Emission-wave length	Time	Absorbance Fluorescence	4	5	Second
Phosphorescence EEM-time decay	Excitation-wave	Emission-wave	Time	Phosphorescence			

	length	length					

<table>
<tr><td>OI=3</td><td></td><td>#samples with Multi-channel
DO = IO +NVar = 3+1 =4</td></tr>
<tr><td>

Tensor of third order
Size: [RT(1), RT(1), #wavelengths #]</td><td></td><td>

Tensor of Fourth order
Size: [RT(1), RT(1), #wavelengths #samples]</td></tr>
<tr><td colspan="3">Assumption : If Resp = LinFn(C,S)</td></tr>
<tr><td>
</td><td></td><td>
</td></tr>
</table>

Fourth order instruments (OI: 4)

Fourth order Instruments	**X**					**Data tensor per**	
					y(Response)	**Sample**	**#Sample**
Second-order liquid chromatographic data with UV–vis fluorescence	Liquid chromatographic Elution-1	Elution-2	UV–vis-	Fluorescence	Absorbance	5	6

Concatenation of instrumental data

The concatenation of instrumental data for a number of samples (from different sites, time period, manufacturers), different experimental influential factors (concentration, compounds, pH) or physico-chemical variables (Temp., ionic strength, dielectric constant) results in multi-way tensorial data.

Order of instruments		Concatenation		Data			
Zero	Zero	Vertical		Horizontal			
pH meter	Conductometer	[1; 2]		[1,2]			No advantage
First	First	Vertical		horizontal			
DAT	IR	[DAD; IR]		[DAD, IR]	*Matrix not possible ,as vector sizes are different Cell structure		No advantage

Data fusion: Borràs et al. [2] reviewedthe advantages of populardata fusion strategies that are in practice in authentication, calibration, quality assessment and adulteration probing of food and beverages.

Fusion of Data, information, parameters and knowledge

Hyphenated instruments

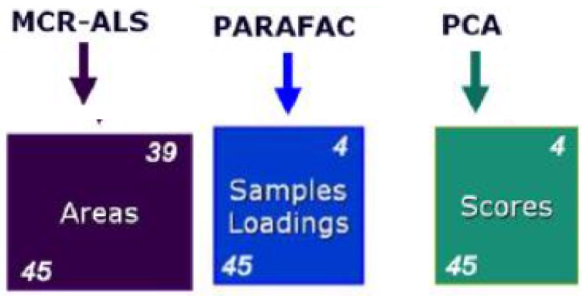
Hyp.Inst

HPLC-DAD **EEM** **^{1}H-NMR**

Primary data

Data Fusion

Models



EDA
CDA

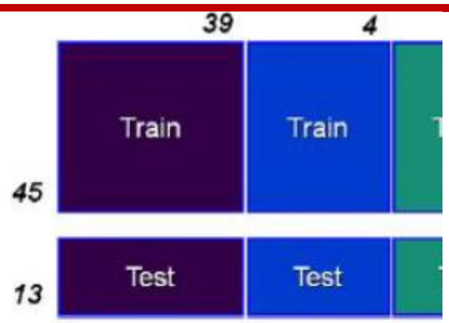
Modeling
Validation
Prediction

Ensemble
Forest

Parameters **Parameter fusion**

Information **Information fusion**

Knowledge **Knowledge Fusion**



Experimental influencing Variables

Causative/influencing/independent Experimental variables	
Spectrophotometry	pH, Reaction time, concentration of ingredients
Flow injection analysis	Solvent flow rate, length, diameter of tube
Chromatography	Columns, Oven temperature, carrier gas concentration
Equilibrium constant/ Rate constant	Concentration ofmetal, ligand, mineral acid, co-solvents, micelle, concentration, Temperature
	Discipline wise
log K log k	Temperature, ionic strength, dielectric constant, hydro phobicity, molecular descriptors
Kovat index Log Capacity factor	Molecular refractivity, dipole moment, hydrogen bond capacity, basicity of solute, molecular descriptors
Environment	Sites, Compound

OI	Primary. Inst.Data	m-way if Expt.variables	
0	1 if A	2-way if A1	3-way if A2
1	2 if A	3way if A1	4-way if A2
2	3 if A	4way if A1	5-way if A2
3	4 if A	5way if A1	6-way if A2
4	5if A	6way if A1	7-way if A2

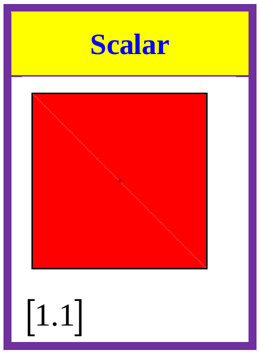
A	if	different samples	
A1	if	different samples & [one] Exp_var	[pH]
A2	if	different samples & [Two] Exp_var	[pH; dielectric constant]
A3	if	different samples & [Three] Exp_var	[pH; dielectric constant; ionic strength]

mth order tensorial Data generation

Scalar: [Tensor0: data(1,1,1), or zero-way] data

Physico-chemical-biological constants

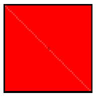
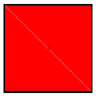
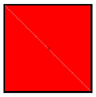
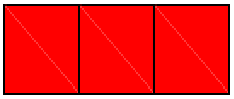
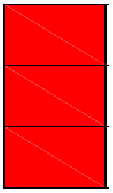
The gravitational constant, Avogadro number, Plank's constant, atomic number of a chemical element or Boltzmann constant, is a single real numerical value; may be integer or floating point number. Each value is unique and they are examples of scalars. In the tensorial notation a scalar is a zero order tensor. In data processing terminology of applied sciences a scalar is a zero-way data.



Development of higher order tensors from X and/or Y

If	Sizes of column vectors are same size
Then	Folding (horizontal fusion) into a matrix or arrays

If	Sizes of column vectors are of different sizes
Then	No folding &
	Stored in a cell
If	Data types are of different
Then	structure

Development of higher order tensors from X and/or Y						
Scalars					-->	Vector
 $[1.1]$	+	 $[1.2]$	+	 $[1.3]$	... =	 $[1.1\quad 1.2\quad 1.3]$
		$[1.1\quad 1.2\quad 1.3]^T = \begin{bmatrix}1.1\\2.1\\3.1\end{bmatrix}$				

1-way tensor:
Vector: [row: Tensor1: data (:, 1,1) or col:Tensor1: data(1,:,1)]

```
%
% vecdev.m
% Development of vectors from scalars
% i.e. first order tensor from zero
% order tensor

clean
 a = 1,b=2,c=3
vecrow = [a b c]
veccol= [a;b;c;]

 pause
a(1) = 1;
 a(2) = 2;
 a(3) = 3;
 a(1)
 a(2)
 a(3)
vecrow = [a(1) a(2) a(3)]
veccol= [a(1); a(2); a(3)]
 pause
fori = 1:3
    a(i,1) = i;
end
 a(1)
```

```
a =
     1
b =
     2
c =
     3
vecrow =
     1     2     3
veccol =
     1
     2
     3

ans =
     1
ans =
     2
ans =
     3
vecrow =
     1     2     3
veccol =
     1
     2
     3
```

```
 a(2)
 a(3)
vecrow = [a(1) a(2) a(3)]
veccol= [a(1); a(2); a(3)]
```

```
ans =
     1
ans =
     2
ans =
     3
vecrow =
     1     2     3
veccol =
     1
     2
     3
```

2-way tensor

The variation of the response (dependent variable) as a function of only one explanatory/causative (independent) variable is described by a conventional two dimensional figure. Here the two basis vectors are orthogonal. In the case of a function with several variables, experiments are performed wherein one variables, experiments are performed wherein one variable is varied keeping all the others at fixed (optimum or non-optimum) values. These are called one variable at a time (OVAT) experiments.

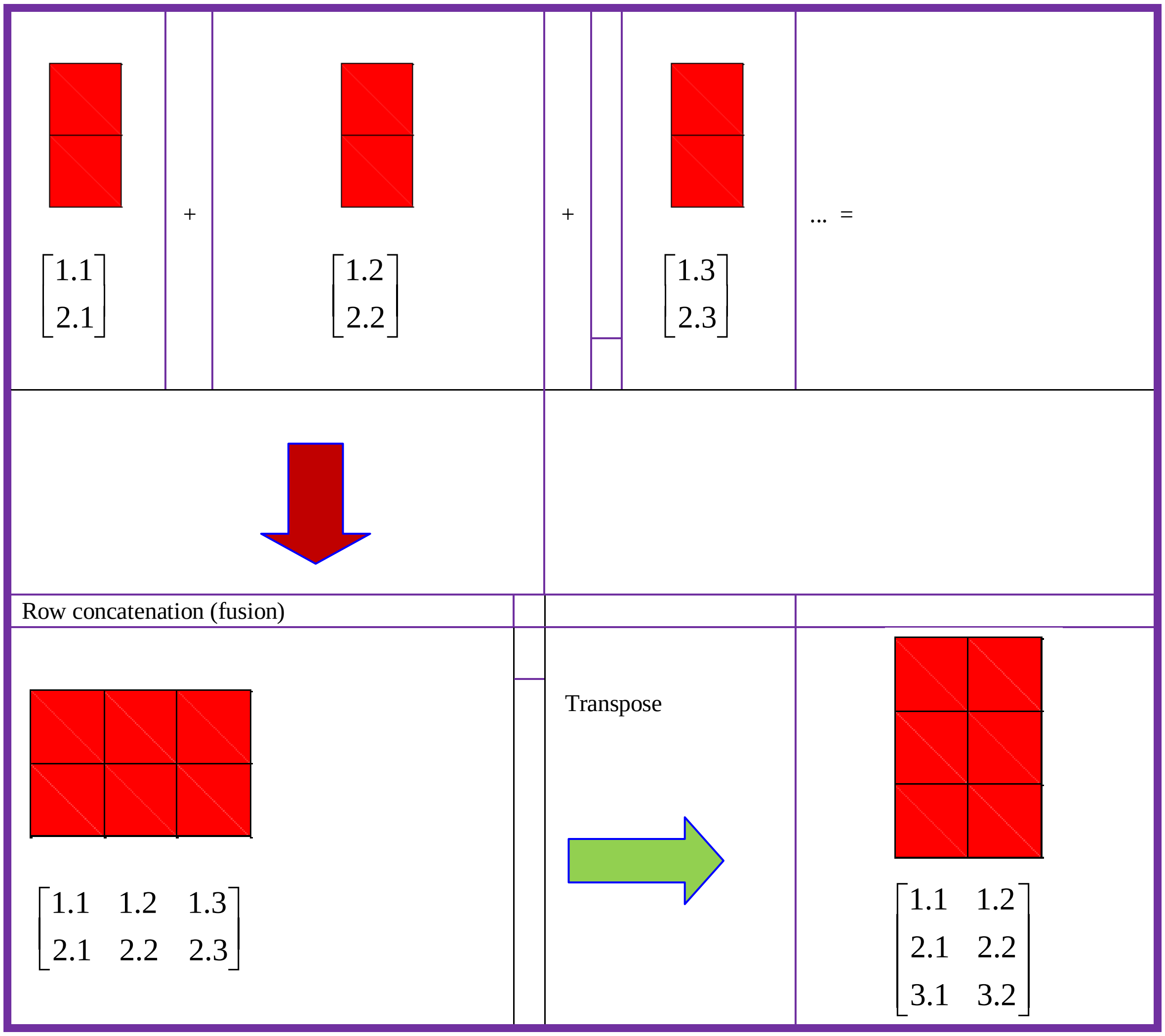
KB. Zz:

If	Sizes of column vectors are same size
Then	Folding (horizontal fusion) into a matrix or arrays
If	Sizes of column vectors are of different sizes
Then	No folding & Stored in a cell
If	Data types are of different
Then	Structure or cell

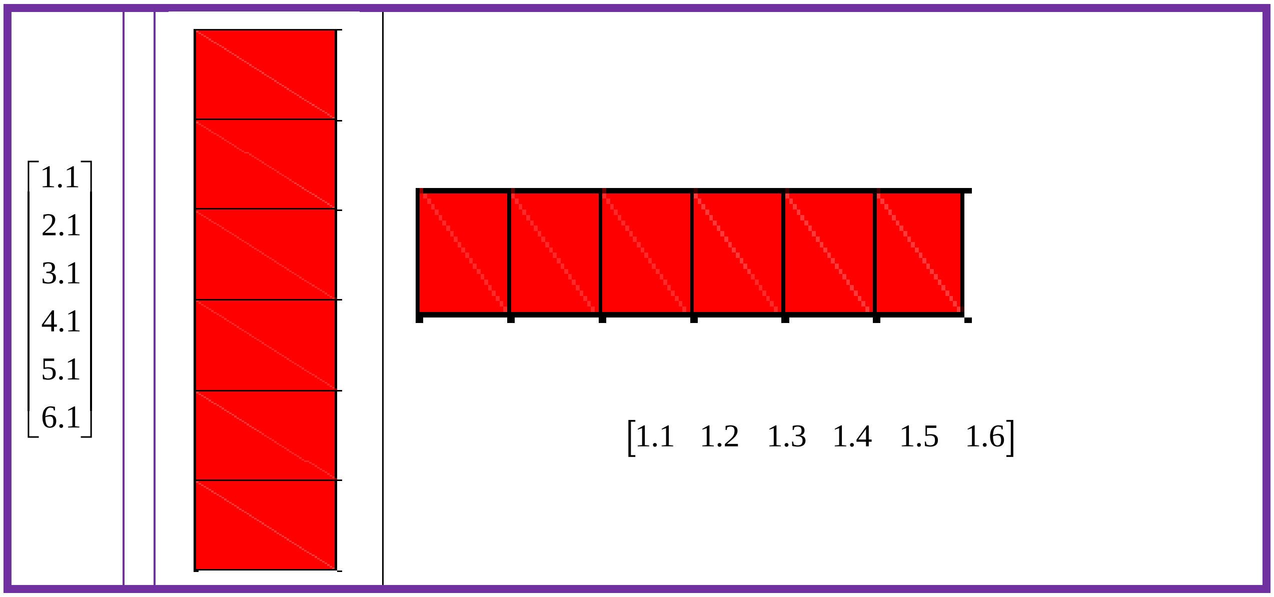
If	IR spectrum & DAD spectrum
Then	Matrices of different sizes&
	No 3-way structure & Cell structure [At each wave number of IR, no possibility for a UV spectrum]
If	HPLC & DAD spectrum
Then	2way- instrumental data [**Explanation:** At each elution time, a DAD spectrum is available]

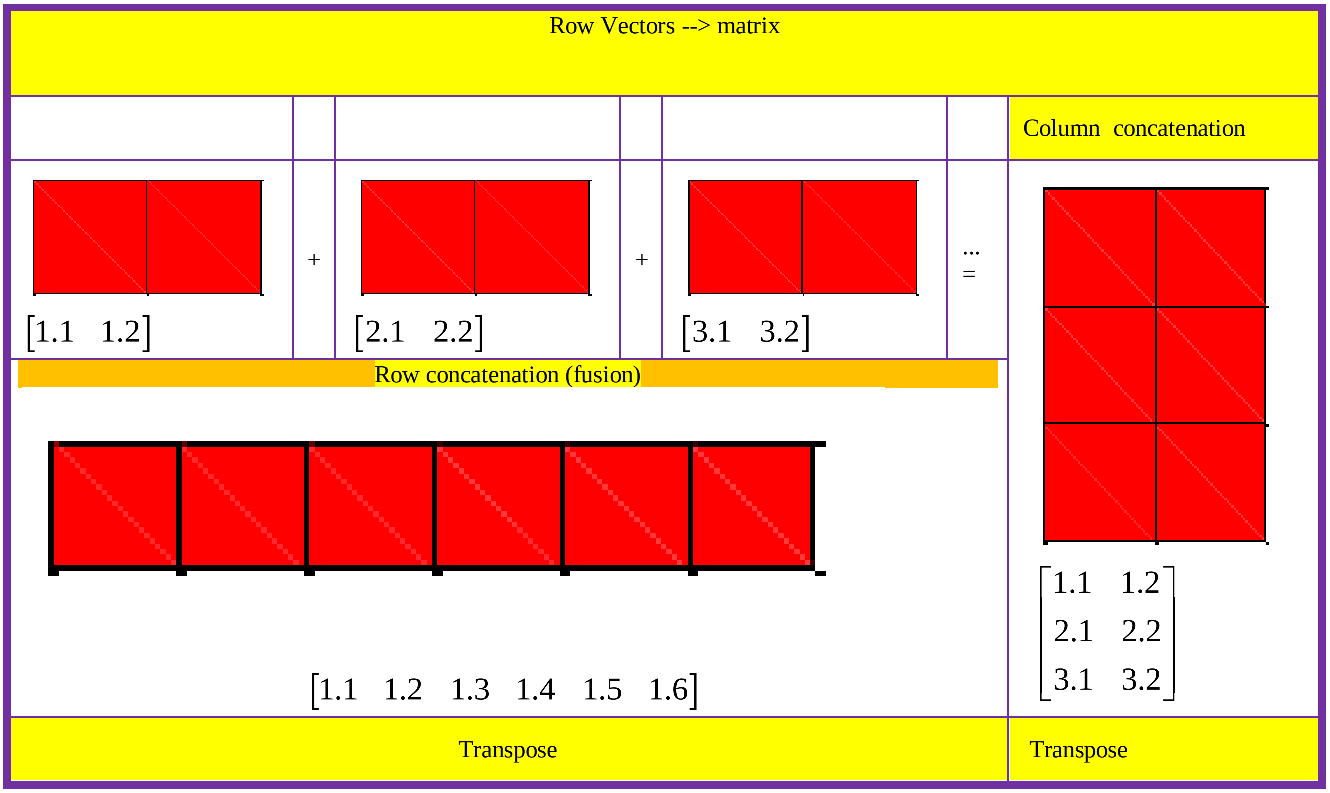
If	HPLC & DAD spectrum & Number of samples (ns)
Then	k- number of 2way-data structure & fold into 3-way data structure [**Explanation:** horizontal or vertical fusion is not more informative]

	Colum Vectors --> matrix	

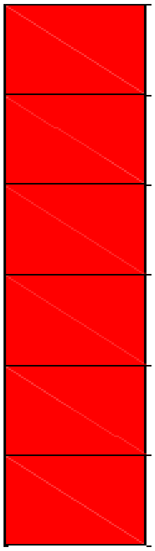


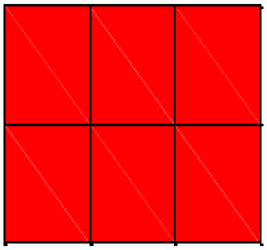
		Colum concatenation	Transpose


$\begin{bmatrix}1.1\\2.1\\3.1\\4.1\\5.1\\6.1\end{bmatrix}$
$\begin{bmatrix}1.1 & 1.2 & 1.3 & 1.4 & 1.5 & 1.6\end{bmatrix}$


Row Vectors --> matrix
Column concatenation
$\begin{bmatrix}1.1 & 1.2\end{bmatrix}$
+
$\begin{bmatrix}2.1 & 2.2\end{bmatrix}$
+
$\begin{bmatrix}3.1 & 3.2\end{bmatrix}$
...
=
$\begin{bmatrix}1.1 & 1.2\\2.1 & 2.2\\3.1 & 3.2\end{bmatrix}$
Row concatenation (fusion)
$\begin{bmatrix}1.1 & 1.2 & 1.3 & 1.4 & 1.5 & 1.6\end{bmatrix}$
Transpose
Transpose

$$\begin{bmatrix} 1.1 \\ 2.1 \\ 3.1 \\ 4.1 \\ 5.1 \\ 6.1 \end{bmatrix}$$





$$\begin{bmatrix} 1.1 & 1.2 \\ 2.1 & 2.2 \\ 3.1 & 3.2 \end{bmatrix}$$

Examples

y = fnLin(x;par)		$\text{Resp} = \begin{bmatrix} one & x \end{bmatrix} * par$ Eqn. z.xx
	x	**y**
Univariate calibration	Concentration of analyte	Absorbance at landamax
Vant Hoff equation	Temperature (^{o}K)	log(rate constant) log(equilibrium constant)

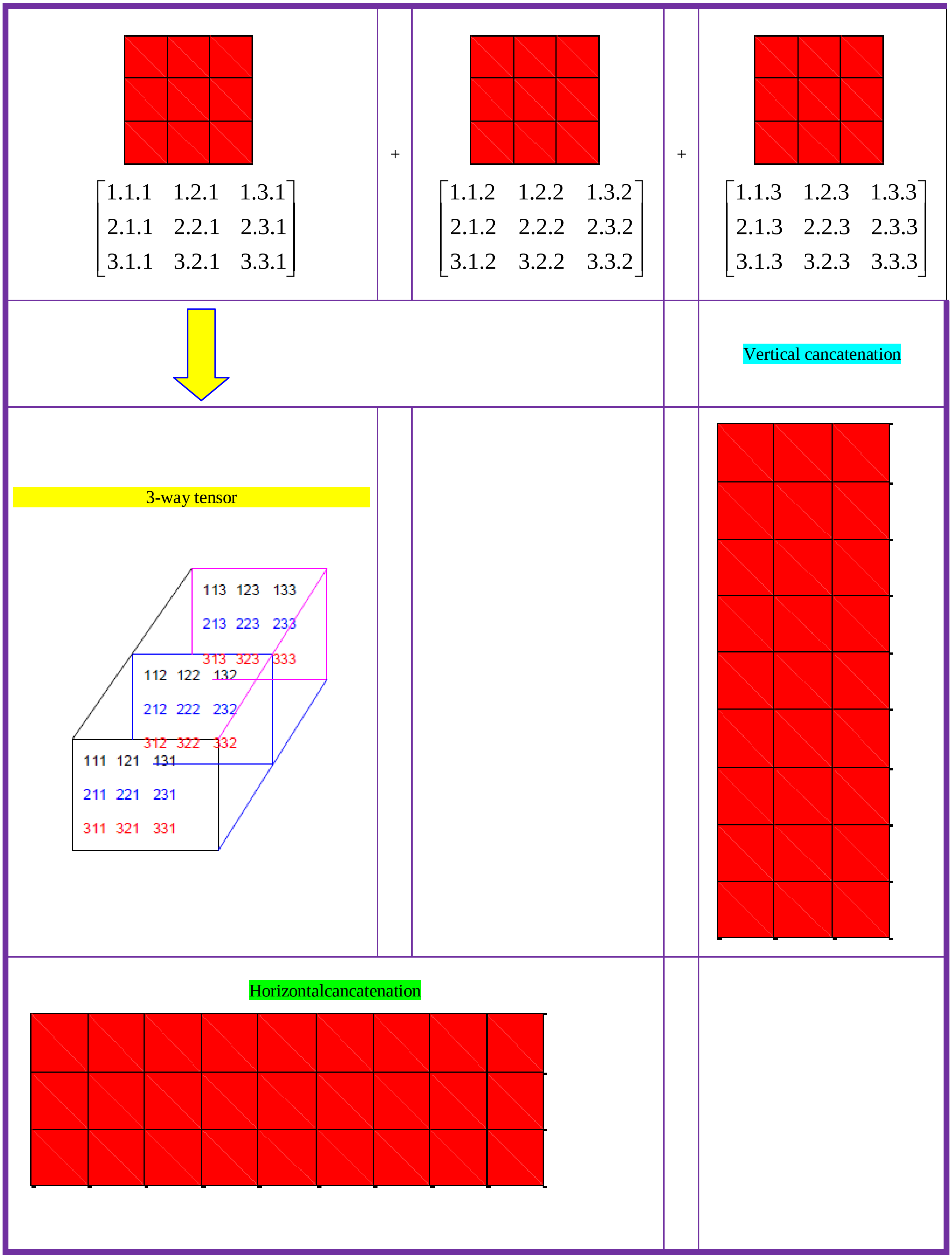
X (:,2)		**[x1,x2]**
	x1	**x2**
Two explanatory variables		
	Temperature	Dielectric constant
	Ionic strength	Dielectric constant

Y (:,2)		**[y1, y2]**
Two responses	**y1**	**y2**
	Reaction time for maximum rate	Maximum yield

3-way tensor:

If	Ex-Em-TimeDelay fluorescence
Then	3-way instrumental data
If	Ex-Em-TimeDelay fluorescence & Number of samples (ns)
Then	4-way data

Matrices --> 3-way tensor



$$\begin{bmatrix} 1.1.1 & 1.2.1 & 1.3.1 \\ 2.1.1 & 2.2.1 & 2.3.1 \\ 3.1.1 & 3.2.1 & 3.3.1 \end{bmatrix} + \begin{bmatrix} 1.1.2 & 1.2.2 & 1.3.2 \\ 2.1.2 & 2.2.2 & 2.3.2 \\ 3.1.2 & 3.2.2 & 3.3.2 \end{bmatrix} + \begin{bmatrix} 1.1.3 & 1.2.3 & 1.3.3 \\ 2.1.3 & 2.2.3 & 2.3.3 \\ 3.1.3 & 3.2.3 & 3.3.3 \end{bmatrix}$$

```
%
% threeway.m        R Naga satish (09/14/2013  03:25 AM;
%                          05/19/2001  08:38 PM)
%     HOT(Hands on tutorial) version
%     not optimum code, No_ knowledge_ base

% Housekeeping
clean
cla
  grid off, box off


 %% display of elements of matrix
%   Layer 1

  subplot(221),axis([0 1 0 1]);,
   text(0.05,0.46,  '111  121   131  ') ,
   text(0.05,0.38, '211  221   231  ', 'color','b');
   text(0.05,0.3,  '311  321   331  ','color','r');
   hold on
   plot([0.03 0.03 0.25   0.25 0.03],[0.25 0.51 0.51 0.25 0.25] ,'k-')
   pause,

%   Layer 2
    st4 = '112  122   132';
   st5 = '212  222   232  ';
   st6 = '312  322   332  ';
  subplot(222),axis([0 1 0 1]);,
   text(0.16,0.66,st4);
   text(0.16,0.58,st5,'color','b')
   text(0.16,0.5,st6,'color','r');
   hold on
   plot([0.14 0.14 0.4 0.4 0.18],[0.51 0.71 0.71 0.45 0.45],'b-')
   pause,

%   Layer 3
   st7 = '113  123   133';
   st8 = '213  223   233  ';
   st9 = '313  323   333  ';
   subplot(223),axis([0 1 0 1]);
   text(0.27,0.86,st7)
   text(0.27,0.78,st8,'color','b')
   text(0.27,0.7,st9,'color','r')
   hold on
    plot([0.25 0.25 0.5 0.5 0.29],[0.71 0.91 0.91 0.65 0.65],'m-')
   pause
                                                    %%
%%
% Drawing boxes
  subplot(224),  axis([0 1 0 1]);hold on,
  plot([0.03 0.25],[0.51 0.91],'k--'),pause
  plot([0.3 0.5],[0.51 0.91],'m--'),pause
  plot([0.3 0.5],[0.25 0.65],'b--'),pause

% Housekeeping
  clean
cla
  grid off, box off
  axis([0 1 0 1]);

 %% display of elements of matrix
```

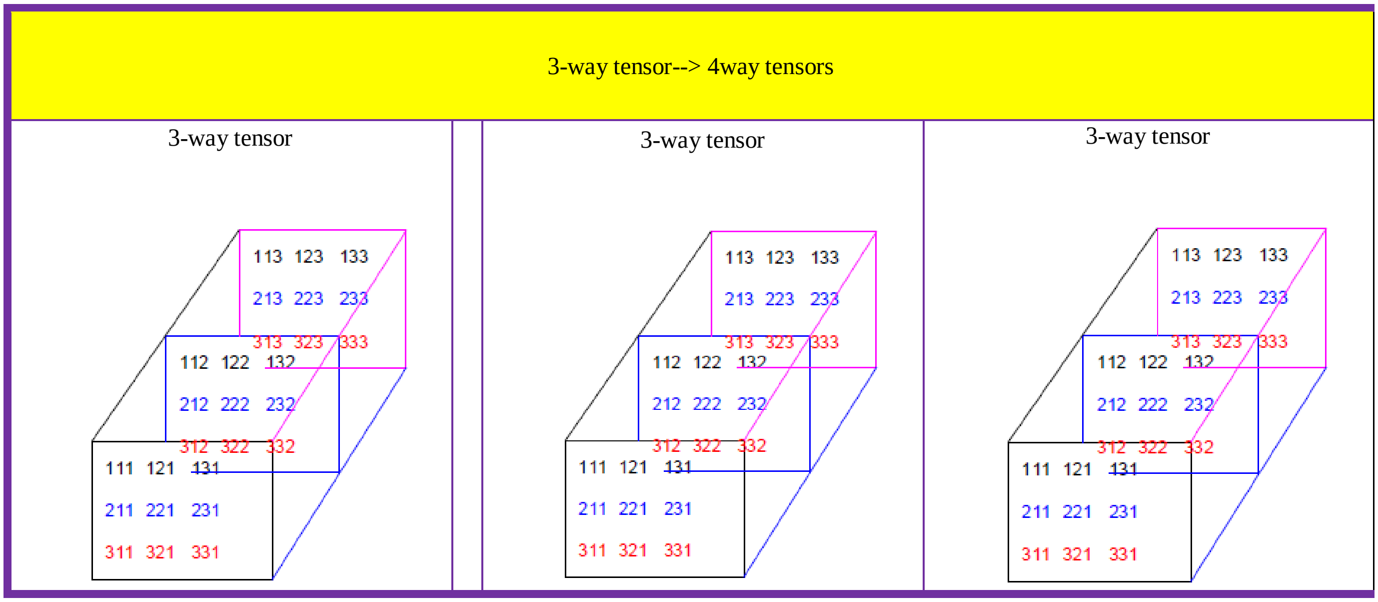
```
%   Layer 1
  hold on
   text(0.05,0.46,  '111  121   131  ') ,
   text(0.05,0.38, '211  221   231  ', 'color','b');
   text(0.05,0.3,  '311  321   331  ','color','r');
pause,plot([0.03 0.03 0.3   0.3 0.03],[0.25 0.51 0.51 0.25 0.25] ,'k-')

%   Layer 2
   st4 = '112  122   132';
   st5 = '212  222   232  ';
   st6 = '312  322   332  ';
pause,text(0.16,0.66,st4);
   text(0.16,0.58,st5,'color','b')
   text(0.16,0.5,st6,'color','r');
   pause, plot([0.14 0.14 0.4 0.4 0.18],[0.51 0.71 0.71 0.45 0.45],'b-')

%   Layer 3
   st7 = '113  123   133';
   st8 = '213  223   233  ';
   st9 = '313  323   333  ';
  pause, text(0.27,0.86,st7)
   text(0.27,0.78,st8,'color','b')
   text(0.27,0.7,st9,'color','r')
pause,plot([0.25 0.25 0.5 0.5 0.29],[0.71 0.91 0.91 0.65 0.65],'m-')
                              %%
% Connecting three matrices into 3-way-tube

  pause, plot([0.03 0.25],[0.51 0.91],'k--')
  pause, plot([0.3 0.5],[0.51 0.91],'m--')
  pause, plot([0.3 0.5],[0.25 0.65],'b--')
     axis off
% plot([0.03 0.25],[0.51 0.91],'k--')
% pause, plot([0.25 0.45],[0.51 0.91],'m--')

% pause, plot([0.25 0.45],[0.25 0.65],'b--')

   axis off

                                         %%
% Interactive position setting of titles of layers
break
gtext('Layer I')
gtext('Layer II')
gtext('Layer III'),
gtext('Rows'),gtext('Colums'),gtext('Tube')
       %%
  hold off
```

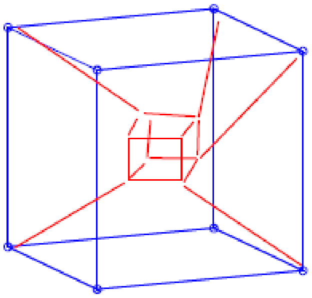
X = tensor(A,SIZ)	creates a tensor from the multidimensional array A. The SIZ argument specifies the desired shape of A.
Ex.: Z = tensor(rand(a*b*c,1),[a b c]) Z is a tensor of size 3 x 2 x 2	Z(:,:,1) = 0.8173 0.3998 0.8687 0.2599 0.0844 0.8001 Z(:,:,2) = 0.4314 0.2638 0.9106 0.1455 0.1818 0.1361

	X = tensor(rand(3,4,2))	X is a tensor of size 3 x 4 x 2 X(:,:,1) = 0.8147 0.9134 0.2785 0.9649 0.9058 0.6324 0.5469 0.1576 0.1270 0.0975 0.9575 0.9706 X(:,:,2) = 0.9572 0.1419 0.7922 0.0357 0.4854 0.4218 0.9595 0.8491 0.8003 0.9157 0.6557 0.9340
	n = 4; Y = tensor(rand(n,1),n)	Y is a tensor of size 4 Y(:) = 0.0318 0.2769 0.0462 0.0971
	row=12; w1= 2;w2=3;w3=2; Z = tensor(rand(row,1),[w1 w2 w3])	Z is a tensor of size 2 x 3 x 2 Z(:,:,1) = 0.8235 0.3171 0.0344 0.6948 0.9502 0.4387 Z(:,:,2) = 0.3816 0.7952 0.4898 0.7655 0.1869 0.4456

Four-way data are created by joining the set of third-order data arrays for the calibration samples and for each of the analyzed test samples. Application of the PARAFAC model to the latter four-way data arrays requires fitting the following expression:

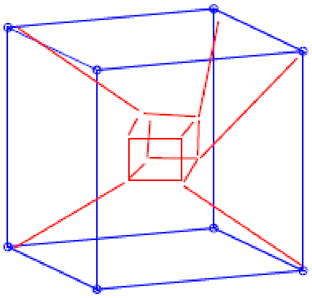
$$X_{ijkl} \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} d_{ln} + E_{ijkl}$$



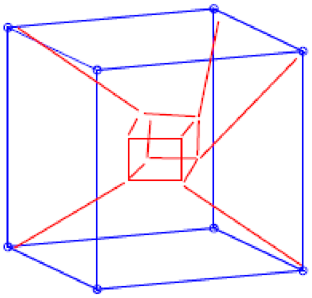


```
a(:,:,1,1) =
   4    5    6
   5    6    7
   6    7    8
   7    8    9
a(:,:,2,1) =
   5    6    7
   6    7    8
   7    8    9
   8    9   10
a(:,:,1,2) =
   5    6    7
   6    7    8
   7    8    9
   8    9   10
a(:,:,2,2) =
   6    7    8
   7    8    9
   8    9   10
   9   10   11
```

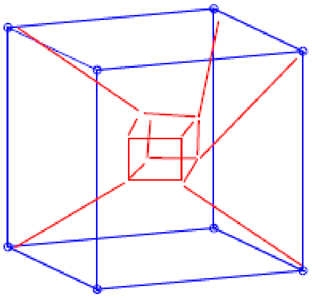
4-way tensors --> 5way tensors



+



+

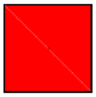
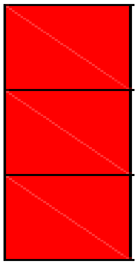
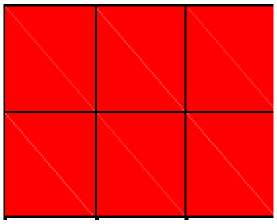
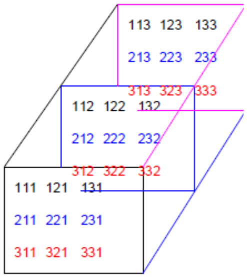
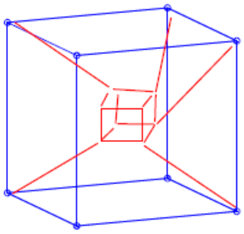


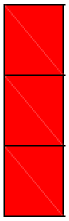
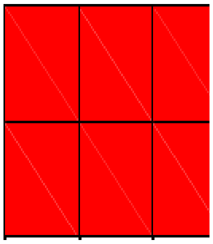
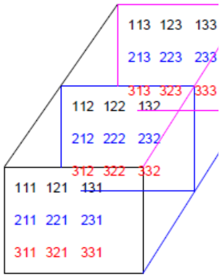
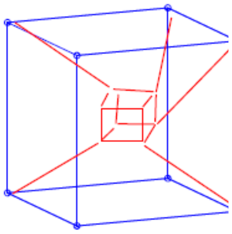
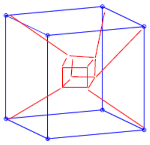
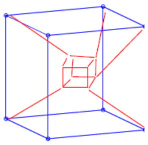
```
a(:,:,1,1,1) =
    5    6    7
    6    7    8
    7    8    9
    8    9   10
a(:,:,2,1,1) =
    6    7    8
    7    8    9
    8    9   10
    9   10   11
a(:,:,1,2,1) =
    6    7    8
    7    8    9
    8    9   10
    9   10   11
```

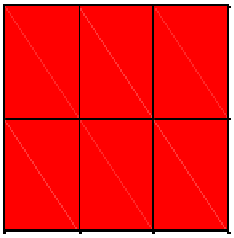
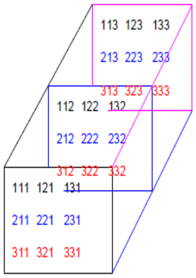
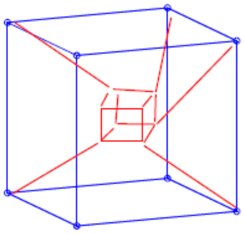
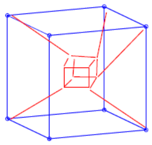
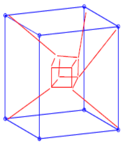
```
a(:,:,2,2,1) =
    7    8    9
    8    9   10
    9   10   11
   10   11   12
a(:,:,1,1,2) =
    6    7    8
    7    8    9
    8    9   10
    9   10   11
a(:,:,2,1,2) =
    7    8    9
    8    9   10
    9   10   11
   10   11   12
```

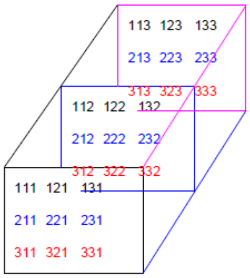
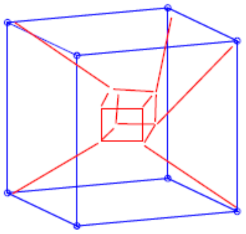
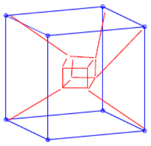
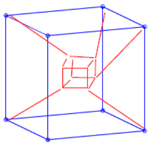
```
a(:,:,1,2,2) =
    7    8    9
    8    9   10
    9   10   11
   10   11   12
a(:,:,2,2,2) =
    8    9   10
    9   10   11
   10   11   12
   11   12   13
```

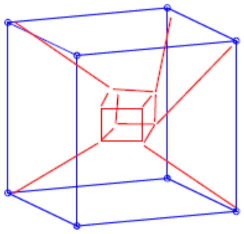
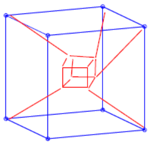
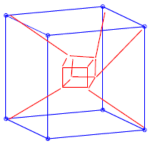
a(1,1,1) = 1.1.1 a(1,1,2) = 1.1.2 a(1,2,1) = 1.2.1 a(1,2,2) = 1.2.2 a(1,3,1) = 1.3.1 a(1,3,2) = 1.3.2	a(2,1,1) = 2.1.1 a(2,1,2) = 2.1.2 a(2,2,1) = 2.2.1 a(2,2,2) = 2.2.2 a(2,3,1) = 2.3.1 a(2,3,2) = 2.3.2	a(3,1,1) = 3.1.1 a(3,1,2) = 3.1.2 a(3,2,1) = 3.2.1 a(3,2,2) = 3.2.2 a(3,3,1) = 3.3.1 a(3,3,2) = 3.3.2	a(4,1,1) = 4.1.1 a(4,1,2) = 4.1.2 a(4,2,1) = 4.2.1 a(4,2,2) = 4.2.2 a(4,3,1) = 4.3.1 a(4,3,2) = 4.3.2

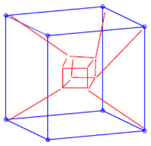
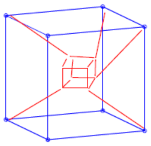
	m-way tensor				
	Variables				
NV -->	**1**	**2**	**3**	**4**	**5**
OI=0					
m-way Data OI+NV	**1**	**2**	**3**	**4**	**5**

NV-->	**1**	**2**	**3**	**4**	**5**
OI=1					 
OI+NV-1	**1-way**	**2-way**	**3-way**	**4-way**	**5-way**

NV-->	1	2	3	4	5
OI=2				 	5-way
OI+NV-1	2-way	3-way	4-way	5-way	6-way

NV-->	1	2	3	4	5
OI=3			 	5-way	6-way
OI+NV-1	3-way	4-way	5-way	6-way	7-way

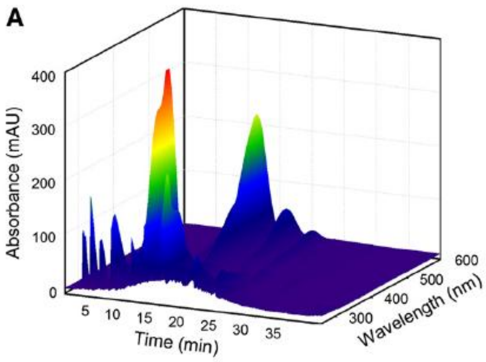
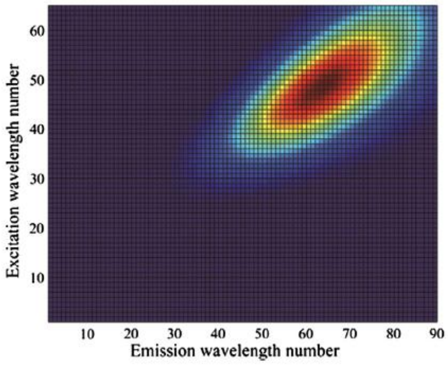
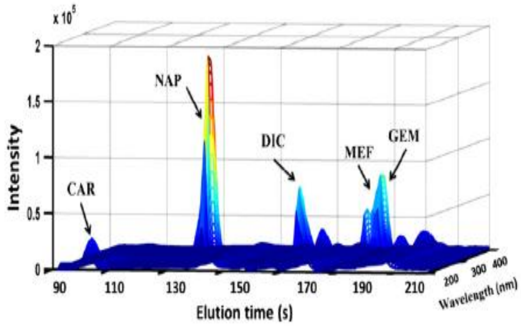
NV-->	1	2		3	4	5
OI=4				6	7	8
OI+NV-1	4-way	5-way		6-way	7-way	8-way

NV-->	1		2	3	4	5
OI=5			6-way	7-way	8-way	9-way
OI+NV-1	5-way		6-way	7-way	8-way	9-way

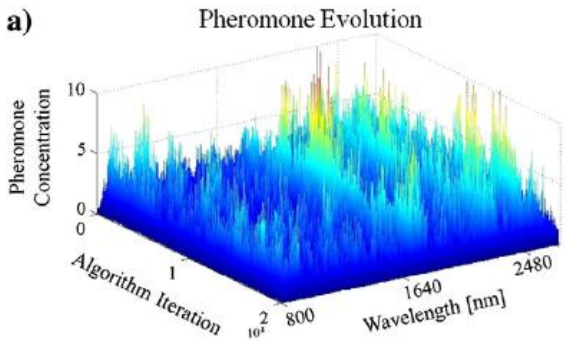
NV-->	1	2	3	4	5
OI=6	6	7	8	9-way	10way
OI+NV	6	7	8	9-way	10way

Powerful information extraction from instruments + experimental variables + chemometric methods					
Data	Higher order instrumental data	+	Number of experimental variables	+	Chemometric methods
	GC-GC-MS		N- samples		Multivariate curve resolution methods
					Multivariate calibration methods
	LC-LC-MS		N- samples		
m-Way	3		1	4	Second order advantage Third order advantage

Appendix A3:
3D-surfaces and 2D-contours of 3-way data-tensors from hyphenated instruments

Chromatographic-wavelength landscape	Non-trilinear background
 Courtesy of Ref: 44	 Courtesy of Ref: 43
HPLC–DAD for five selected pharmaceuticals30.0 ng mL^{-1}  Courtesy of Ref: 42	

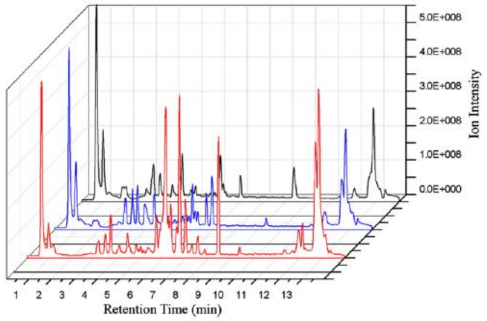
Protein prediction with NIR



Courtesy of Ref: 41

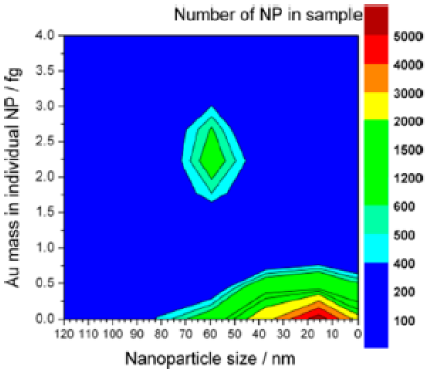
Ion chromatograms of urine samples

black line: first day; blue line: second day; red line: third day



Courtesy of Ref: 47

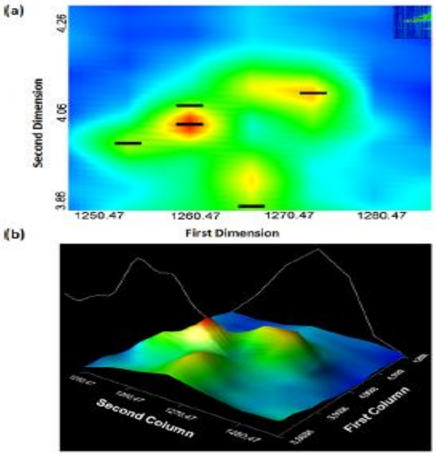
3D- hydrodynamic chromatogram
gold nanoparticles



Courtesy of Ref: 18

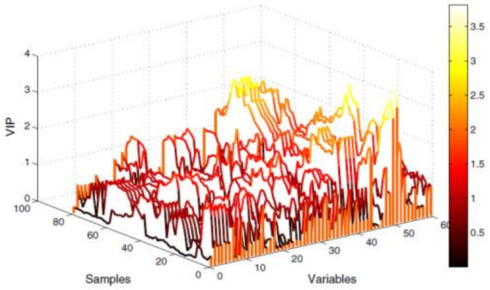
GC × GC-TOFMS tensor for North Sea oil

MCR-ALS

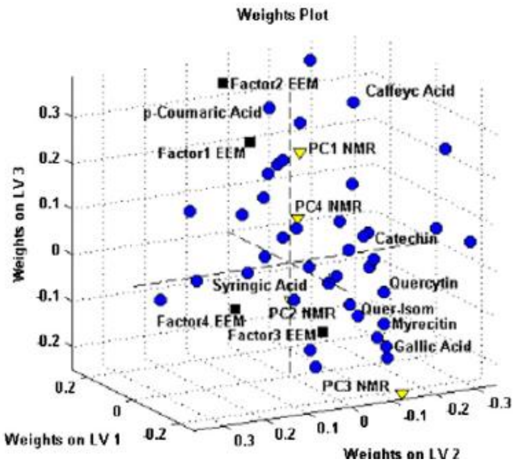
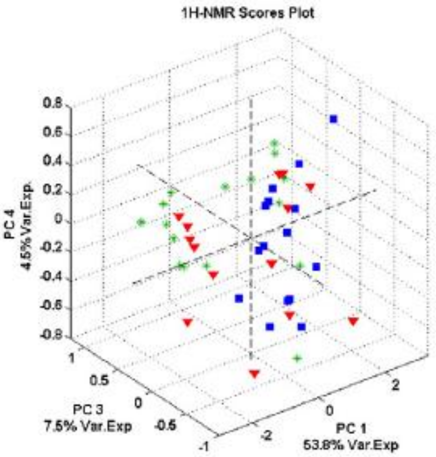


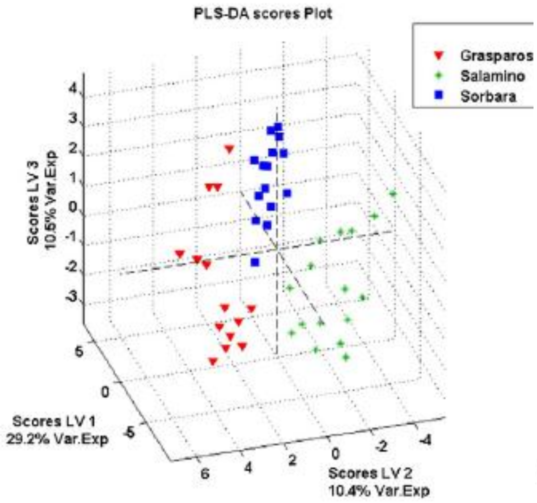
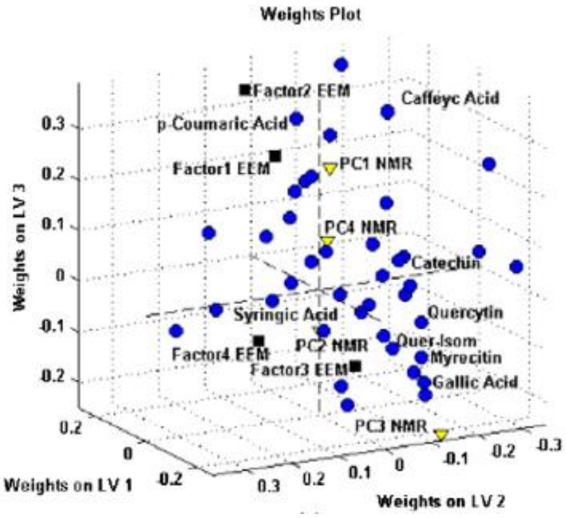
Courtesy of Ref: 19

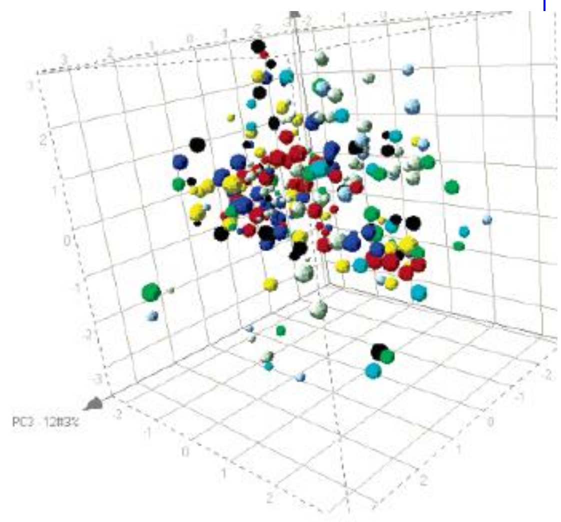
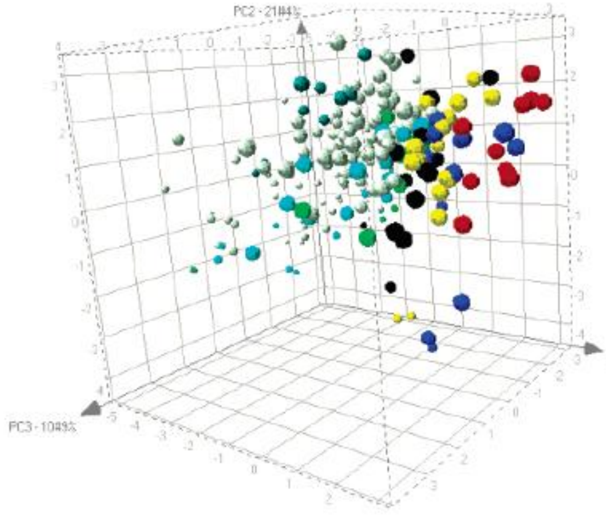
Variable importance (influence) in projection



Courtesy of Ref: 44

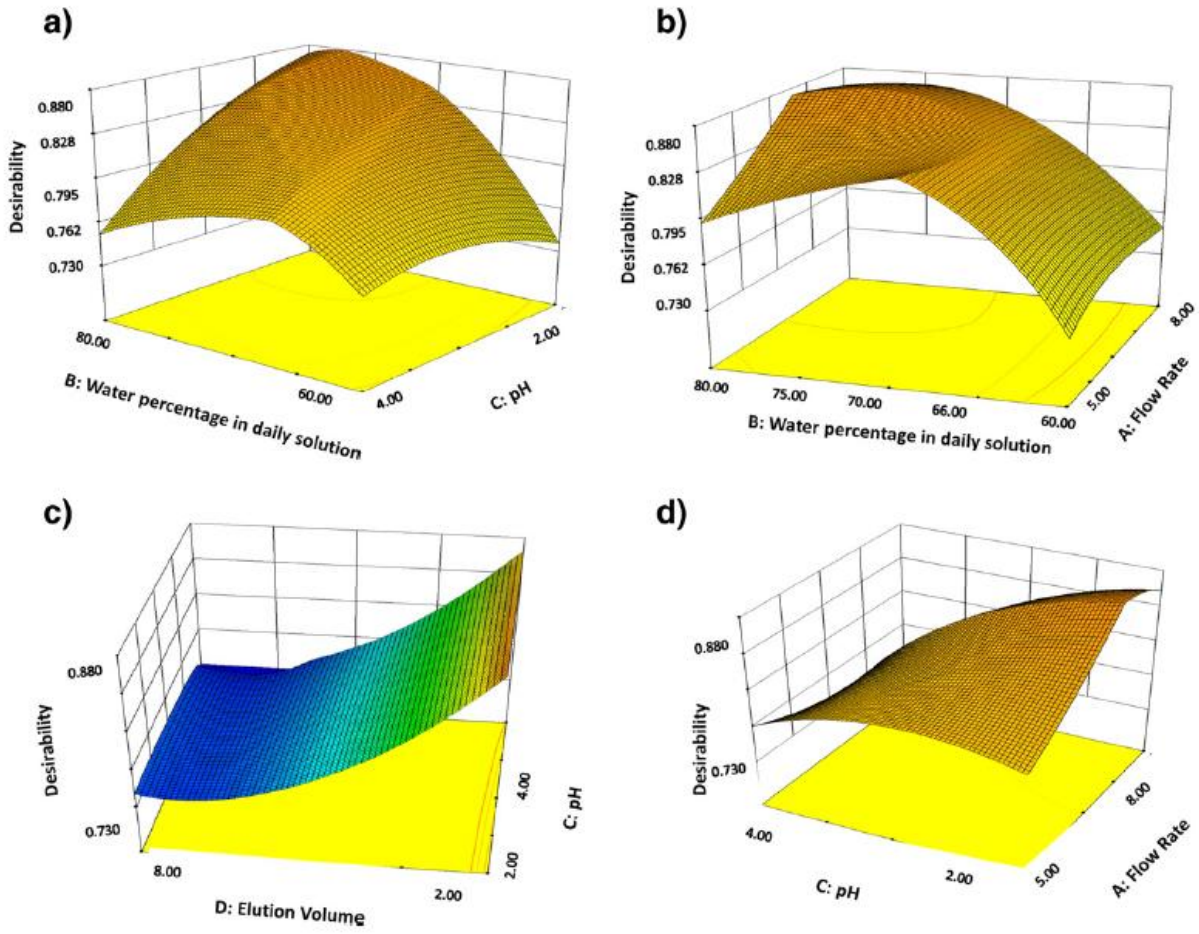
 Courtesy of Ref: 46	 Courtesy of Ref: 46

PLS-DA model for fused dataset	
Scores plot	weights plot
 Courtesy of Ref 44	 Courtesy of Ref ~~44~~

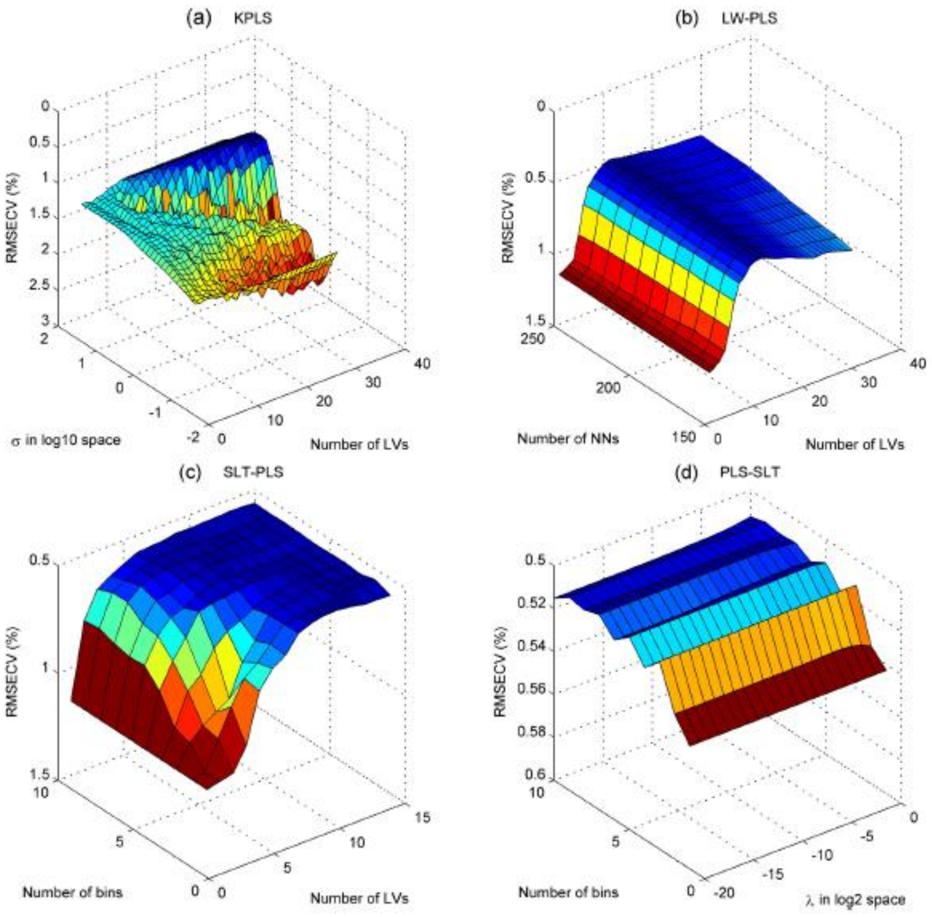
Serum	Yeast supernatant
	

Loadings of the first (samples) mode of the PARAFAC model for EEM dataset

(a) [water percentage in daily solution, pH] (b)[water percentage in daily solution, flowrate of loading step]
(c) [elution volume, pH] (d) [pH, flowrate of loading step]



CV error surfaces for (a) KPLS, (b) LW-PLS, (c) SLT-PLS and (d) PLS-SLT

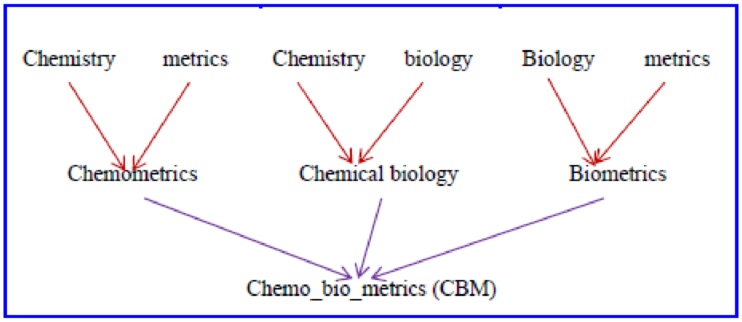


Appendix A4:
m-way Models

Models for m-way (1-way-, 2-way, 3way- and 4-way)datamodeling

The preliminary step is to look at the scatter diagrams in 1-D, 2D- and 3-dimensions (appendix A5). The next leap is exploratory data analysis using ANOVA, linear correlation and (multi-) linear cause-effect models. In case of multiple X- and/or Y- tensors both unsupervised and super-vised model strategies are employed. Here is the jump to probe into appropriate set of model paradigms in data-space, transformed-(FT), orthogonalized, projected spaces and/or filtering (Kalman) methods to land on near appropriate cost-effective models for the task on hand. This it point, where a decision to probe into CPU- and expert-time intensive exhaustive iterative cycle of modeling/simulation and data collection again with ED and so on.

The subtle goals in mathematical parlance are classification/discrimination, minimum of number of factors, curve resolution, calibration, prediction etc. The curve resolution consists of arriving at spectra/chromatogram/time evolution of pure compounds or component (with large number of compounds). The multi-component calibration in presence of un-calibrated interferents/matrices is achieved through second order advantage of hyphenated instrumental data.



Discipline	**Task**	**Methods**
Environment	Pollution concentration = fn([Var])	Scatter profiles
Kinetics	Evolution, extent of reaction product conc. = fn(.)	Scatter profiles
Detection of number of clusters of species/compounds/routes	Response = fn(.)	m-way clustering
Multivariate Calibration in presence of inference's, background, masking agents, matrices		m-way calibration methods
Complex equilibria	PL or ML stability constant	

Calibration	Complicating factors
- Single analyte - Multiple analytes	Chemical system - Overlap of response profiles - Interferents - Matrices Instrument - Noise - Weak signal - Drift

Simulation - Complex equilibria	kinetics	
Effect of errors in instrument readings Ingredient concentrations	Log betas % species	Max error in y Safe area limits
Ingredient conc	% species (major/minor)	

Mathematical tasks	**Sub_goals**	**Primary_goals**
	Mathematical	

- Curve resolution
- Clustering
- Classification
- Discrimination
- Pattern recognition
- Calibration (cause effect Model)
	✓ Curve fitting
	✓ Model
	  ✓ Mathematical
	  ✓ Physico-chemical
		🕮 Understanding
		🕮 Control
		🕮 Prediction

Calibration		Billinear Non-bilinear	
Inverse calibration			
Factor analysis	2-way	Varimax rotation, Target factor, Evolving_FA, RAFA,	
	3-way	PARAFAC	
Orthogonalization	2-way	PCA,PLS	
Elimination of collinearity	3-way	PCA,PLS	Set of 2-way matrices 2-way PCA,PLS
	4-way	PARAFAC	Quadrilinear

Typical dimension reduction techniques **Input:Data Tensor (Xtensor or Ytensor)**

Abbreviation	$$$Factor Analysis (FA)	
Lin_FA	Linear	FA
NL_FA	Non-linear	
AFA	Abstract	
Varimax_FA	Rotation	
TFA	Target	
EFA	Evolving	
RAFA	Rank annihilation	
PARAFAC	Parallel	

Principal component Analysis (PCA)	
Linear	PCA
Non-linear	
Robust	

PCA	PLSC
PC_X	PLSC_XY
PC_Y	PLSC_YX
PC_XY	

Eigen vector (EV)	
Row wise	EV
Column wise	

Singular value decomposition (SVD)	

Principal component Analysis (PCA)	
Linear	PCA
Non-linear	
Robust	

Recursive hierarchical adaptive PCA (RecHier.Adap.PCA) algorithm

S. Rännar, J.F. MacGregor, S. Wold, Chemom. Intell. Lab. Syst. 41 (1) (1998) 73–81. Adaptive batch monitoring using hierarchical PCA

	Chemical Rank	
Rank	Instrument	
1	UV-Vis	Rank-1 analyte
1 to 5	NMR	Medium rank
>5	MS-MS	High rank
	Residual nonbilinear	

Linear	Bi-linear
	Tri-linear
	Quadru-linear
NL	Non-bilinear
Residual	Residual bilinear
	Residual nonbilinear

- ❖ Just-in time learning

Partial least squares analysis (PLSA)

Input: Both X and y data

- ❖ Linear and non-linear PLSA
- ❖ Hot hierarchically ordered taxonomic Hot_PLSA
  - + Employs hierarchical structure
  - + Models linear or non-linear subtle details,
  - + Uses extra information

Translation of a task in a discipline into mathematical parlance

Discipline	**Task**	**Methods**
Cause-effect relationships	Analytical chemistry	Calibration
		Inverse calibration
	Physical chemistry/biochemistry, Pharmacy	SXR

Criteria for optimized models

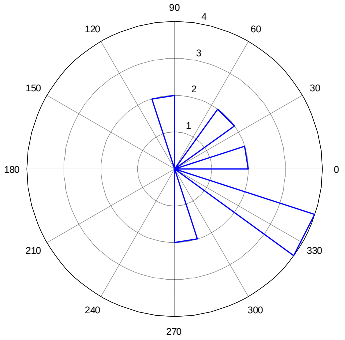
Normal distribution		Base
	Errors in y	
Least squares	Homo sedastic	
Weighed LS	Hetero sedastic	
Robust	**To Outliers**	
Least median squares		
Double median		
Least absolute deviation		

Bayesian concept	
Deviations and residuals are used rather as synonyms.	

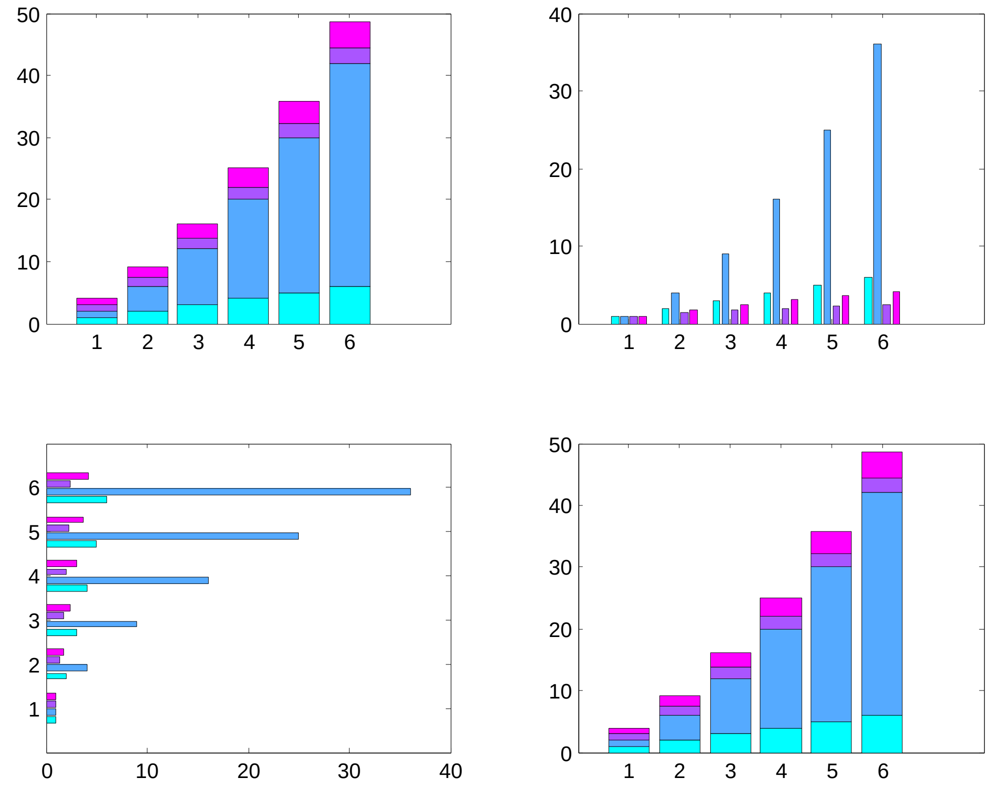
AppendixA5:
Graphic Display of Data and its products

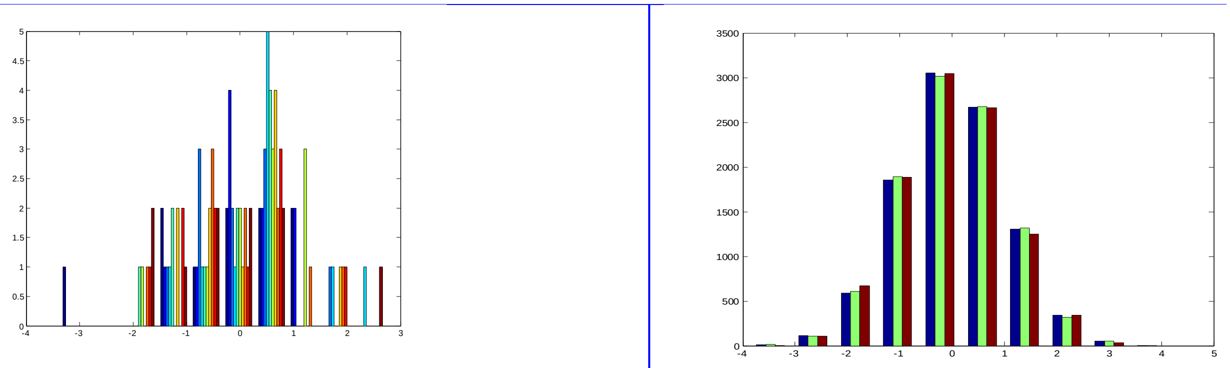
Scatter Graphic Display diagram

- **Data**
  - **Primary**
  - **Transformed**
- **Parameters**
- **Residuals**
- **Effect of errors on**
  - **Data**
    - **Primary**
    - **Transformed**
  - **Parameters**
  - **Residuals**

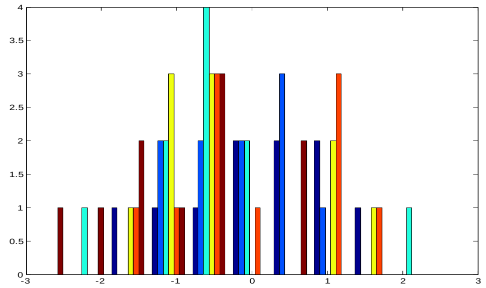


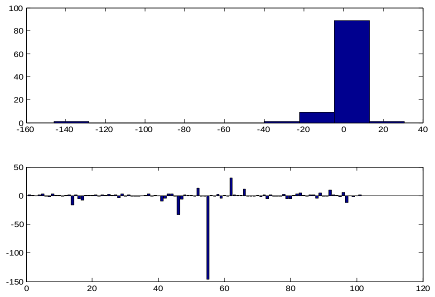
Histograms in Polar Coordinates

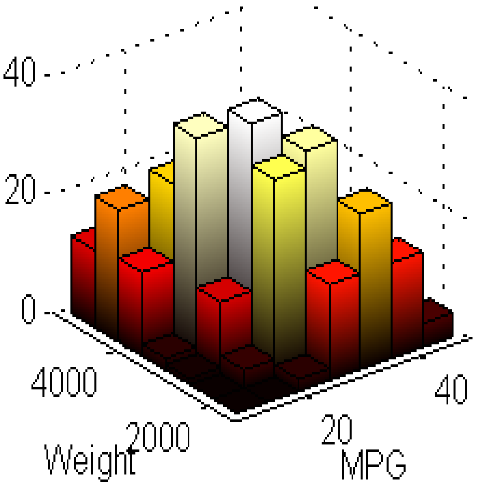


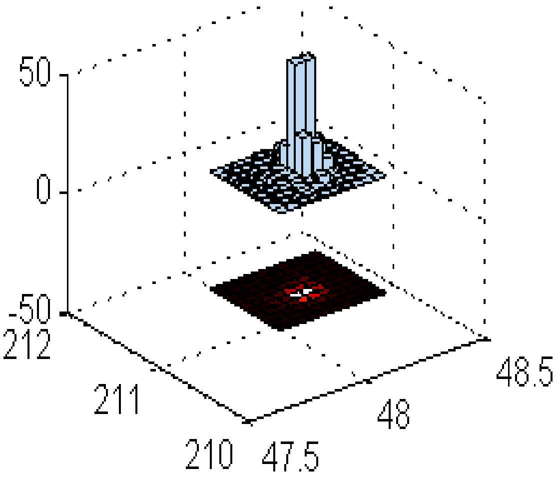


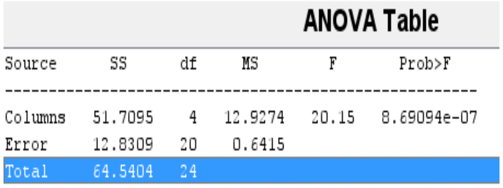
Histogram of matrices





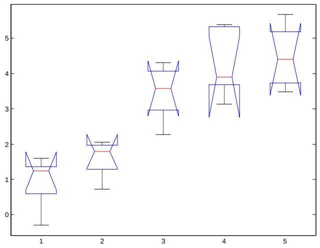


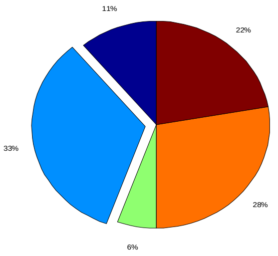


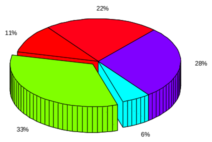


```
                         ANOVA Table
Source    SS      df    MS       F      Prob>F
-----------------------------------------------------------
Columns   51.7095   4   12.9274   20.15   8.69094e-07
Error     12.8309  20    0.6415
Total     64.5404  24
```

```
X = meshgrid(1:5)
X = X + normrnd(0,1,5,5)
p = anova1(X)
```



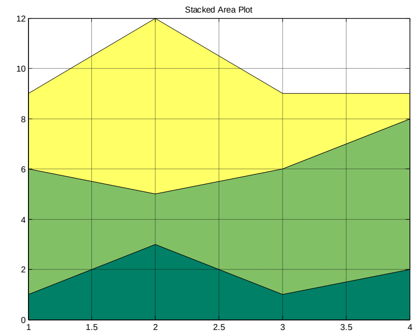


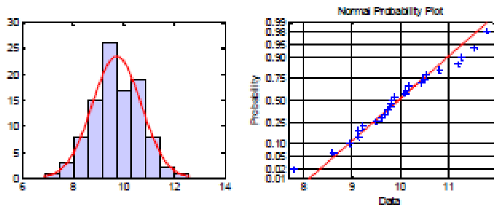


```
clean
Y = [1, 5, 3;
 3, 2, 7;
 1, 5, 3;
 2, 6, 1];
area(Y)
grid on
colormapsummer
set(gca,'Layer','top')
title ('Stacked Area Plot')
```

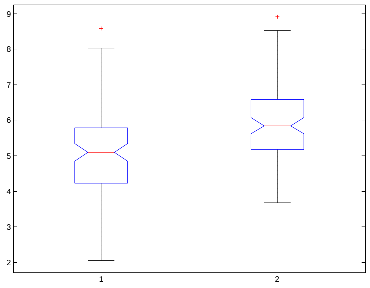
```
clean
x = [1 3 0.5 2.5 2];
explode = [0 1 0 0 0];
pie(x,explode)
colormapjet
```

```
%Slice in the pie chart
%Assign corresponding
explode element to 1:
x = [1 3 0.5 2.5 2];
explode = [0 1 0 0 0];
pie3(x,explode)
colormaphsv
```



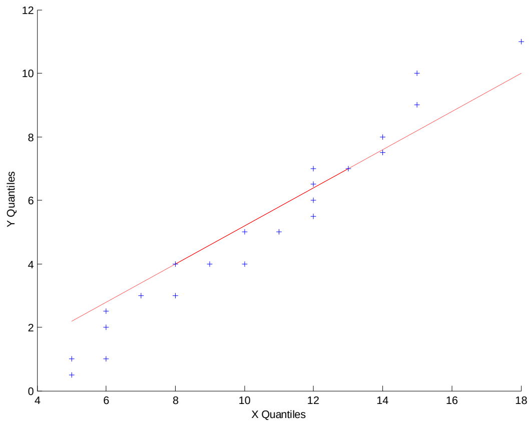


```
Clean
r = normrnd(10,1,100,1);
subplot(221),histfit(r)
h = get(gca,'Children');
set(h(2),'FaceColor',[.8 .8 1])
subplot(222),x = normrnd(10,1,25,1);

normplot(x)
```



Boxplot

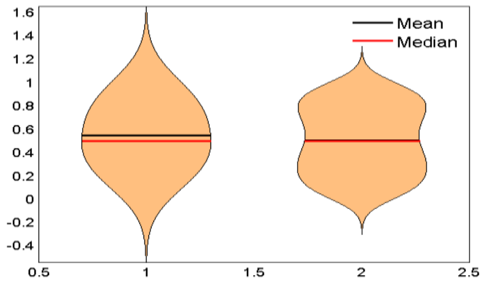
```
clean
x1 = normrnd(5,1,100,1);
x2 = normrnd(6,1,100,1);
boxplot([x1,x2],'notch','on')
```



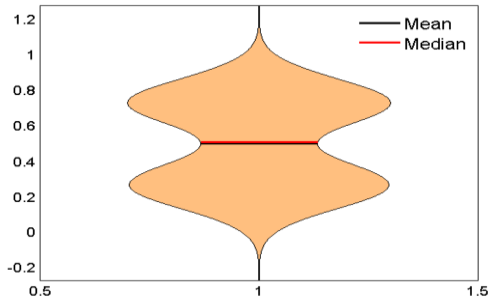
qqplot

```
clean
x = poissrnd(10,50,1);
y = poissrnd(5,100,1);
qqplot(x,y);
```

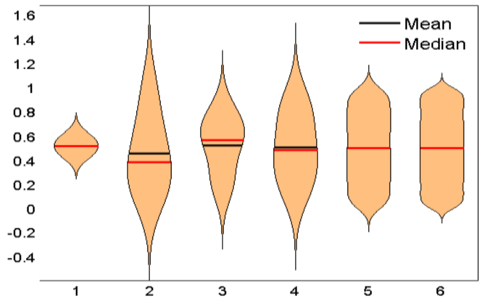
Violin plot



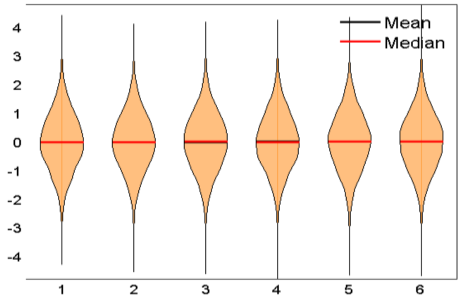
```
Y{:,1}=rand(10,1);
Y{:,2}=rand(1000,1)
 violin(Y,2)
```



```
rn = randn(1000,1)*0.38+0.5;
        rn2 =
[randn(500,1)*0.1+0.27;randn(500,1)*0.1+0.73]
;
        rn2=min(rn2,1);rn2=max(rn2,0);
 violin(rn2,2)
```

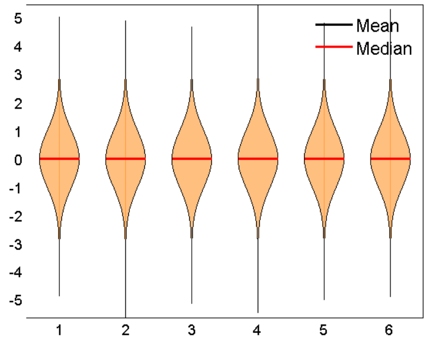


```
 Y{:,1} = rand(2 ,1); Y{:,2} = rand(6,1);
 Y{:,3} = rand(10,1); Y{:,4} = rand(50,1);
 Y{:,5} = rand(10000 ,1); Y{:,6} =
rand(100000,1);
% Y{:,7} = rand(1e6 ,1); Y{:,8} =
rand(1e7,1);
 violin(Y,6)
```

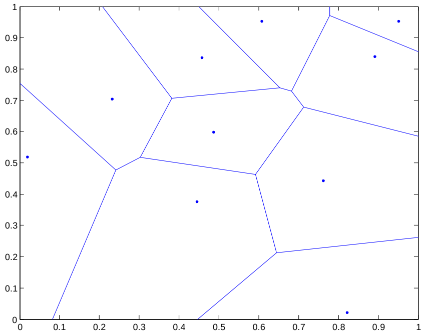


```
 n =10000;
 Y{:,1} = randn(n,1); Y{:,2} = randn(n,1);
 Y{:,3} = randn(n,1); Y{:,4} = randn(n,1);
 Y{:,5} = randn(n ,1); Y{:,6} = randn(n,1);
```

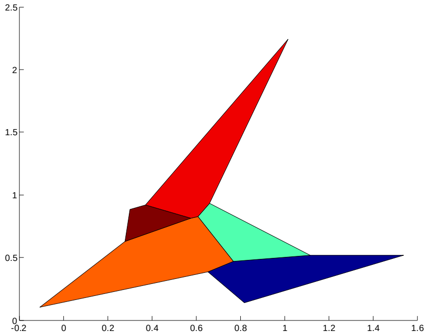
```
n =1e6
 Y{:,1} = randn(n,1); Y{:,2} = randn(n,1);
 Y{:,3} = randn(n,1); Y{:,4} = randn(n,1);
 Y{:,5} = randn(n ,1); Y{:,6} = randn(n,1);
 violin(Y,6)
```



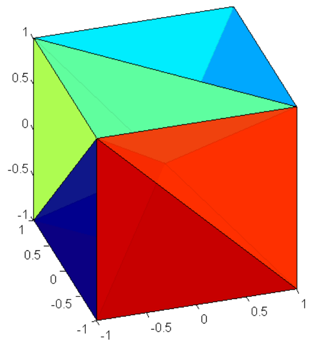
```
mean(Y)
-0.0003    -0.0015     0.0008     0.0010    -0.0005
0.0007
std(Y)
0.9997     1.0003     0.9994     0.9990     0.9999
1.0004
```



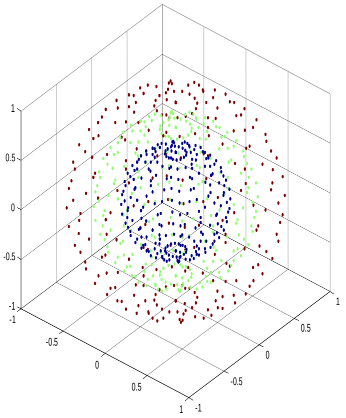
```
clean
x = gallery('uniformdata',[1 10],0);
y = gallery('uniformdata',[1 10],1);
voronoi(x,y)
```

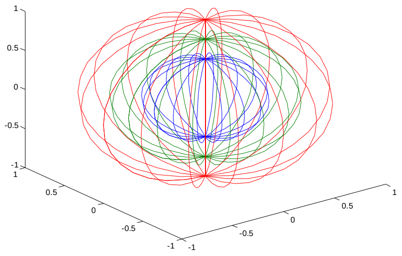


```
x = gallery('uniformdata',[10 2],5);
[v,c]=voronoin(x);
fori = 1:length(c)
if all(c{i}~=1)    % If at least one of
the indices is 1,
% then it is an open region and we
can't
% patch that.
patch(v(c{i},1),v(c{i},2),i); % use
color i.
end
end
```



[y,[X2]]	Response in [probe1-probe2] instruments
	Fn(Conc) vs variables in Env (sites, ..)
	Response,Fn(ED_design matrix)
After model build up	
2D-contour	Y vs (parameters) Y vs(residuals)





Appendix A6:
Output and matlab programs for Canonical analysis

```
   Knowledge Based System for Canonical analysis
                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 --> Inference based on sign disc (=4)
contours are concentric parabolic
          since disc > 0

 --> Inference based on sign canonical coefficients
     1     0     1
ellipses
          since a12 == 0 & a22 >0 & a11 >0


                 Knowledge Based System for Canonical analysis
                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 --> Inference based on sign disc (=4)
```

```
contours are concentric parabolic
          since disc > 0

 --> Inference based on sign canonical coefficients
    -1      0    -1
ellipses
          since a12 == 0 & a22 <0 & a11 <0

                 Knowledge Based System for Canonical analysis
                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 --> Inference based on sign disc (=-4)
contours are hyperbolic
          since disc < 0

 --> Inference based on sign canonical coefficients
    -1      0     1
 hyperbolic
          since a12 == 0 & a22 <0 & a11 <0

                  Knowledge Based System for Canonical analysis
                 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 --> Inference based on sign disc (=-4)
contours are hyperbolic
          since disc < 0

 --> Inference based on sign canonical coefficients
     1      0    -1
 hyperbolic
          since a12 == 0 & a22 <0 & a11 <0

 %
%    dem_kbcan.m    R S Rao 4/9/95
%
  clean
kb_can(1,0,1)
kb_can(-1,0,-1)
kb_can(-1,0,1)
kb_can(1,0,-1)
```

```
%
%             kb_can   R S Rao  12-8-15;  4/9/95
%

functionkb_can(a11,a12,a22)

     heading = 'Knowledge Based System for Canonical analysis';
     center02(heading);
%
          ant11 = 'disc < 0';
          conseq11 = 'contours are hyperbolic';
%
          ant12 = 'disc > 0';
          conseq12 = 'contours are concentric parabolic';
%
st   = [blanks(10) 'since '];
%
 %%      KB based on sign disc
%
          disc = 4 * a22 * a11 - a12 * a12;

disp([' --> Inference based on sign disc (=', num2str(disc),')'])
ifeval(ant11)
```

```
disp(conseq11),disp([st ant11])
end
ifeval(ant12)
disp(conseq12),disp([st ant12])
end%%
%%


%
    %%          KB based on sign canonical coefficients
%
disp(' '),disp(' --> Inference based on sign canonical coefficients')
disp( [a11 a12 a22])
%
         ant1 = 'a12 == 0 & a22 >0 & a11 >0'; conseq1 =  'ellipses';
ifeval(ant1)
disp(conseq1), disp([st ant1])
end
%
         ant2 = 'a12 == 0 & a22 <0 & a11 <0';conseq2 =  'ellipses';
ifeval(ant2)
disp(conseq2), disp([st ant2])
end
%
         ant3 = 'a12 == 0 & a22 >0 & a11 <0'; conseq3 =  ' hyperbolic';
ifeval(ant3)
disp(conseq3),disp([st ant2])
end
%
         ant4 = 'a12 == 0 & a22 <0 & a11 >0'; conseq4 =  ' hyperbolic';
ifeval(ant4)
disp(conseq4),disp([st ant2])
end
disp(' ')

           %%
```

```
 %
%    kb_landa.m  4/9/97
%
functionkb_landa(landa)
     %%
if any(landa<0)
disp([' - Chemically invalid values'])
disp(['   Remedy : repeat expt with -ve values making zero'])
end
     %%
%
       conseq1 = 'Parabolic bowl opening upwards';
       conseq2 = 'Parabolic bowl opening downwards';
       conseq3 = 'Flattened parabolic bowl opening downwards';
       conseq4 = 'Ridge';
       conseq5 = 'Saddle region';

       ant1 = 'landa(1) > eps &landa(2) > eps';
       ant2 = 'landa(1) < eps &landa(2) < eps';
       ant3 = 'landa(1) < eps &landa(2) > eps';
       ant4 = 'landa(1) > eps &landa(2) < eps';

st   = [blanks(10) 'since '];
      %%
```

```
ifeval([ant1])
disp ([conseq1]), disp ([st ant1])
end
ifeval(ant2)
disp ([conseq2]),disp ([st ant2])
end
ifeval(ant3)
disp ([conseq4]),disp ([st ant3])
end
ifeval(ant4)
disp ([conseq5]),disp ([st ant4])
end
      ant6 = 'landa(1) == 0';
ifeval(ant6)
disp ([ant6]),disp ([landa(2)])
end
      ant7 = 'landa(2) == 0';
ifeval(ant7)
disp ([ant7]),disp ([landa(1)])
end
      eps
    %%
```

```
%
%  dem_landa.m      R S Rao 24-8-15; 4/9/97
%
clean
landa  = [4;2];    kb_landa(landa)
landa  = [-4;-2];  kb_landa(landa)
landa  = [-4;2];   kb_landa(landa)
landa  = [4;-2];   kb_landa(landa)
```

```
               Knowledge Based System based on Eigen analysis
               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

landa = [4,2]
Parabolic bowl opening upwards
          since landa(1) > eps &landa(2) > eps

landa = [-4,-2]
 - Chemically invalid values
   Remedy : repeat expt with -ve values making zero

Parabolic bowl opening downwards
          since landa(1) < eps &landa(2) < eps


landa = [-4, 2]
 - Chemically invalid values
   Remedy : repeat expt with -ve values making zero
Ridge
          since landa(1) < eps &landa(2) > eps
landa = [4, -2]
 - Chemically invalid values
   Remedy : repeat expt with -ve values making zero
Saddle region
          since landa(1) > eps &landa(2) < eps

eps = 2.2204e-16
```

$$Let \begin{bmatrix} s1 \\ s2 \end{bmatrix} = \begin{bmatrix} x1 \\ x2 \end{bmatrix}$$

s corresponds to co-ordinates of stationary point.

$$[e,l] = eig(s)$$

$$[eig\_vec, landa] = eig\left(e^{T}\right)$$

$$Xcan = x1\_tran * e(1,1) + x2\_tran * e(2,1)$$

$$Xcan = y1\_tran * e(1,2) + y2\_tran * e(2,2)$$

$$x\_trans = x - one * S$$

$$x\_can = eig\_vec * x\_tran$$

$$zcan = a0 + l(1,1) * x\_can^2 + l(2,2) * y\_can^2$$

$$zcan - a0 = l(1,1) * x\_can^2 + l(2,2) * y\_can^2$$

$$z\_can = landa^{T} * x\_tran$$

The response at the stationary pint is calculated as

$$\text{Resp} = \begin{bmatrix} a0 + a1 * s1 + a2 * s2 + \\ a12 * s1 * s2 + \\ a11 * s1^2 + a22 * s2^2 \end{bmatrix}$$

Negative Eigen values are made zero

Response with only positive and zero Eigen values

```
%
%    eigenAnal.m     24-8-15; 4/9/97
%
functioneigenAnalysis(range,a0,a1,a2,a11, a22,a12)
     S = [2*a11     a12
          a12     2*a22];
     f = [a1 a2     ]';
     s = - pinv(S) * f;
```

```
     heading = 'Knowledge Based System based on Eigen analysis';
     center02(heading);

disp('  Eigen values &eigen vectors')
     [e,l] =eig(S),landa =diag(l);

kb_landa(landa);

%  Coordinates of Stationary points

disp([' Coordinates of Stationary points : '])
xopt = s;disp([s'])
z_sta = a0 + a1 * s(1) + a2 * s(2) + a11 * s(1).^2 + a22 *s(2).^2 ...
                 + a12*s(1).*s(2);

disp([' Response at Stationary Point : ',num2str(z_sta)])

fori = 1:length(s)
if s(i) < 0
          s(i) = 0;
end
end
z_sta_ = a0 + a1 * s(1) + a2 * s(2) + a11 * s(1).^2 + a22 *s(2).^2 ...
                 + a12*s(1).*s(2);
disp([' Response at all s positive/zero  : ',num2str(z_sta_)])

  %%

     n1 = range(1);     n2 = range(2);     n3 = range(3);
     n4 = range(4);     n5 = range(5);     n6 = range(6);
     [x,y] = meshgrid(n1:n2:n3,n4:n5:n6);
     [rx,cx] = size(x);onex = ones(rx,cx);
     [ry,cy] = size(y);oney = ones(ry,cy);
     x1 =   (x - s(1) *onex);
     y1 =   (y - s(2) *onex) ;
%clear onexoney
xcan =x1*e(1,1) + y1 *e(2,1);
ycan =x1*e(1,2) + y1 *e(2,2);
% clear x1 y1
     z = a0 + l(1,1) * xcan.^2 + l(2,2) * ycan.^2;
     figure, plot_can(z,xcan,ycan)
```

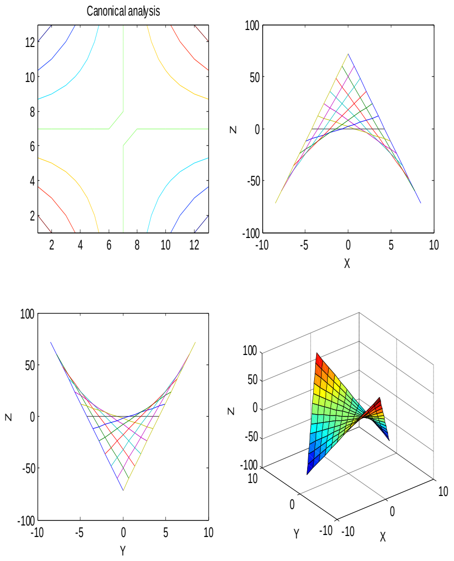
```
%
%  dem_can.m      R S Rao 24-8-15; 4/9/97
%
clean
range = [-6 1 6 -6 1 6]'; range'

par = [0 1 0.  0 0 0]';can(par,range)
par = [0 0 1.  0 0 0]';can(par,range)

par = [0 0 0.  0 1 0]';can(par,range)
par = [0 0 0.  0 0 1]';can(par,range)
par = [0 0 0.  1 0 0]';can(par,range)
par = [0 0 0.  0 1 1]';can(par,range)
par = [1 1 1.  1 1 1]';can(par,range)
```

```
par = [0 0 0.  0 1 1]';can(par,range)
par = [0 0 0.  0 -1 -0.9]';can(par,range)
par = [0 0 0.  0 -1 -1]';can(par,range)
par = [0 0 0.  0 -0.91 -0.02]';can(par,range)
par = [1 1 1.  0 0.7 -1]';can(par,range)

par = [0 1 1. 0 0 0 0]';can(par,range)

par = [3.264 1.537 0.5664 -0.05875 -0.1505 -0.0273]';
range = [0 1 10 0 1 10]'
can(par,range)
```

```
a0     x     y   x*y    x^2   y^2   Coef of full quadratic model
  0     0     1     0       0       0
 !!! Error: Canonical Analysis is not possible
            since coefiicients of second order terms are all zero
               So, Run aborted
```

```
************ Canonical analysis *********
                   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    a0     x     y   x*y    x^2   y^2   Coef of full quadratic model
     0     0     0     0     0     1
 NC_Canonical Analysis: Coefficients of second order terms <> 0 is satisfied
```

```
    Knowledge Based System based on Eigen
analysis

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  Eigen values & eigen vectors
e = eigen vectors
   -0.7071    0.7071
    0.7071    0.7071

l = eigen values
    -1     0
     0     1
-Chemically invalid values
   Remedy : repeat expt with -ve values making
zero

Ridge
since landa(1) < eps & landa(2) > eps

Coordinates of Stationary points : 0      0
Response at Stationary Point : 0
Response at all s positive/zero   : 0
```



AUTHORS' ADDRESSES

1.**K RamaKrishna**
Department of Chemistry,
Gitam Institute of Science,
Gitam University,
Visakhapatnam, A.P
2. **R. Sambasiva Rao**
School of Chemistry,
Andhra University,
Visakhapatnam 530 003, A.P