



Designs in constrained experimental regions

Design and Analysis of Mixture Experiments

*Prof Eddie Schrevens
Biosystems Department
Faculty of Bioscience Engineering
K. Universiteit Leuven
Belgium*

Eddie.schrevens@biw.kuleuven.be

Part I

General introduction to design and analysis of mixtures

What's expected to be understood

- **Sampling strategy versus experimental design**
- **What is experimentation. Basic principles and properties.**
- **'Classical' experimental design theory**
- **Multiple regression in Response Surface Methodology. Dependent and independent variables are numeric. The predicted response is studied.**
- **Specific designs for RSM**
- **Randomised Complete Designs**

What's next?

Generalisation to constrained experimental regions

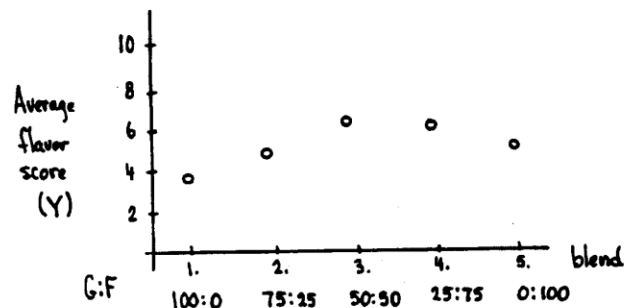
Generalisation to optimisation in constrained experimental regions: What are mixtures?

Tea Sweetener Example

Suppose we wished to study blends of two sweeteners (glucose, fructose) in making iced tea. The total amount (by volume) of the sweeteners in the tea is 2%. The measured response is flavor.

Design strategy: Select 5 different blends of glucose and fructose and make 10 batches of each blend.

Glucose	Fructose	Blend	Average Flavor Score (1 to 10)
100	0	1.	3.5
75	25	2.	4.7
50	50	3.	6.2
25	75	4.	6.0
0	100	5.	5.2



Note that flavor varies not as a result of changing the amount of sweetener but rather with the ratio of G/F.

Generalisation to optimisation in constrained experimental regions: What are mixtures?

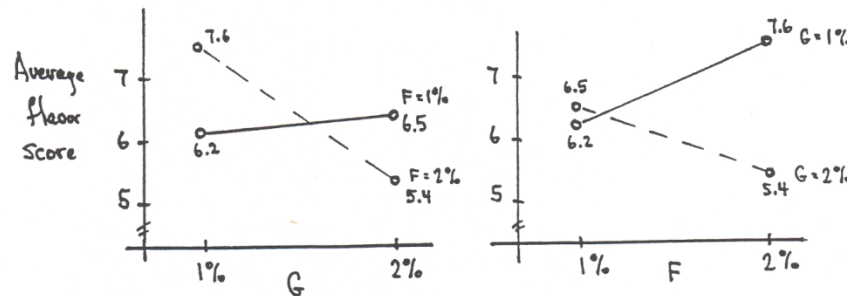
Tea Sweetener Example (cont.)

Now, suppose we take our two sweeteners and study each at two levels, say 1% and 2% (by volume). If we setup the $2 \times 2 = 4$ combinations of G and F, we would have

G	F	Total %	Ave. flavor score
1%	1%	2	6.2
2%	1%	3	6.5
1%	2%	3	7.6
2%	2%	4	5.4

How would we analyze and interpret these data values? To begin, we might seek answers to the following questions:

- Does increasing G increase the response? How about F?
- Does increasing both G and F increase flavor? If not, why not?



Question: How does the above experimental strategy (the 2^2 design) differ from the design chosen on the previous page?

Generalisation to optimisation in constrained experimental regions

Design and analysis of mixture systems

Definition

In the general mixture problem, the response that is measured is only a function of the proportions of the ingredients present in the mixture and not of the amount of the mixture.

For a mixture system consisting of q components, with x_i the fraction of the i -th component the following equations are valid :

$$0 \leq x_i \leq 1 \quad \text{for } i = 1, 2, 3, \dots, q$$

$$\sum_{i=1}^q x_i = 1$$

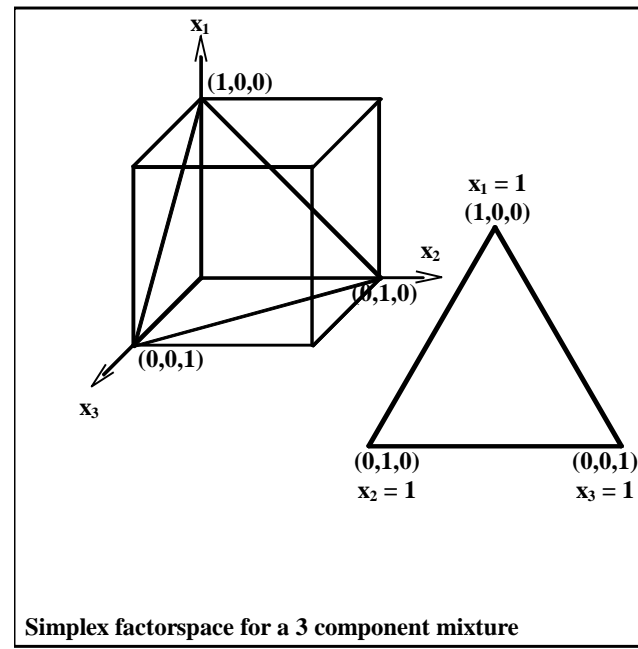
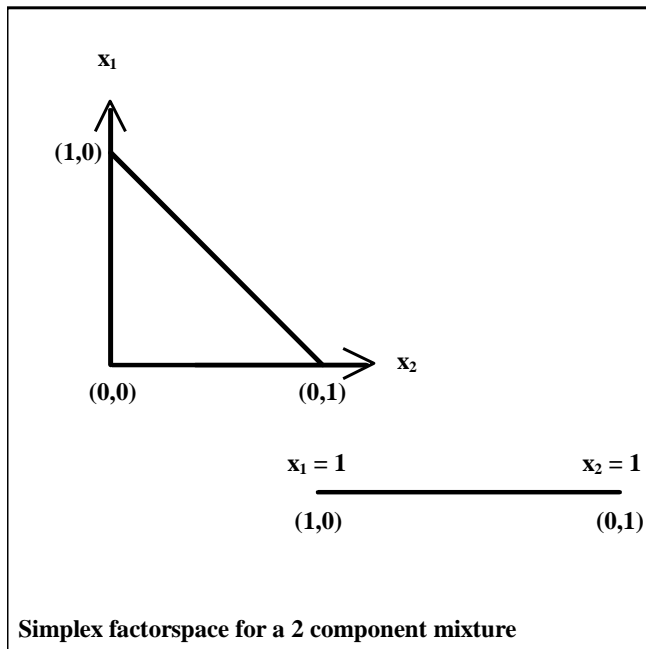
Design and analysis of mixture systems

The q components of a mixture system are called “mixture variables”. The proportion of each mixture variable can vary from 0 (the component is not present) to 1, a mixture with only one component, called a ‘pure mixture’. If in a q component mixture the proportion of $q-1$ mixture variables is determined, then the proportion of the q_{th} mixture variable is also determined:

$$x_q = 1 - \sum_{i=1}^{q-1} x_i \quad \longrightarrow \quad \text{One exact multicollinearity constraint}$$

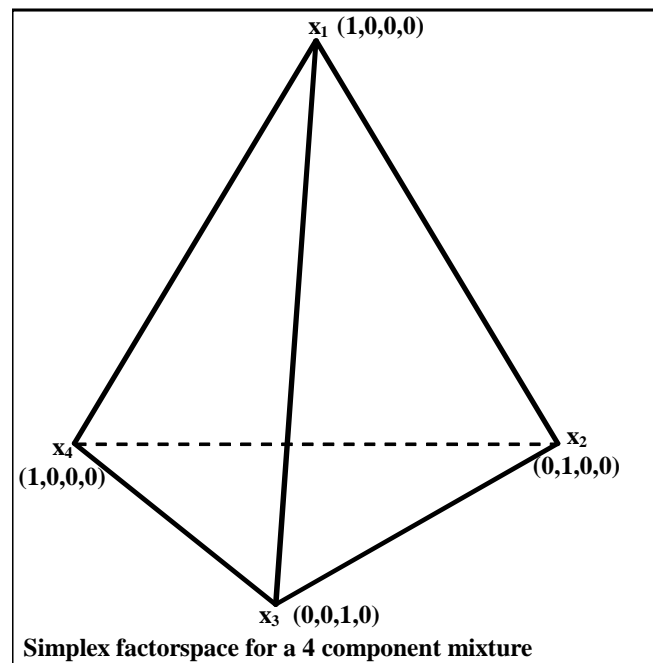
Design and analysis of mixture systems

In this way the mixture equations reduce the q dimensional factor space to a $q-1$ dimensional simplex.



Design and analysis of mixture systems

In this way the mixture equations reduce the q dimensional factor space to a $q-1$ dimensional simplex.



Design and analysis of mixture systems

As a result of the mixture constraints classical orthogonal experimental designs can not be used. Confounding between mixture components is inherent to the mixture problem

Specific designs are developed

Specific 'mixture models' are fitted with Response Surface Methodology.

Emphasis on predicted response estimation.

Design and analysis of mixture systems

Some examples of mixtures

Food sciences: dough mixes, chocolate, ice cream, wine blending, fruit juices, fish paties, ...

Chemical industry: gasoline blending, textile fiber blends, explosives, ...

Ceramics industry

Pharmaceutical industry

Agriculture: nutrient solutions, fertilizers, fodder, multiple cropping, ...

Medical sciences: media for in vitro culture, ...

Consumer sciences: organoleptic attributes, ...

Design and analysis of mixture systems

Taking care of the mixture constraints

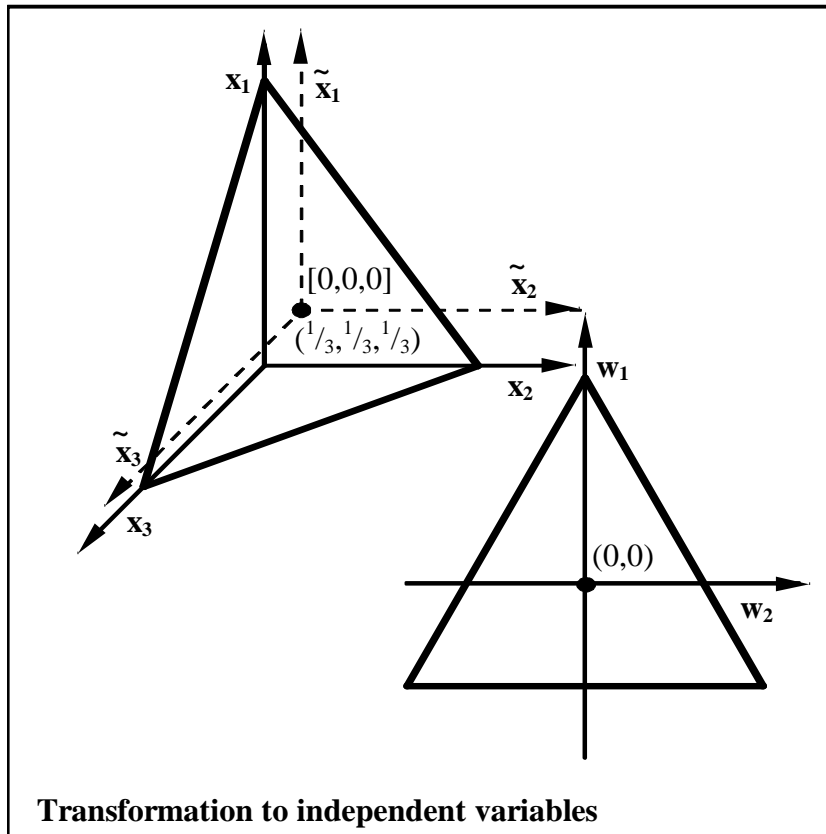
Two possibilities to take in account the mixture constraints

1. Transform the q mixture components to $q-1$ independent, orthogonal factors. In this way the mixture problem is reduced to classical experimentation with orthogonal factors
2. Incorporate the mixture constraints into the models. This leads to specific 'mixture' models

Design and analysis of mixture systems

Taking care of the mixture constraints

Transform the q mixture components to $q-1$ orthogonal factors.



Orthogonal transformation matrix
Rotate with $\text{svd}(v)$ (PCA)
Translate to centroid
Center and scale

Design and analysis of mixture systems

Taking care of the mixture constraints

Transforming the q mixture components to $q-1$ orthogonal factors is carried out by translation and rotation of the original axes system.

The use of mathematically independent variables has the major advantage of classical design and response surface fitting, the interpretation of the obtained coefficients is quite difficult, due to the fact that the independent variables are linear combinations of the mixture variables and have no practical meaning.

Specially for interpreting interactions between transformed variables, interpretation in the original mixture components is impossible.

Multiple mixtures, mixtures with additional constraints:
multicollinearity problems persist

Design and analysis of mixture systems

Taking care of the mixture constraints

Incorporate the mixture constraints into the models. This leads to specific 'mixture' models

RSM first and second order models are adapted to the mixture situation leading to Scheffe canonical mixture models

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$



$$\sum_{i=1}^q x_i = 1$$

$$y = (\beta_0 + \beta_1) x_1 + (\beta_0 + \beta_2) x_2 + \dots + (\beta_0 + \beta_k) x_k$$

$$y = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_k x_k$$

Design and analysis of mixture systems

Mixture models

Scheffe canonical mixture models

First order mixture model

$$y = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_k x_k$$

Second order mixture model

$$y = \gamma_1 x_1 + \dots + \gamma_k x_k + \gamma_{12} x_1 x_2 + \dots + \gamma_{(k-1)k} x_{k-1} x_k$$

Other mixture models

Cox models

Models with inverse terms

Adaptation of known theoretical models

Mixture experimental design

- Whole simplex experiments
 - Experiments over the full mixture space
- Homomorphic experimental regions
 - Experimental region has the same shape as the overall simplex
- Complex constrained experimental regions
 - Irregular convex hyper-polyhedron

Design and analysis of mixture systems

Specific mixture designs over the whole simplex

{q,m} Simplex Lattice Design is a design to fit a m-th order mixture model in q components

In this design each of the q mixture component varies with m+1 equally spaced values from 0 to 1

$$x_i = 0, 1/m, 2/m, \dots, 1$$

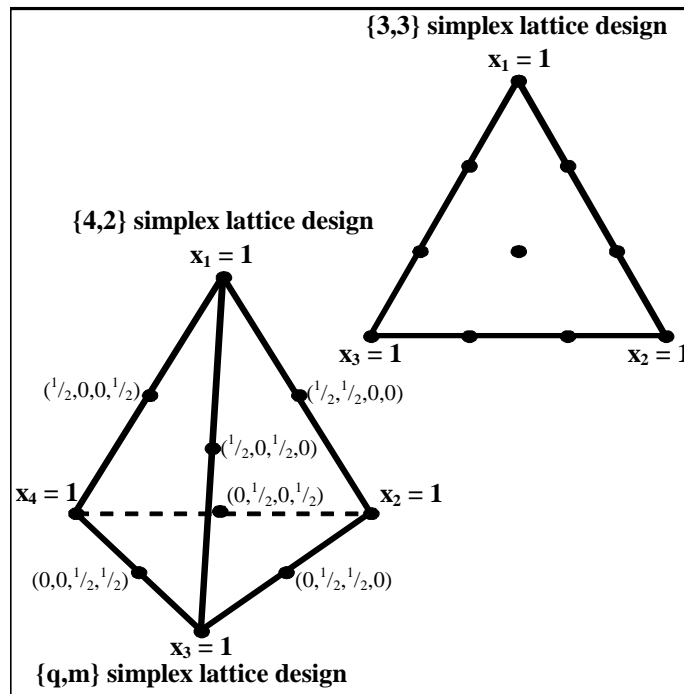
This design has the following number of treatments

$$\binom{q + m - 1}{m} = \frac{(q + m - 1)!}{m! (q - 1)!}$$

Design and analysis of mixture systems

Specific mixture designs over the whole simplex

{q,m} Simplex Lattice Design



Design and analysis of mixture systems

Specific mixture designs over the whole simplex

A Simplex Centroid Design consists of $2^q - 1$ treatments:

q pure components:

q permutations of $(1,0,0,0,\dots,0)$

$\binom{q}{2}$ binary mixtures with equal proportions:

permutations of $(1/2, 1/2, 0, 0, \dots, 0)$

$\binom{q}{3}$ ternary mixtures with equal proportions:

permutations of $(1/3, 1/3, 1/3, 0, \dots, 0)$

...

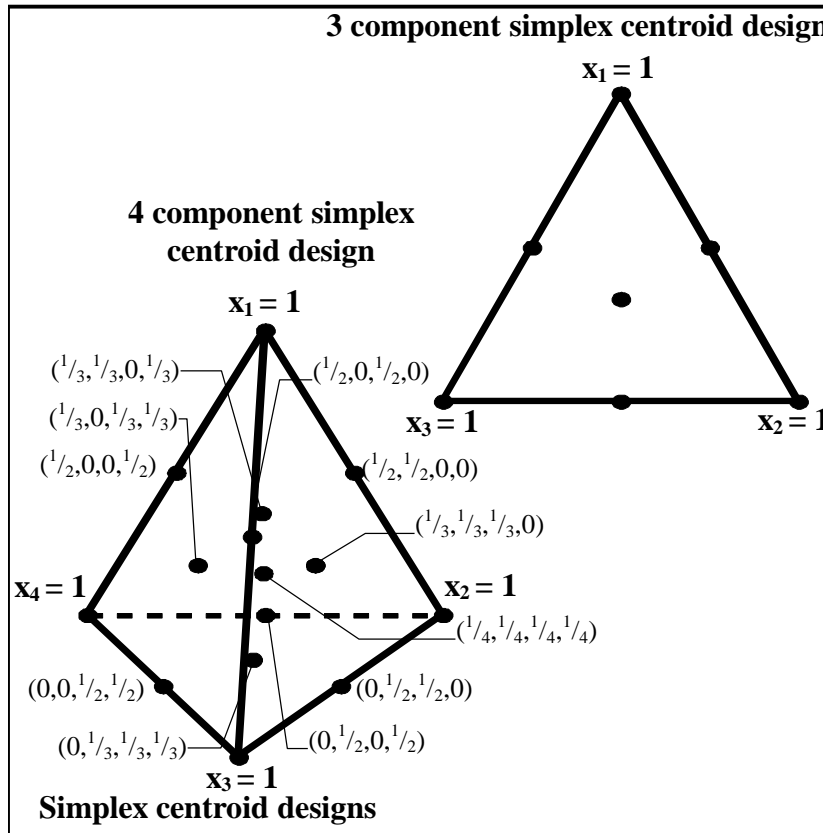
one q -nary mixture with equal proportions:

the mixture $(1/q, 1/q, 1/q, 1/q, \dots, 1/q)$

Design and analysis of mixture systems

Specific mixture designs over the whole simplex

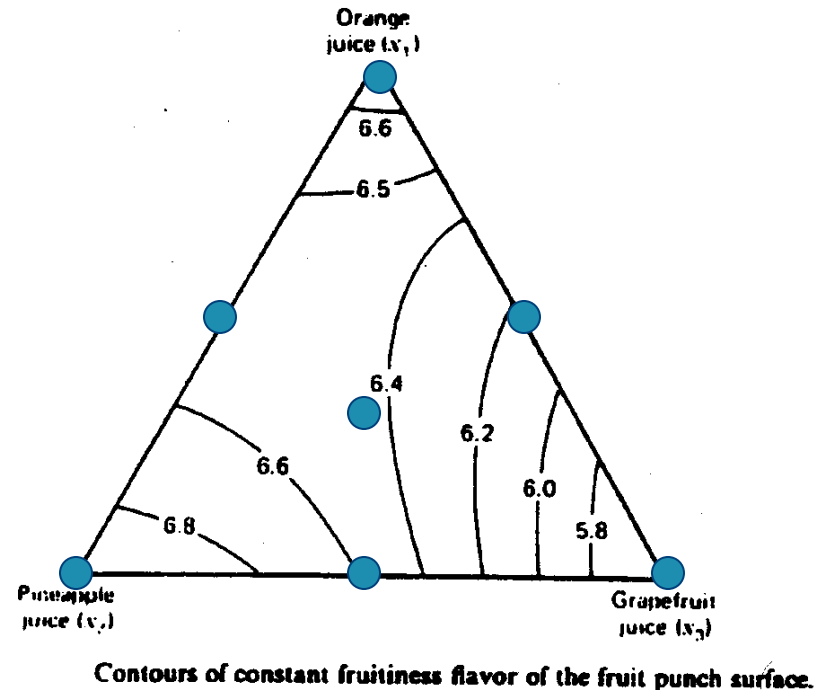
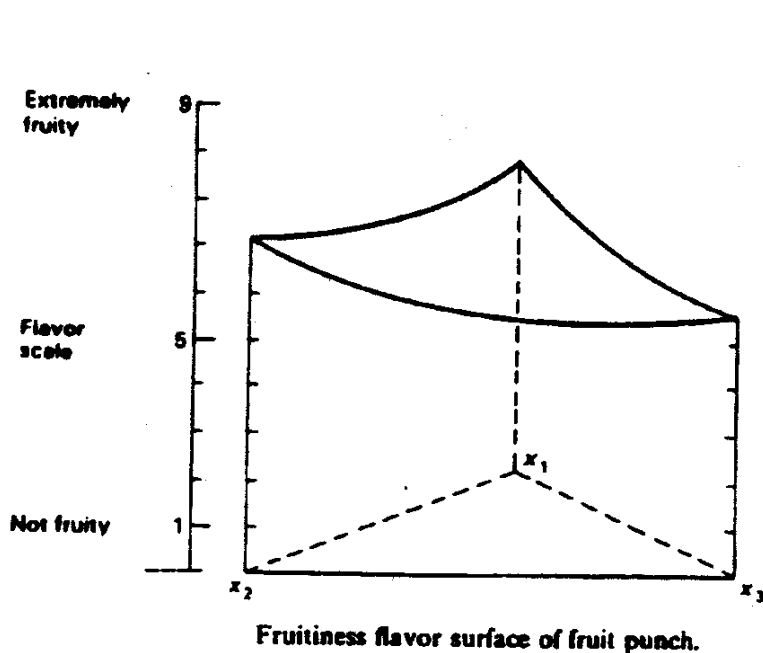
A Simplex Centroid Design



Design and analysis of mixture systems

An example of a whole simple mixture experiment

Fruitiness flavor of a fruit punch optimised in relation to proportions orange juice, pineapple juice and grapefruit juice (simplex centroid design, 10 replications per treatment blend)



Design and analysis of mixture systems

Additional constraints

Physiological
Economical
Chemical
Physical

Production of explosives

Single component constraints

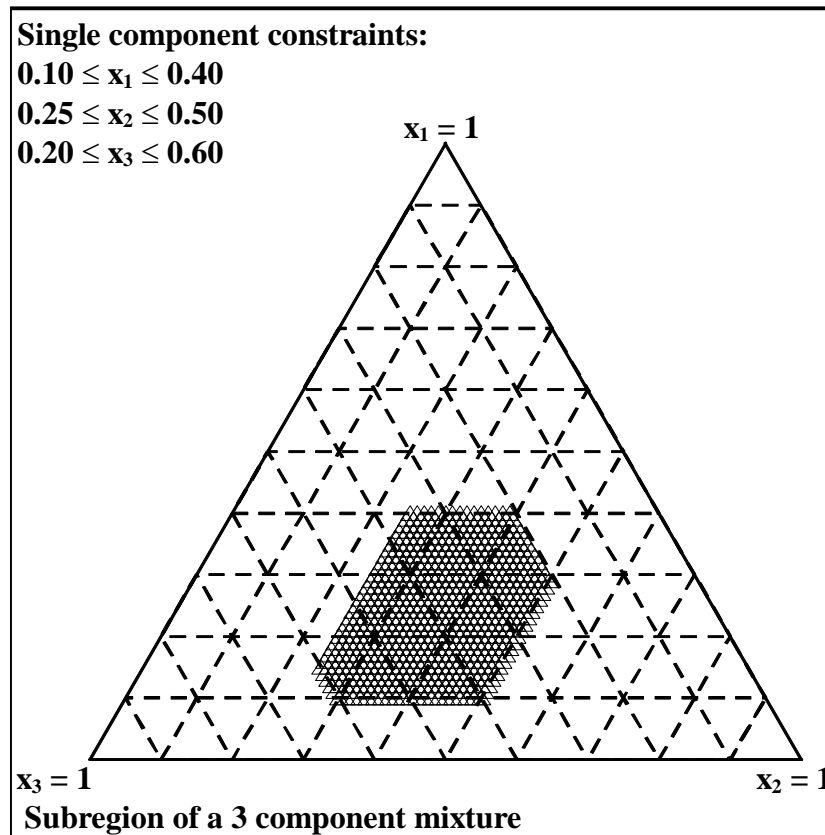
Imposing a certain lower (l_i) and/or upper (u_i) bound on the proportion of the mixture variables, reduces the factor space to a sub-region of the $(q - 1)$ - dimensional simplex, defined by the following equations:

$$0 \leq l_i \leq x_i \leq u_i \leq 1$$

with l_i and u_i respectively the lower and upper bound imposed on the component x_i .

Design and analysis of mixture systems

Additional constraints



Design and analysis of mixture systems

Additional constraints

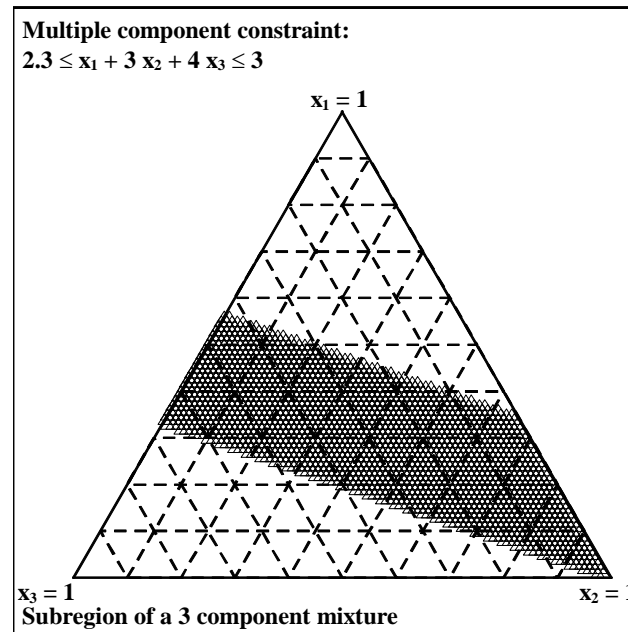
Multiple component constraints

Multiple component or multiple variable constraints are linear constraints of the form:

$$a_j \leq c_{1j} x_1 + c_{2j} x_2 + \dots + c_{qj} x_q \leq b_j$$

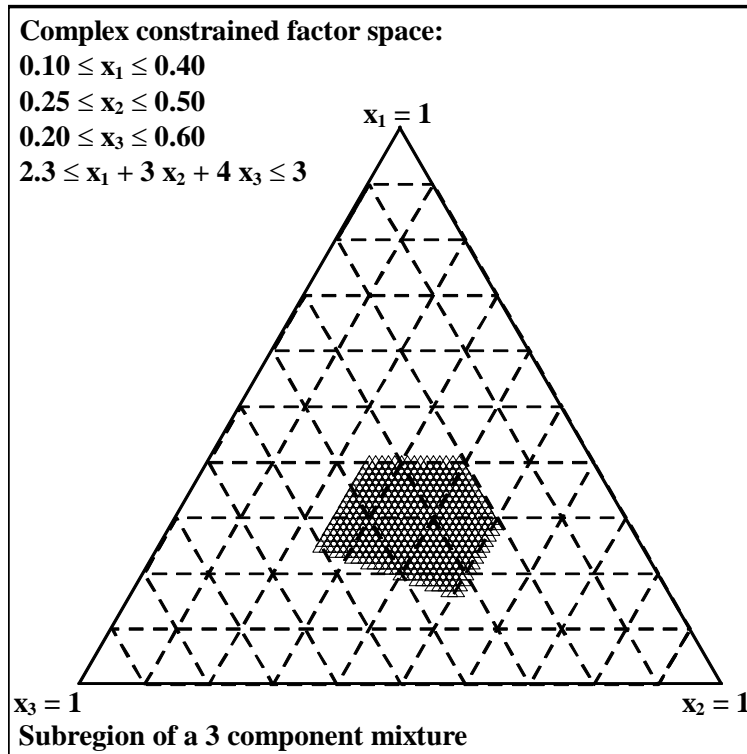
Design and analysis of mixture systems

Additional constraints



Design and analysis of mixture systems

Additional constraints: single and multiple



How to design an optimal experimental design in these experimental regions

Design and analysis of mixture systems

Additional constraints: the experimental design problem

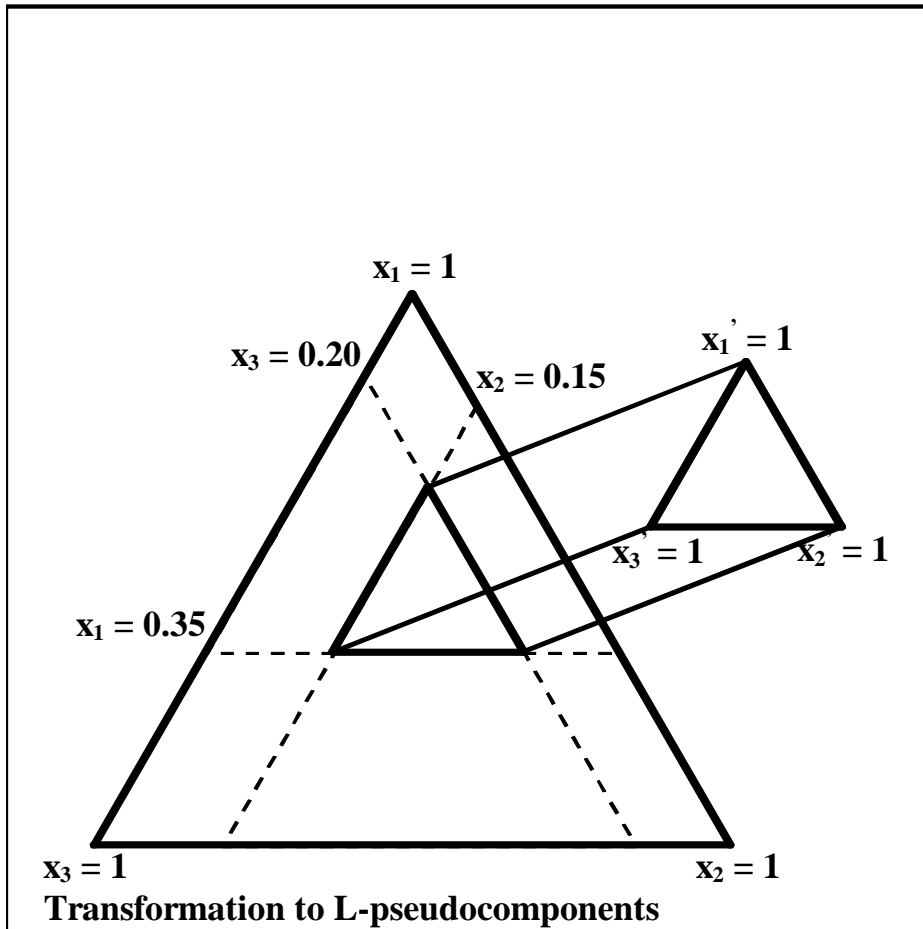
Two situations occur

1. The experimental region is homomorphic with the whole simplex
2. The experimental region is a convex, irregular hyperpolyhedron

Design and analysis of mixture systems

Additional constraints: the experimental design problem

The experimental region is homomorphic with the simplex



Transformation of x_i to pseudocomponents x_i'

$$x_i' = \frac{x_i - l_i}{1 - \sum_{i=1}^q l_i}$$

Projecting the experimental region on the whole simplex

Whole simplex methodology can be used

Whole simplex experiments in raw mixture components or pseudocomponents: an example

- Most 'optimal' approaches for design and analysis of mixture experiments
- Straightforward design and analysis, even for complex problems
- Examples
 - Multigrain crackers

Multigrain crackers

Objectives

- Develop a multiple grain cracker where a proportion p of the wheat flour is replaced with buckwheat, oats, barley and rye
- Optimising the flour composition of a multiple grain cracker in relation to consumer preference, scored by magnitude estimation

Multigrain crackers

Defining constraints of the experimental region

- All components of the dough mixture are constant except flour
- Proportion p of the wheat flour can be replaced by buckwheat, oats, barley and/or rye
 - $1-p \leq P_{\text{wheat}} \leq 1$
 - $0 \leq P_{\text{buckwheat}} \leq p$
 - $0 \leq P_{\text{oats}} \leq p$
 - $0 \leq P_{\text{barley}} \leq p$
 - $0 \leq P_{\text{rye}} \leq p$
- Homomorphic experimental region with 5 components
- Pseudocomponents transformation to whole simplex
- $\{5,2\}$ simplex lattice is proposed

Multigrain crackers

{5,2} simplex lattice design in pseudocomponents

Mixture	Wheat	Buckwheat	Oats	Barley	Rye
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1
6	0.5	0.5	0	0	0
7	0.5	0	0.5	0	0
8	0.5	0	0	0.5	0
9	0.5	0	0	0	0.5
10	0	0.5	0.5	0	0
11	0	0.5	0	0.5	0
12	0	0.5	0	0	0.5
13	0	0	0.5	0.5	0
14	0	0	0.5	0	0.5
15	0	0	0	0.5	0.5

Multigrain crackers

Full model: quadratic Scheffé canonical polynomial

- 5 linear terms
- 10 cross product terms

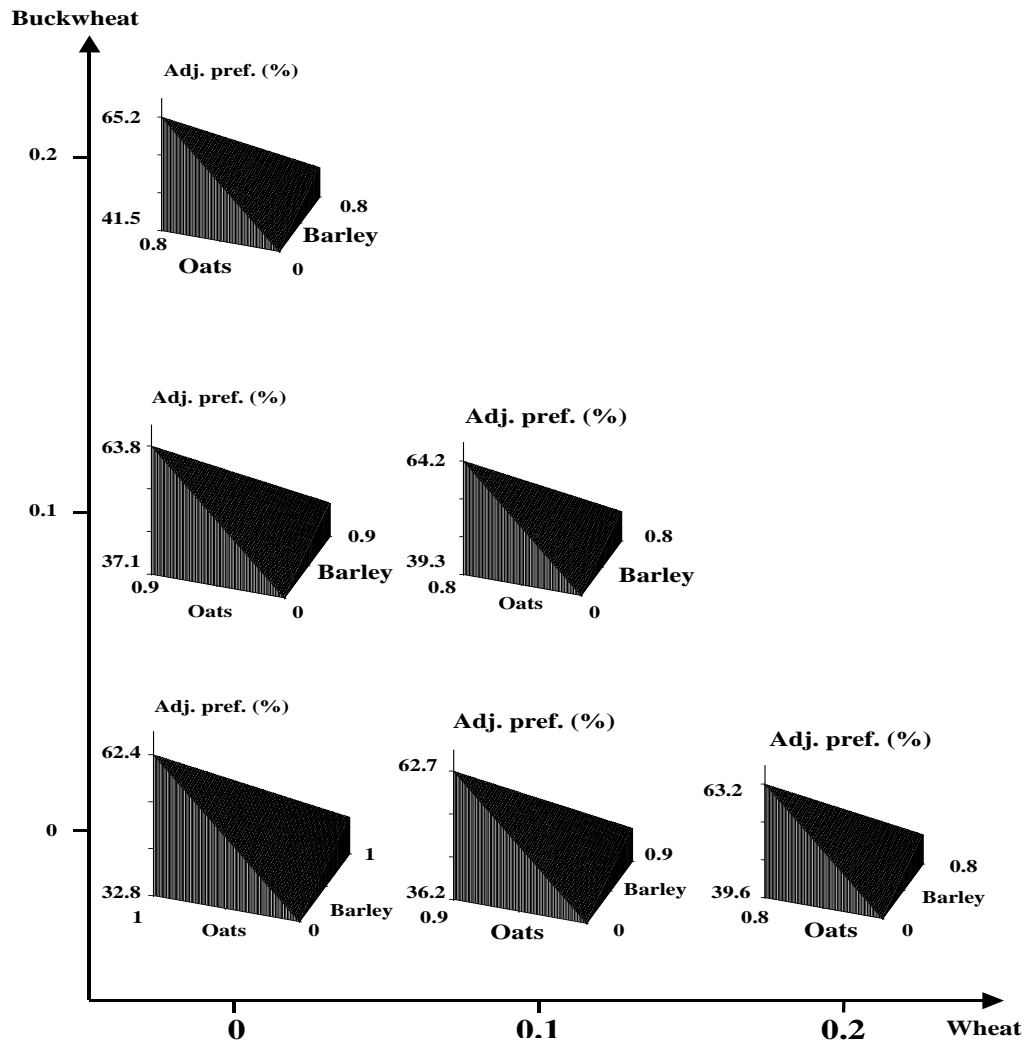
$$\begin{aligned} \text{Pref.} = & \beta_{10000} \text{wheat} + \beta_{01000} \text{buckwheat} + \beta_{00100} \text{oats} + \beta_{00010} \text{barley} + \\ & \beta_{00001} \text{rye} + \beta_{11000} \text{wheat} * \text{buckwheat} + \beta_{10100} \text{wheat} * \text{oats} + \\ & \beta_{10010} \text{wheat} * \text{barley} + \beta_{10001} \text{wheat} * \text{rye} + \beta_{01100} \text{buckwheat} * \text{oats} + \\ & \beta_{01010} \text{buckwheat} * \text{barley} + \beta_{01001} \text{buckwheat} * \text{rye} + \beta_{00110} \text{oats} * \text{barley} + \\ & \beta_{00101} \text{oats} * \text{rye} + \beta_{00011} \text{barley} * \text{rye} \end{aligned}$$

Multigrain crackers

Model reduction based on all possible models and maximum R^2

pref. = 66.62 wheat + 76.51 buckwheat + 62.41
oats + 40.99 barley + 32.81 rye – 129.39
wheat*buckwheat

$R^2 = 0.98$ CV=6 %



Multigrain crackers

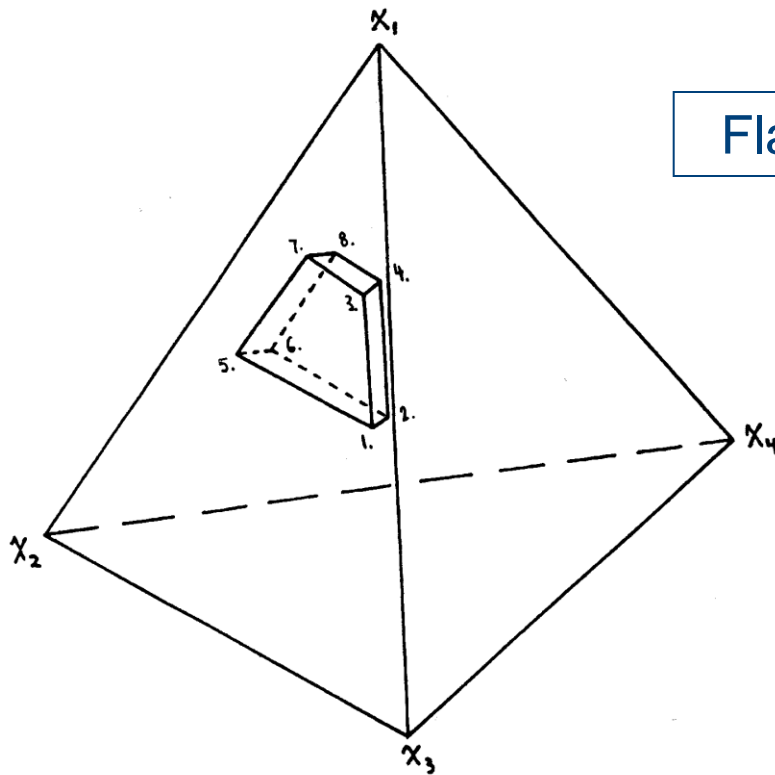
Conclusions

- Replacing the wheat flour partially with buckwheat and oats had a positive effect on the consumers preference
- A antagonistic interaction between wheat and buckwheat persists
- Rye and barley addition resulted in low preference

Design and analysis of mixture systems

Additional constraints: the experimental design problem

The experimental region is an irregular hyper-polyhedron

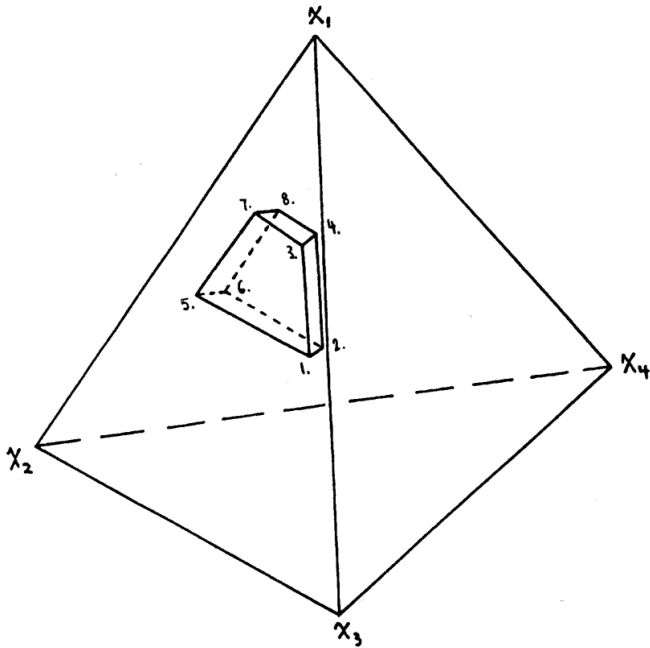


Flare manufacturing experiment

Design and analysis of mixture systems

Additional constraints: the experimental design problem

The experimental region is an irregular hyper-polyhedron



A specific optimal experimental design has to be developed for each individual case

Computer Aided Design of Experiments is necessary (CADEX)

Extremely computational intensive in high dimensionality

Additional constraints: the experimental design problem

The experimental region is an irregular hyper-polyhedron
CADEX approach

Construction of a list of candidate treatments

Extreme vertices and centroids of all the lower dimensional boundary hyperplanes, planes and edges of the convex, irregular hyper-polyhedron

For some models equidistant grids can be used as candidate list

Assuming a model form

Defining an optimal design criterion

D-optimality

G-optimality

Optimisation algorithm

Extremely computational intensive exchange algorithms

Run the optimisation for different numbers of treatments

Additional constraints: the experimental design problem

The experimental region is an irregular hyper-polyhedron
CADEX approach (continued)

Choose a design with optimal properties

Implement possible randomisation schemes

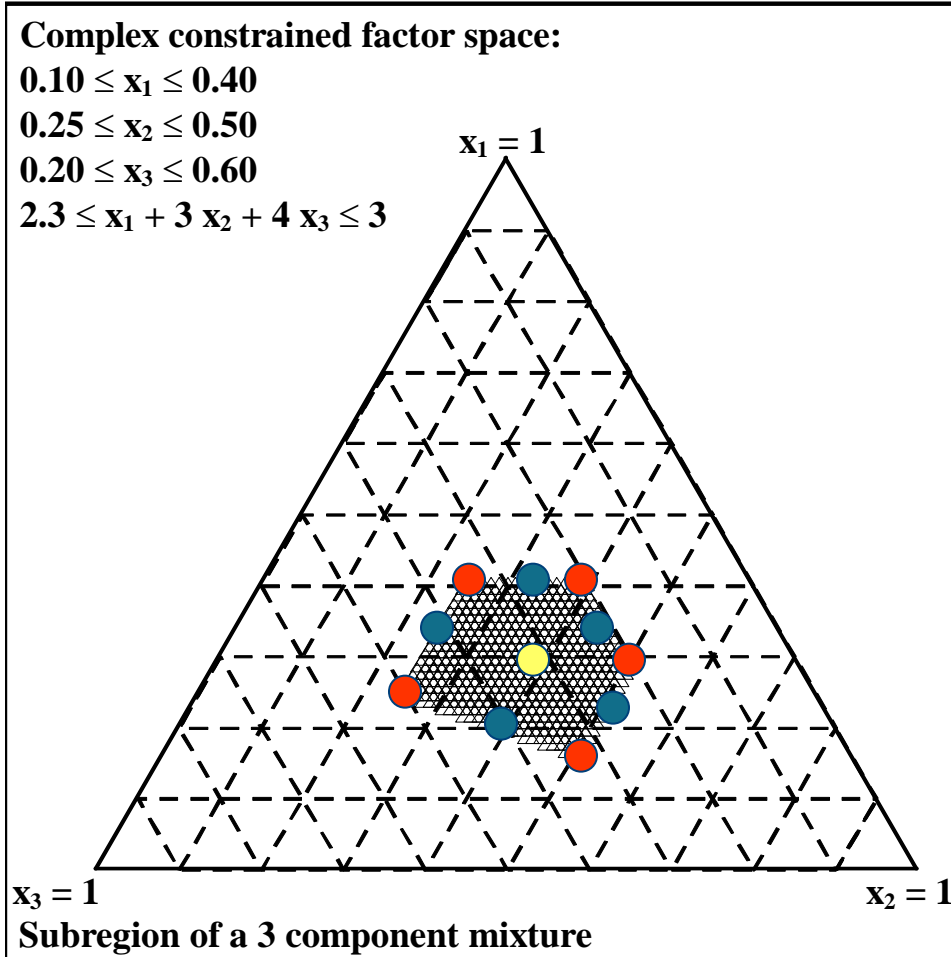
Carry out the experiment and do the measurements

Fit the model by classical least squares (GLM)

Interpretation

Additional constraints: the experimental design problem

The experimental region is an irregular hyper-polyhedron
Calculation of the candidate list



- Vertices
- Edge centroid
- Plane centroidvertices

The method of least squares

The polynomial models can be formulated in matrix notation:

$$Y = X\beta + \varepsilon$$

Where Y is an $n \times 1$ vector of observations on the independent variable, X equals an $n \times p$ matrix of known factor levels for each individual component, including cross-product and quadratic terms, β is a $p \times 1$ vector of unknown parameters, n is the number of experimental units and ε is an $n \times 1$ vector of random errors.

In the case where the $(X'X)$ matrix is not singular, the least squares estimation of the parameters b of β is given by:

$$b = (X'X)^{-1}X'y$$

The variance-covariance matrix of b is expressed in the following equation:

$$\text{var}(b) = \sigma^2 (X'X)^{-1}$$

with σ^2 the error variance. The elements of the matrix $(X'X)^{-1}$ are proportional to the variance and the covariances of the elements of b .

The variance of the prediction in a specific point x is given by:

$$\text{var}(xb) = \sigma^2 x'(X'X)^{-1}x$$

Optimisation criteria for selecting optimal designs in constrained regions

Two situations occur:

Comparing different experimental designs in their precision to estimate parameters in a RSM

$$\text{var}(\mathbf{b}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Minimizing $\det\{(\mathbf{X}'\mathbf{X})^{-1}\}$ \longrightarrow D-Optimality

Comparing different experimental designs in their precision to estimate predicted response in a RSM

$$\text{var}(x\mathbf{b}) = \sigma^2 \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}$$

Minimizing $\max\{\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\}$ \longrightarrow G-optimality

Design and analysis of mixture systems

Process variables

If a mixture experiment consists besides of the mixture components out of other factors, not bound by the mixture constraint, these are called “process variables”. Changing the process variables may effect the blending properties of the mixture components.

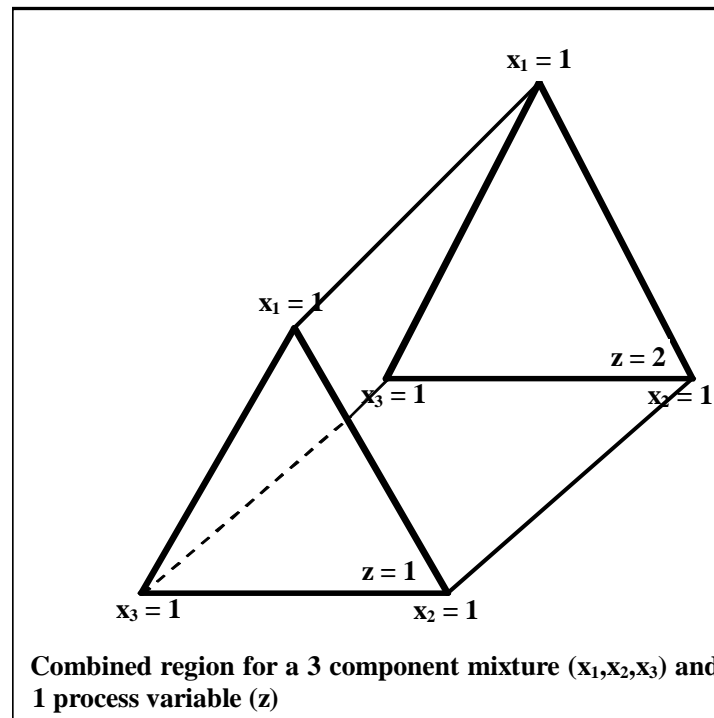
The experimental region of a mixture experiment with process variable(s) is the combined region of the mixture components and the process variable(s).

The dimensionality of the combined experimental region equals the sum of the dimensionality of the separate experimental regions

Design and analysis of mixture systems

Process variables

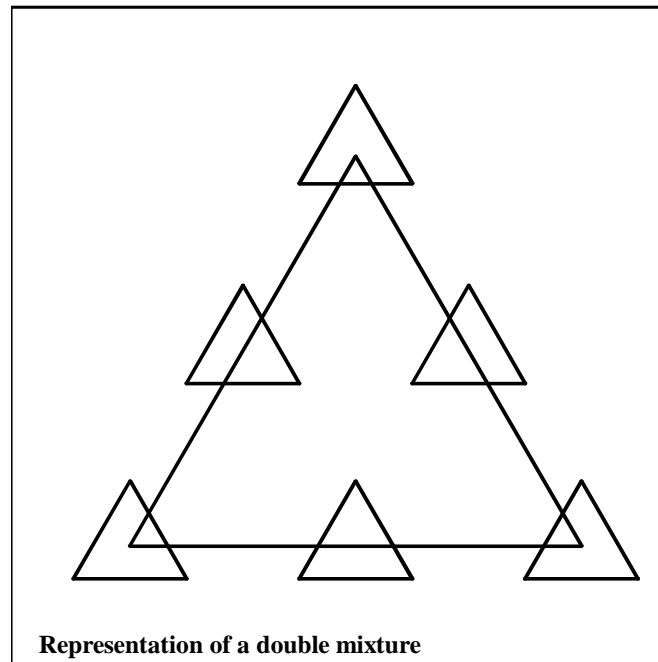
The 3 dimensional prism is the combined experimental region of the two-dimensional simplex (x_1, x_2, x_3) and the one-dimensional experiment in process variable z .



Design and analysis of mixture systems

Multiple mixture systems

A multiple mixture can be defined as a mixture of different other mixtures



Design and analysis of mixture systems

Multiple mixture systems

Process variables

Designs for sub-experiments are combined

Specific adaptations of mixture models are necessary

Composed liquid fertilizer as a constrained, double mixture amount

Objectives

Optimize the composition of a composed liquid fertilizer in relation to plant response (Ca content of Rye grass)

Composed liquid fertilizer as a constrained, double mixture amount

Defining constraints

Aqueous solutions of inorganic ions



with C = total concentration in units of charge (meq/l)

Prepared by dissolving salts



Ionic balance constraint: balance of charge

$$[\text{K}^+] + [\text{Ca}^{2+}] + [\text{Mg}^{2+}] = [\text{NO}_3^-] + [\text{H}_2\text{PO}_4^-] + [\text{SO}_4^{2-}] = C/2$$

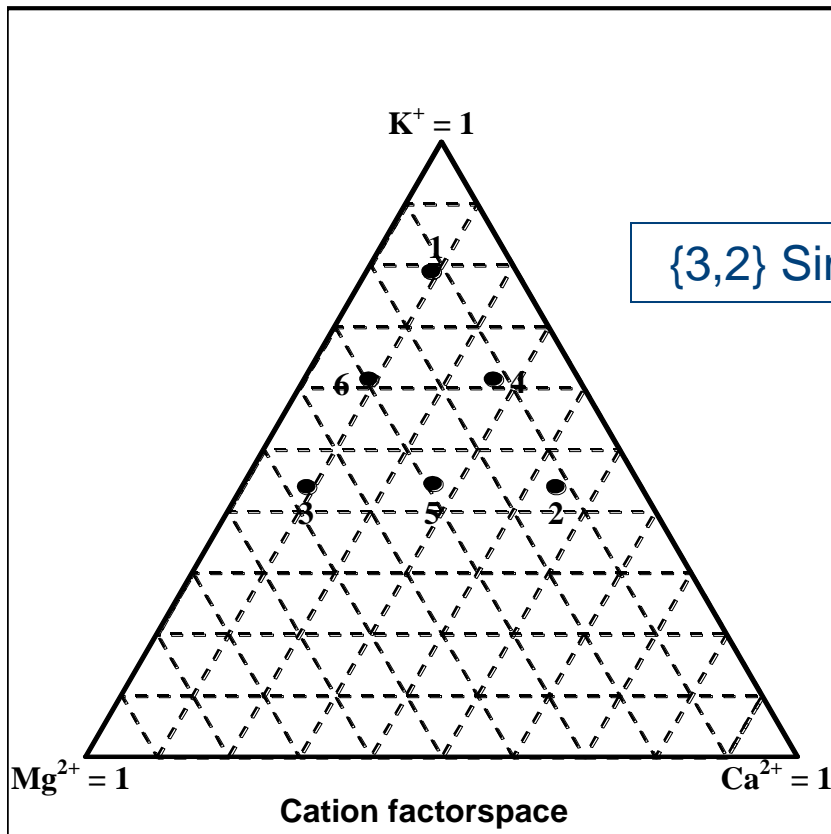


Double mixture constraints: multicomponent equality constraints

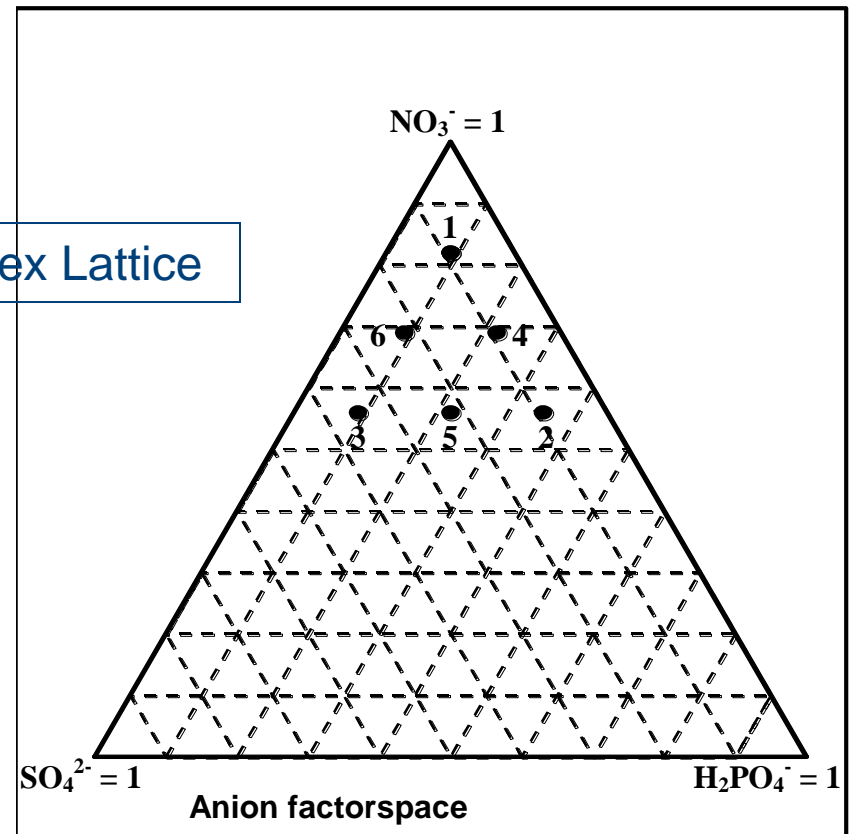
Composed liquid fertilizer as a constrained double mixture amount

$$\begin{aligned} 0.44 &\leq K^+ \leq 0.79 \\ 0.09 &\leq Ca^{2+} \leq 0.44 \\ 0.12 &\leq Mg^{2+} \leq 0.47 \end{aligned}$$

$$\begin{aligned} 0.56 &\leq NO_3^- \leq 0.82 \\ 0.09 &\leq H_2PO_4^- \leq 0.35 \\ 0.09 &\leq SO_4^{2-} \leq 0.35 \end{aligned}$$



{3,2} Simplex Lattice



Composed liquid fertilizer as a constrained, double mixture amount

- Mixtures ($\{3,2\}$ SimLat) in the separate mixtures are transformed to pseudocomponents
- Total concentration (C) is defined as a process variable
- The mixture design is the combination of the two constituent mixtures (Cations, Anions), mixed with proportion 0.5
- This design is repeated at the levels of the process variable (C = total concentration or amount)

Composed liquid fertilizer as a constrained, double mixture amount

- Pseudocomponents transformations

$$k = \frac{\text{K}^+ - 0.44}{0.35}$$

$$n = \frac{\text{NO}_3^- - 0.56}{0.26}$$

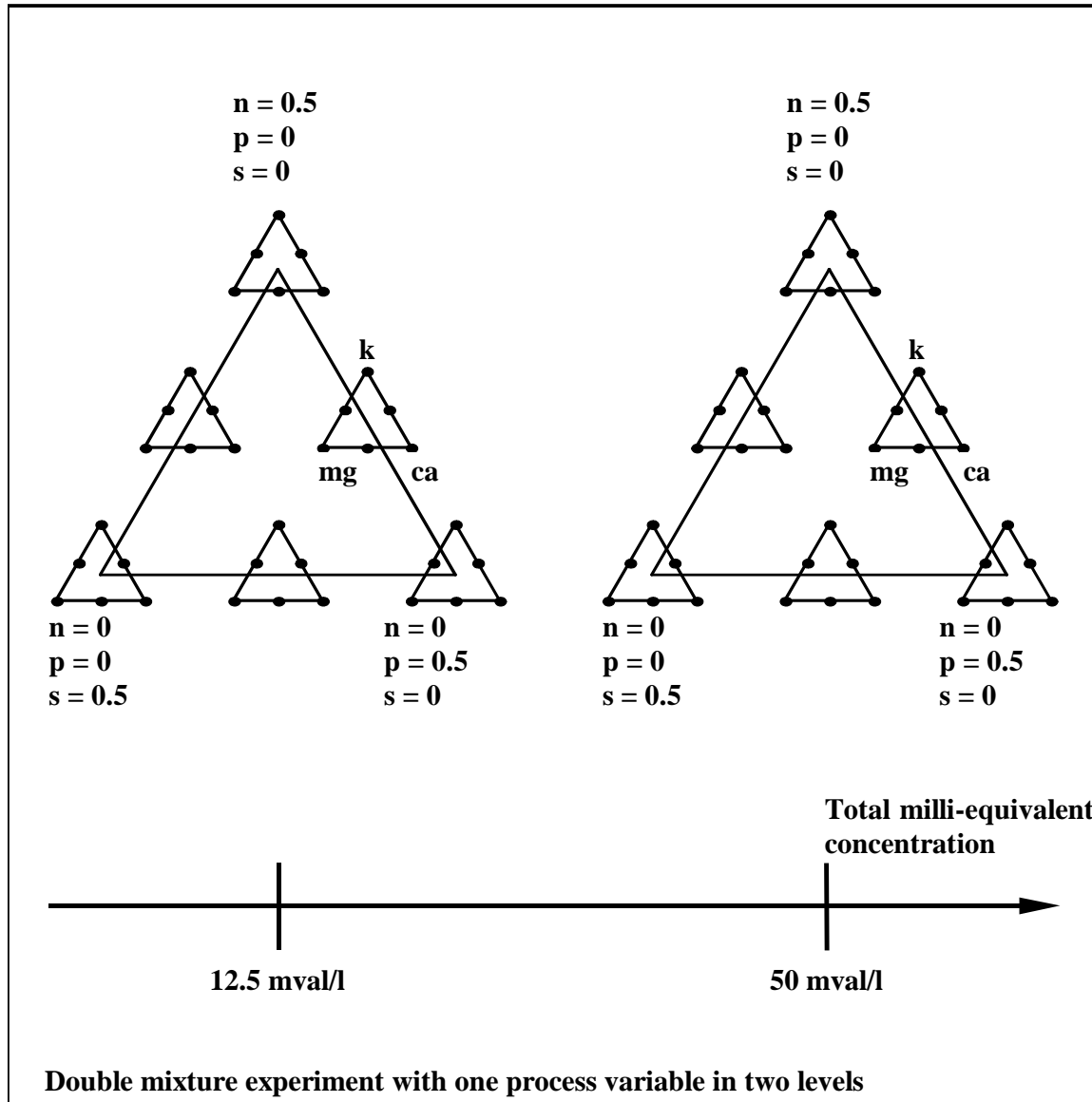
$$ca = \frac{\text{Ca}^{2+} - 0.09}{0.35}$$

$$p = \frac{\text{H}_2\text{PO}_4^- - 0.09}{0.26}$$

$$mg = \frac{\text{Mg}^{2+} - 0.12}{0.35}$$

$$s = \frac{\text{SO}_4^{2-} - 0.09}{0.26}$$

An example of a double mixture with one process variable



Composed liquid fertilizer as a constrained, double mixture amount

Number of treatment combinations

- Cations x anions: $6 \times 6 = 36$
- 2 levels in the process variable: $36 * 2 = 72$
- 5 replications per treatment: 360 exp units

- Checkpoints?

Composed liquid fertilizer as a constrained, double mixture amount

Model formulation

- The second degree double mixture model with process variable is obtained by combining the process variable model with a second degree double mixture model, resulting in a total of 72 terms

Composed liquid fertilizer as a constrained, double mixture amount: model formulation

$$\begin{aligned}
 f(x,z) = & b^0_{100100} kn + b^0_{100010} kp + b^0_{100001} ks + b^0_{100110} knp + b^0_{100101} kns + b^0_{100011} kps + \\
 & b^0_{010100} can + b^0_{010010} cap + b^0_{010001} cas + b^0_{010110} canp + b^0_{010101} cans + b^0_{010011} caps + \\
 & b^0_{001100} mgn + b^0_{001010} mgp + b^0_{001001} mgs + b^0_{001110} mgnp + b^0_{001101} mgns + b^0_{001011} mgps + \\
 & b^0_{110100} kcan + b^0_{110011} kcap + b^0_{110001} kcas + b^0_{110110} kcanp + b^0_{110101} kcans + b^0_{110011} kcaps + \\
 & b^0_{101100} kmgn + b^0_{101010} kmgp + b^0_{101001} kmgs + b^0_{101110} kmgnp + b^0_{101101} kmgns + \\
 & b^0_{101011} kmgps + b^0_{011100} camgn + b^0_{011010} camgp + b^0_{011001} camgs + b^0_{011110} camgnp + \\
 & b^0_{011101} camgns + b^0_{011011} camgps + b^1_{100100} knz + b^1_{100010} kpz + b^1_{100001} ksz + b^1_{100110} knpz + \\
 & b^1_{100101} knsz + b^1_{100011} kpsz + b^1_{010100} canz + b^1_{010010} capz + b^1_{010001} casz + b^1_{010110} canpz + \\
 & b^1_{010101} cansz + b^1_{010011} capsz + b^1_{001100} mgnz + b^1_{001010} mgpz + b^1_{001001} mgsz + b^1_{001110} \\
 & mgnpz + b^1_{001101} mgnsz + b^1_{001011} mgpsz + b^1_{110100} kcanz + b^1_{110010} kcapz + b^1_{110001} kcasz + \\
 & b^1_{110110} kcanpz + b^1_{110101} kcansz + b^1_{110011} kcapsz + b^1_{101100} kmgnz + b^1_{101010} kmgpz + \\
 & b^1_{101001} kmgsz + b^1_{101110} kmgnpz + b^1_{101101} kmgnsz + b^1_{101011} kmgpsz + b^1_{011100} camgnz + \\
 & b^1_{011010} camgpz + b^1_{011001} camgsz + b^1_{011110} camgnpz + b^1_{011101} camgnsz + b^1_{011011} camgpsz
 \end{aligned}$$

$b^z_{kca mg n p s}$: parameters

With x : mixture variable
 z : process variable
 k, ca, mg, n, p and s : pseudocomponents in proportions

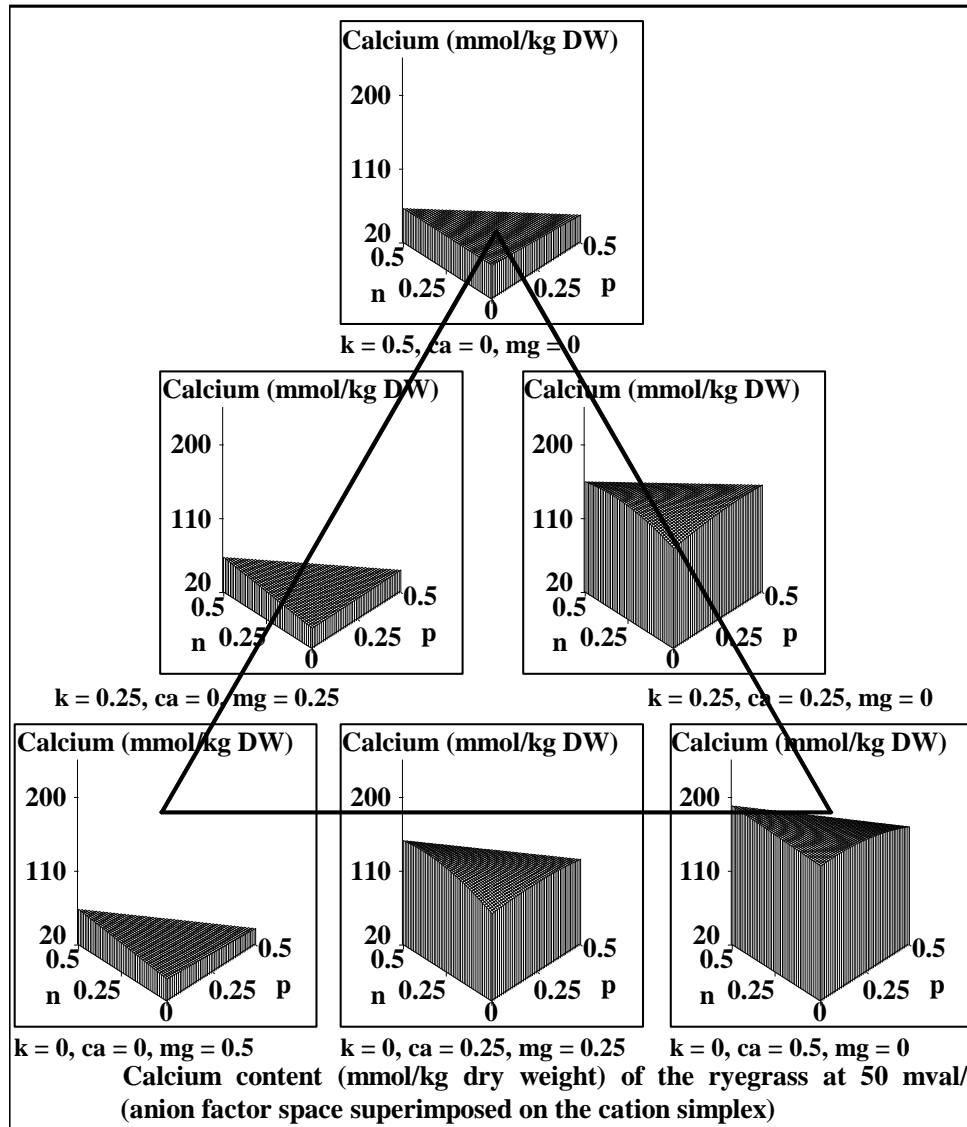
Composed liquid fertilizer as a constrained, double mixture amount

Reduced Model 25 terms:

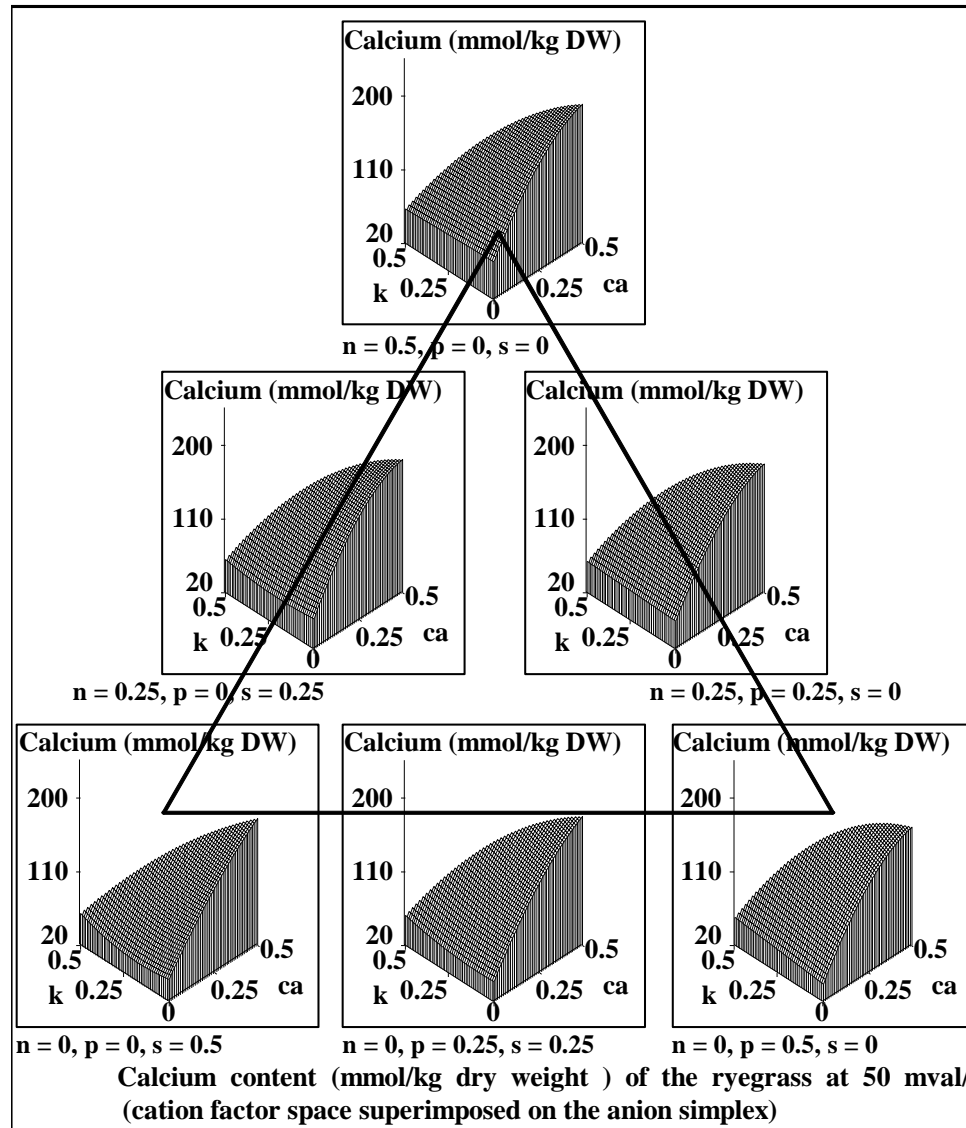
$$\begin{aligned} \text{calcium} = & 211.65 \text{ kn} + 217.64 \text{ kp} + 237.35 \text{ ks} + 672.47 \text{ can} \\ & + 657.87 \text{ cap} + 700.02 \text{ cas} + 245.40 \text{ cap} \\ & + 176.66 \text{ mgn} + 125.13 \text{ mgp} + 123.59 \text{ mgs} \\ & + 928.30 \text{ kcan} + 542.34 \text{ kcap} + 2437.85 \text{ kcans} \\ & - 282.50 \text{ kmgs} + 661.36 \text{ camgn} + 716.75 \text{ camgp} \\ & + 318.91 \text{ camgs} + 2418.63 \text{ camgns} + 37.37 \text{ knz} \\ & + 88.21 \text{ canz} + 78.72 \text{ mgnz} + 36.88 \text{ mgpz} \\ & + 60.66 \text{ mgasz} + 795.67 \text{ kcapz} + 573.17 \text{ kcasz} \end{aligned}$$

$$R^2 = 0.99, \text{ CV} = 4.9$$

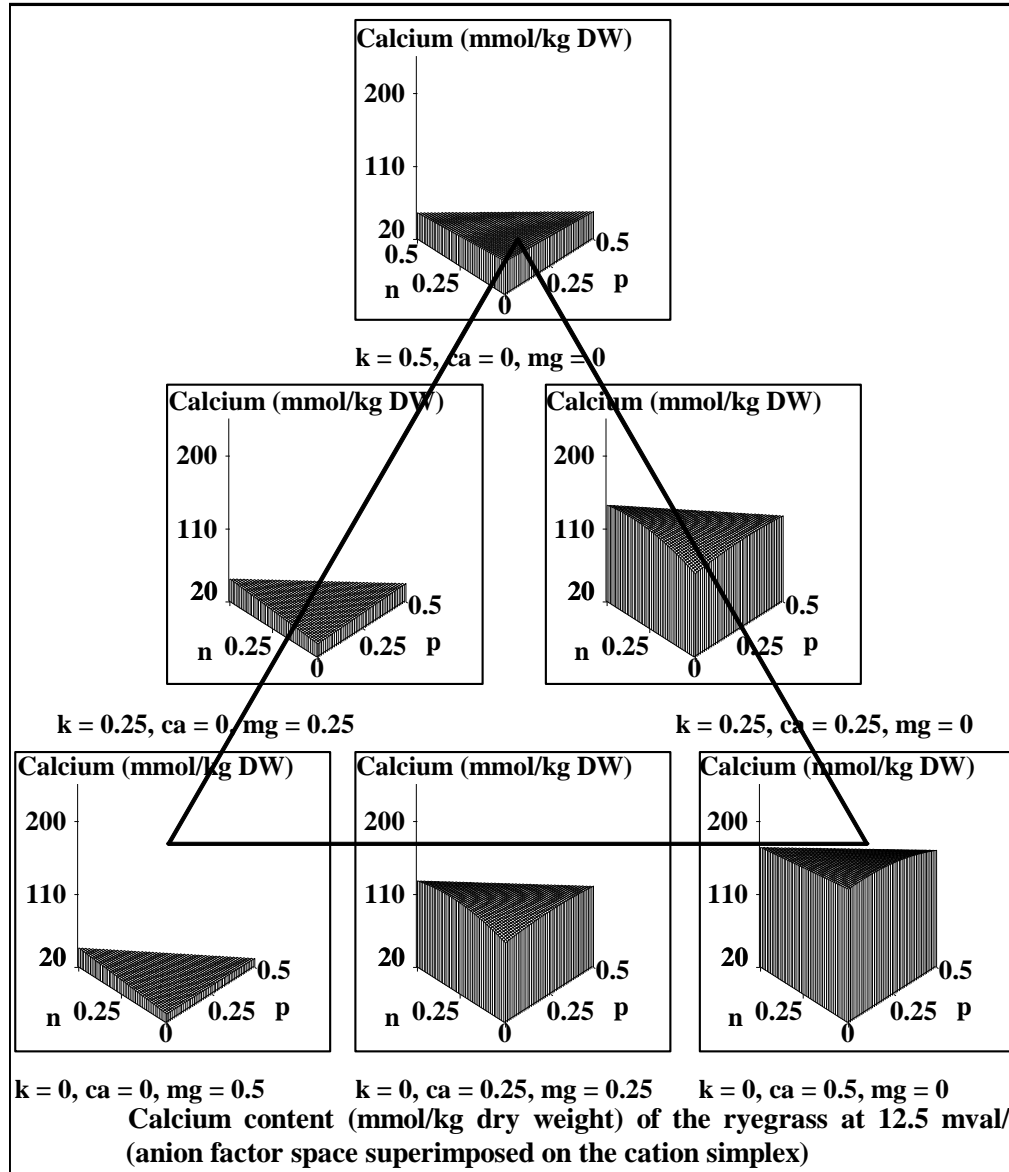
An example of a double mixture with one process variable: 50 mval/l



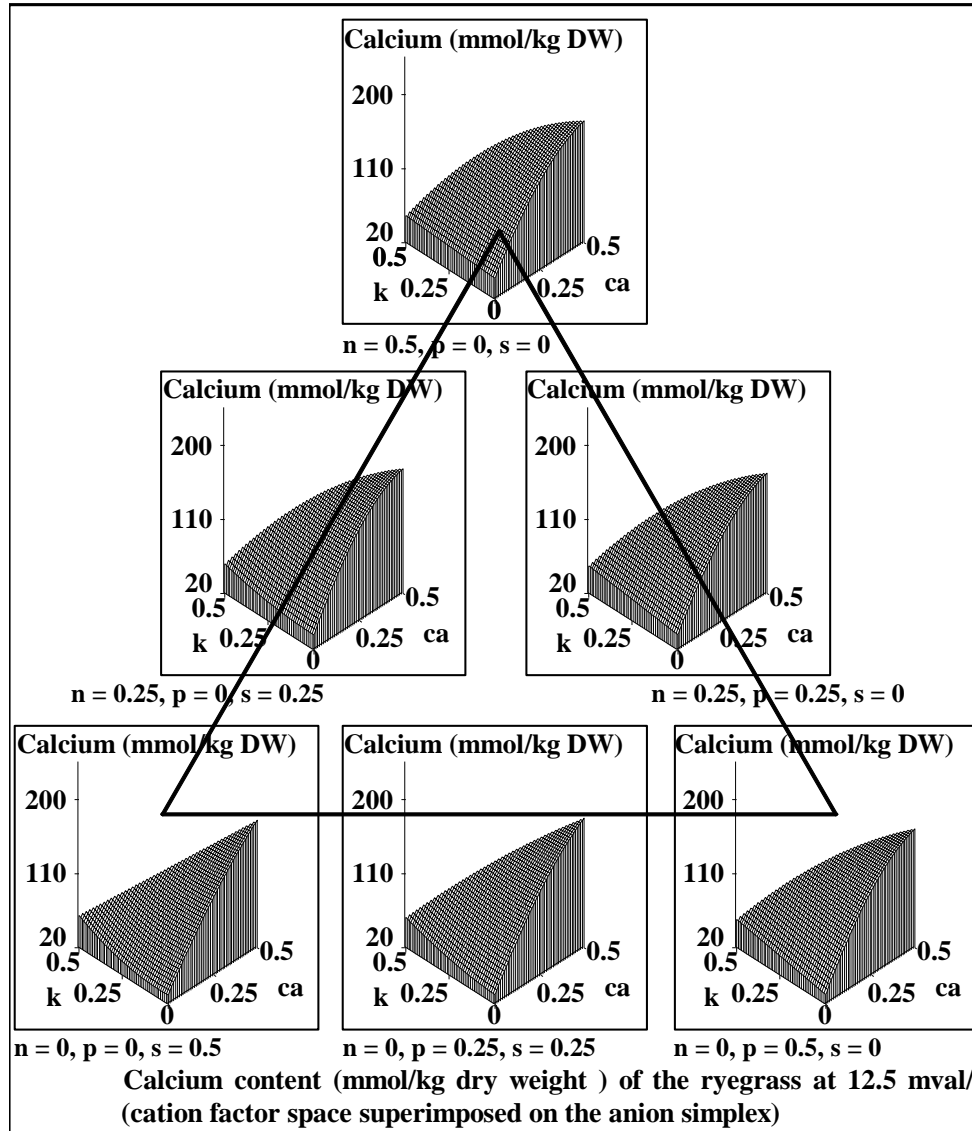
An example of a double mixture with one process variable: 50 mval/l



An example of a double mixture with one process variable: 12.5 mval/l



An example of a double mixture with one process variable: 12.5 mval/l



Composed liquid fertilizer as a constrained, double mixture amount

Conclusions

- Ca in the fertilizer increases the Ca content in the Rye grass
- Marked Ca*K antagonistic effect
- Small Mg*Ca antagonistic effect
- No cation – anion interactions
- Total concentration has an additive effect, not interacting with composition

Design and analysis of mixture systems

Conclusions

Experimental factors in constrained regions of interest are always confounded

Experimentation in constrained experimental regions demands adapted strategies to develop optimal experimental designs and specific model forms

Estimation of model parameters is straightforward RSM

Multiple mixtures, process variables, additional constraints, ... result in complicated designs and models