**A comparison of methods for analyzing multivariate sensory data in designed experiments – a case study from salt reduction in liver paste.**

**Tormod Næs\*^, Oliver Tomic\*, Kirsti Greiff+, Kari Thyholt#**

\*Nofima mat, Oslovegen 1, 1430 Ås, Norway

^Dept of Food Science, University of Copenhagen, Denmark

+SINTEF, Fisheries and Aquaculture Ltd., P.O.Box, 7465, Trondheim, Norway

#Mills DA, PO-Box 4644 Sofienberg, 0506 Oslo, Norway

**Abstract**

This paper is a comparison of different methods for analyzing designed experiments. The tools used are based on PCA, PLS and ANOVA, either used separately or in combination. Special emphasis will be on how to obtain information about the medium and less important factors in the presence of very dominating ones. It will be shown that there exist possibilities for this based on splitting the dataset in two. The method comparison will be based on a data set obtained for studying the effect of salt reduction in liver paste.

**Key words.** PCA, ANOVA, PLS regression, ASCA, designed experiments.

**1. Introduction**

Designed experiments play an important role when developing new products in industry. Typically one is interested in how different formulations and different process conditions influence the properties of the end product. These properties can be technological, health related or related to taste and odour. In most cases one is interested in several aspects of the output at the same time, i.e. one is interested in multivariate output data. In addition to improved insight one is typically also interested in optimizing the responses for the purpose of satisfying for instance consumer needs and wishes. The focus of the present paper is methodology for obtaining improved insight.

A number of different methods exist for analyzing multivariate output in designed experiments. The classical approach is multivariate analysis of variance (MANOVA, see e.g. Mardia et al. (1979)) which provides tests of significance of the different input factors on the whole vector of responses. This is useful as a starting point, but it provides little insight into correlation structures among the output variables and similarities and differences between the objects. For this purpose, it is often advantageous to use data compression methods which project the response data onto a subspace and relate these few components to the design of the experiment by analysis of variance (ANOVA). A possible way of doing this is to use PC-ANOVA (principal components-analysis of variance, see e.g. Ellekjær et al. (1996), Luciano and Næs (2009) and Næs et al. (2010)), which is based on first using principal component analysis (PCA) of the multivariate response values followed by regular ANOVA of the first few components vs. the design. In some cases, a simple visual inspection of the PCA plots may be sufficient (Baardseth et al. (1992)), but often the extra ANOVA step makes interpretation easier. The advantages of this type of methodology are many as described in for instance Luciano and Næs (2009). The main drawback is that if the different design factors give rise to very different correlation structures among the responses, the information in the first few components may be too complex. A possible remedy to this is to use the ASCA (analysis of variance – simultaneous components analysis, see e.g. Jansen et al. (2005)) method which reverses the two operations of ANOVA and PCA. The method first estimates the effects of the factors for each response variable using regular ANOVA averaging techniques and then uses a PCA on the individual effects matrices separately. The method provides plots that are easy to interpret, but provides more limited information about significance of the effects. A strongly related method (ANOVA-PCA, Harrrington et al. (2005)) adds residuals to the effects before running the PCA and can in this way be used to obtain an indication also of the uncertainty in the PCA plots. A third possibility is to use partial least squares (PLS) regression of the response variables onto the design variables represented by dummy variables. This gives one score plot and two loadings plots, one for the responses and one for the design variables. Cross-validation can be used for assessing overall performance of the model, but more specific information about the significance of the design factors on the component scores is more difficult to obtain by this approach.

The purpose of this paper is to compare and discuss advantages and disadvantages of the approaches mentioned in a situation of product development in the food sector. In addition to discussing and comparing these established approaches, we will also propose a couple of new variants which can be useful when the focus is on special aspects of the design and how these interact with the other factors. Focus will also be on how further fine-tuning of methodology can provide improved insight into the effect of moderately important factors in the presence of very dominating ones. The case study presented is based on a designed experiment conducted for the purpose of investigating the effect of reducing salt level in liver paste without losing important sensory properties (based on sensory profiling). The design used is a full factorial design based on 16 combinations of 4 factors. Special focus will be on the salt effect and possible interactions with the other factors. Some replicates are also present and some attention will be given to how these can be used for assessing the validity of the experiment.

**2. Methodology**

**2.1 General framework.**

The response data matrix is here written as **Y.** The columns represent the sensory variables and the rows represent the experimental runs or products. The design matrix is called **X.** The **Y** matrix can as usual be seen as a function of the design of the experiment **Y=f(X)** plus noise **E.** Since the design considered is a two-level design, we will here only consider the linear model situation, i.e. the model

 **Y=XB+E (1)**

In some cases we will, however, also add interaction terms to the model, which corresponds to extending the **X** matrix with some new columns. Although focus will here be on factorial designs, some of the methods discussed can easily be extended to other designs.

**2.2. PC-ANOVA**

The main ingredients of PC-ANOVA are PCA and ANOVA (Ellekjær et al. (1996)). The PCA is first run for all the response variables **Y** or only a subset of them if one is interested in a special focus. In this way one obtains scores **T** and loadings **P,** which can be plotted and interpreted directly by using standard scatter plots. The scores from the PCA model **Y=TP’+F** are then used as dependent variables in an ANOVA model with **X** as the independent variables (**T=XA+E)**. Note that one then has full freedom to define the ANOVA model according to which design that is underlying the study (included interactions) and also which error structure that is the right one to use (split-plot, repeated measures etc., see Luciano and Næs (2009)). Note also that all tools available for regular ANOVA will be available here (LS-means, random effects, post-hoc testing etc.). One of the main advantages of this approach is its simplicity and the fact that it can be run in whatever statistical package one uses. In particular in situations with many objects the additional ANOVA step can be very useful to aid interpretation of the effects. If the two models are put together, one can write the result as a regular regression model (as model (1))with **B=AP’**.

In Luciano and Næs (2009) it was also proposed to provide scores plots based on averages for levels of the most interesting factors in order to highlight the factor effects in the same multivariate space as the raw data. Note that this is identical to averaging raw observation vectors for each of the actual design variable levels and then projecting the averages onto the principal components. In the same paper it was also proposed to incorporate line segments associated with the principal component (PC) axes with a length indicating for instance the least significant differences (LSD’s) or the square root of the random error. This extra information can be very useful for assessing visually the importance of the effects seen in the plot. In this paper the latter of these options will be illustrated.

A sophisticated modification of the standard PC-ANOVA was proposed in Langsrud (2002). This is called 50/50 MANOVA and is based on an interesting splitting of the information in both data blocks for the purpose of significance testing. The method can be accompanied by plotting procedures based on PCA in more or less the same way as will be discussed here.

When concerns replicates of some of the samples, there are several possibilities. One possibility is to incorporate all data in the PCA and also in the ANOVA, in this way utilizing the extra data for improved precision (Myers and Montgomery (1995)). This will, however, lead to unbalanced models, possibly with repeated measurement error structure, and one needs to be a bit more cautious in the analysis. If only few replicates are available, the improvement will be marginal. A better possibility is then to simply project the replicates onto the PCA space for the purpose of assessing the variation visually. The focus is then not on improved precision, but merely on information about the uncertainty in the measurements and experimentation. Note that the best is usually to use replicates of the whole experimental procedure from mixing ingredients to measurements. Only looking at measurement replicates (repeated measurements) by for instance taking a new measurement of the same sample is usually not satisfactory.

Note that the use of ANOVA on the scores also implicitly represents a type of validation of the PCA model. If some of the effects are clearly significant and the effects look reasonable, this supports the validity of the PCA model. Cross-validation (CV) is also sometimes used for validation of PCA models, but in small design experiments with possibly unique samples this is questionable. In this case study, however, it seems that there is enough similarity between the objects to give reasonable CV results for the PCA. Another possibility for validation is the bootstrap or other randomization techniques such as permutation tests (Endrizzi et al. (2013)), but this will not be considered further here.

An extra tool that should be mentioned here is the use of passive variables in the analysis. This means that also the design variables (dummy) and interactions are incorporated in the PCA, but in such a way that they are given very low weight. This means that they do not influence the analysis, but they will still play a meaningful role for interpretation of the correlation loadings plot (Martens and Martens (2001)). This will be illustrated here for the PCA of the full data set.

It should be mentioned that because each principal component is a linear combination of original variables, the normality assumption can be expected to be better for the components than for the original variables (see also Luciano and Næs (2009)).

**2.3. ASCA related methods**

The main advantage of ASCA (Jansen et al. (2005), Jansen et al(2008)) as opposed to PC-ANOVA is that it can handle different correlation structure for the different design variables. If for instance the variation of **Y** associated with design variable(s) in **X1** is completely different from the variation associated with design variable(s) in **X2** (where the **X1** and **X2** together span the same space as **X**), the regular PCA plot is a compromise of the two effects and then possibly difficult to interpret or lead to many components to be interpreted. The idea is then instead to switch the order of the operations, i.e. to use ANOVA to first estimate the effects in the model and then to use PCA for each of the effect matrices separately. In more detail, if the model is

**Y=X1B1+X2B2+E** (2)

the first step is to estimate the coefficients **B1** and **B2** and then to use PCA for each of the effect matrices,  and  separately. This means that each  is decomposed as **TiPi** and the results plotted as usual. This method is also very flexible with respect to model used, error structure etc. A similar procedure is proposed in Harrington et al. (2005) (see also Zwanenburg et al. (2011)) in which the residuals are added to the effects before the PCA. This is useful and gives an indication of the uncertainty of the scores, but this is not a real statistical test of significance since the variability of the estimated effects is not explicitly taken into account. Note that additional interaction effects can be easily added to model (2) by just incorporating extra matrices in the model.

In our situation with a two-level factorial design, making a PCA of each estimated effects matrix would represent a PCA of a matrix with essentially only 2 different rows. This analysis is not possible, and in such cases it is natural to merge two or several factor effects for the purpose of doing a PCA. This means that a slight and simple modification of the original ASCA needs to be used in this case. One possibility is to just add two or several estimated main effects matrices and do a PCA of the sum of the effects.

If we take two design variables as an example, the level effects can be written as , for the first factor and and  for the second (different values for each response variable). The four combinations (for each response variable) that correspond to the four level combinations are thus , , and , corresponding essentially to the four “products” (i.e. lines) along the vertical axis of the data table. Each column still corresponds to each of the variables measured.

An even simpler possibility is to calculate the raw average values for the four combinations of the two (or more) factors. The difference from above is that these values will contain information about both main effects and interactions. In this sense it resembles the information obtained in a so-called interaction plot with all the four values represented along the same axis (see e.g. Figure 1 for an example of an interaction plot). After averaging, regular mean centering is done and PCA is run as usual. This approach has the advantage that it looks at all the joint effects of the two (or more) factors at the same time instead of splitting into main effect and interactions. As can be seen, this approach is a kind of compromise between PC-ANOVA and ASCA. It uses the ASCA advantage of using the PCA on different subsets of the design while it resembles PC-ANOVA since it does not estimate effects separately, it only projects averages onto the PCA space. This modification has not been used before as far as we know and will be the approach taken here instead of the original ASCA.

If replicates are present, they are most easily incorporated here for the purpose of estimating the coefficients.

**2.4. PCA for subsets of the data**

The procedure proposed here is applicable only for full factorial designs and in fractional designs when enough design variables have been found to be of no statistical significance. The requirement needed is that when the only the important/significant variables are considered and the other ones disregarded, the design must be a full factorial design. The first step of this procedure is to split the **Y** dataset into two parts according to the level of a factor of particular interest. This creates two Y-matrices **Y1** and **Y2** which are based on the same design except that the variable focused on is at two different levels. All points in the two score plots will thus have a parallel in the other. In this paper the data set is split for the design variable salt level which is the most important design variable for the study and for the protein and carbohydrate blend which is the most dominant (see experimental section below).

The next step is to do a PCA of each of the **Y**-matrices separately. These two PCA’s can also be accompanied with ANOVA testing if wanted and if the design is large enough for running a meaningful ANOVA. Analyzing and interpreting the two separately can be useful in itself, but an extra advantage is obtained if we orthogonalize one of the blocks with respect to the other. In other words, the **Y2** is first orthogonalized with respect to the first few principal components **T1** of **Y1** and then one uses PCA as usual for the orthogonalized matrix . The operation of orthogonalisation can be written as

  (3)

where the  is the projection operator onto the space spanned by the scores **T1** of **Y1**. The orthogonalized **Y2** matrix contains information that is uncorrelated with the scores **T1** of **Y1** and thus contains only information of the additional variability that is introduced when moving from one level to the other of the splitting factor. Note that in its scope this latter method has some similarities with ASCA since it focuses on effects of single variables.

If replicates are available for **Y1**, one can visualize variability as before by simple projection. If the replicates are available for **Y2,** they will need to go through the formula (3) before projection and visualization.

**2.5. Orthogonalize data wrt. design variables**

A somewhat related method to the above is to orthogonalize a whole data set wrt. a design variable (or several). This means that one concentrates on the part of the dataset that cannot be explained by the design variable(s) orthogonalized out. The process is simple and can be obtained by the same formula as in (3) with the only exception that now one uses the whole data set and the orthogonalisation is done with respect to a design variable. Again, the method can be useful for concentrating on the rest of the design variables. The method is similar to the filtering technique in Li et al. (2008). As before, the ANOVA can be used to obtain information about the relative importance of the variables.

**2.6 PLS regression of Y onto X.**

Yet another approach is to use PLS regression of the response vector onto dummy variables representing the design variables and their interactions. In the case of two-level variables there will be one dummy variable for each, with 0 representing the low level and 1 the high level. The advantage of this approach is that it is one-step, but on the other hand it is less transparent how the PLS selects the important information in multivariate designed experiments. Testing is also more difficult and less developed. Some possibilities are were proposed in Martens and Martens (2001) based on the jack-knife (JK).

The PLS method is based on the idea of simultaneously compressing **Y** and **X** using a joint model framework

 **X=TP’+E** and **Y=TQ’+F**. (4)

where the score matrix **T** is the carrier of information from **X** over to **Y**. Note that the symbol **P** is here (as opposed to above) used for the loadings of the input data **X** according to standard PLS notation (see Martens and Næs (1989)). The PLS components are obtained successively, with a deflation between each component, by maximizing the covariance between linear combinations of **Y** and **X.** The scores are obtained first and the loadings are then obtained by regressing **X** and **Y** onto the scores (see Martens and Næs (1989) for more details). If they are put together, the two models can be combined as for the joint model for PC-ANOVA. The difference between this and PC-ANOVA is that here the **T**’s are computed from both **Y** and **X** using a covariance criterion.

A possible shortcoming of this method for some designs is that it cannot handle complex error structures such as split-plot designs. Usually validation is done by CV and some testing possibilities for each individual variable vs. the design can be obtained by the use of the jack-knife (Martens and Martens (2001)). Methods for testing of each factor score vs. the design, as for the PC-ANOVA, have not been developed yet.

Replicates can also here be treated directly or projected down onto the model. In this case, the natural thing to do is to project the replicated **y**-vectors down onto the estimated **Q** in order to get predicted scores for the objects. The formula to use is the projection operator, where **y** now represents a (column) replicate vector. These scores are plotted in the same plot as the fitted scores. Lines should be drawn between the replicates.

**3. Experimental plan**

The design was developed for the purpose of investigating the effect of salt and its interactions with other factors in liver pastes. An experimental design with four factors was set up consisting of 16 experimental runs. The design is presented in Table 1. The runs were randomized in the production. Sample no. 2, 8, 10 and 16 (Table 1) were repeated twice in randomized order in between the experimental runs.

The four factors, each at two levels, consisted of A) amount (weight % of liver paste) of a standardised blend of raw materials rich in proteins and carbohydrates, B) amount (weight % of liver paste) of a standardised blend of raw materials rich in fat, C) amount of salt (weight % of liver paste), and D) baking temperature. The upper and lower levels of A) and B) were adjusted to span the variation of protein, carbohydrate and fat contents of commercial liver pastes in Norway. The factor C) represented the upper level of salt in commercial pastes and no added salt. The rest of the ingredients were kept at a constant level, except from water, which constituted the remaining part of the 100 weight % of the liver pastes. The recipes were optimized in preliminary experiments, in order to avoid irrelevant samples in the design.

Baked liver pastes in aluminum trays were chilled on ice, packaged, stored in a freezing plant (-28⁰C, 3 weeks) and defrosted (5⁰C, 1 day) prior to the sensory analyses. A panel of 6 industrial sensory assessors were calibrated on two reference samples, by agreeing upon scores of attributes. The reference samples were blinded and they were the same as two of the replicated pastes. The following sensory attributes were chosen based on the experience from the preliminary trials: firmness, ordinary salty taste, other salty taste, bitter taste (also called bitterness below), sweet taste (also called sweetness below), spicy/onion taste and liver paste taste. The samples were assessed in 4 rounds, each consisting of 6 samples. There were two reference samples available for the assessors to calibrate against. Only the averages over assessors are here used for the analyses.

**4. Results**

**4.1. Univariate and multivariate ANOVA tests.**

A table of p-values, effect sizes and some of the most important main effects and interactions plots are presented in Table 2 and Figure 1. Only the main effects and the significant interactions are presented in Table 2. As can be seen, interactions are only present for firmness and bitter taste. In both cases there is an interaction between protein+carbohydrate (A) and salt (C). Firmness increases by increased level of both factors. The bitter taste and “other salt” taste decrease by increased protein+carbohydrate level. In other words, firmness and bitter taste go in different directions. As can also be seen from the interaction plot at the bottom of in Figure 1, the effect of salt on firmness is largest for the lowest level of protein+carbohydrate. Likewise for bitterness, the effect of salt is stronger for low level of protein+carbohydrate. As a matter of fact, in both cases, the effect is mainly present for the low level of protein+carbohydrate. Sweet taste does not have any significant relation to the design variables and one could therefore consider eliminating it. It was, however, discovered later on that it showed up quite clearly in some multivariate plots, i.e. when the correlations with other variables are taken into account. It was therefore decided here to keep it as a part of the analysis.

The p-values for the multivariate ANOVA tests are <0.01, 0.25, 0.01, 0.94 for A, B, C and D respectively showing that overall, the factors of main influence are A and C.

**4.2 Standard PC-ANOVA**

The scores and correlation loadings for the PCA are presented in Figure 2. Two components cover most of the variation (91%) in this case. All variables are explained by more than 50% since their correlation loadings fall outside the inner circle (sweetness on the border). Cross-validation (results not shown) supports the results very well here since the two explained variance curves (fitted and cross-validated) follow each other quite closely. The first component explains a contrast between high and low protein+carbohydrate value and also between low salt and high salt content. As can also be seen, the first component is most strongly related to firmness vs. bitterness; the more protein+carbohydrate and salt, the firmer and less bitter liver pastes. The second component is mainly related to low salt vs. high salt level and relates most strongly to the attributes salt and liver paste, but also a bit to“other salt” and spice. The more salt and the less protein+carbohydrate, the more taste of liver paste and salt. These aspects are visible directly in the plot (by comparing with the design in Table 1), but are also verified by PC-ANOVA (Table 3). The significant effects are primarily A and C (strong) for both axes, but also B is significant at 5% level for component 1. As can also been seen, the design variable A has the strongest effect for component 1 and C the strongest effect for component 2. The effect of the protein+carbohydrate and salt effects go in the same direction for component 1 and in the opposite direction for component 2. There are no significant interactions for any of the components, except a small tendency between A and C for the first component. This can be seen visually by the fact that the samples with highest salt content are positioned slightly more to the right for low protein+carbohydrate level than for high level of protein+carbohydrate.

All these results correspond well to the univariate ANOVAs, but in the univariate case, one gets no information about similarity of objects and correlations among variables and about which samples that are characterized by which attributes. This clearly underlines the advantages of a multivariate approach, but a univariate analysis should not be neglected due to its ability to provide detailed information about each single response variable. Also for communication with practitioners, the univariate results are important since they are even easier to explain.

As can be seen the PCA was the main source of information here, but the ANOVA supported the results and helped interpretation. The PC-ANOVA results also supports the validity of the PCA plots as was discussed above. In other cases, where the effects are less pronounced or where the number of samples is higher, the ANOVA can be even more important for interpreting the plots.

In Luciano and Næs (2009), the averages for each of the factor combinations were calculated and plotted in the same plot as the scores. Since the effects are very clear here, this is not needed. With only two levels for each factor, such a practice is also less useful, and one would normally consider two factors simultaneously instead of only one. Note that this is very similar to what will be done below for the modified ASCA method.

Line segments (short lines along the two principal component axes) indicating the random error are also added to the plot in Figure 2 as was suggested in Luciano and Næs(2009). These line segments are important for simultaneously looking at the noise level and variation among the samples. Alternatively one can use the least significant differences (LSD) values for multiple comparisons. In this case, the random noise is quite low as compared to the actual variability and effects seen.

In this experiment some of the samples were repeated. In order to highlight the experimental error, these samples were projected onto the principal component directions. The replicates are joined by line segments in the plot. As can be seen, the replicates are relatively close to each other supporting the validity of the experiment. Compared to the line segments that represent the noise level in the data set, the replicates fall within the range of the line segments.

In order to visualize the method of passive design variables, we added an extra plot at the bottom in Figure 2. As can be seen, in addition to the response loadings in the correlation loadings plot, the input variables also have loading values. Again A and C are seen as the most important, with A and C going in the same (positive) direction along component 1 and in different directions (positive for one, negative for the other) along component 2. This plot may be useful to look at, but it does not give explicit information about significance as does the ANOVA.

**4.3 Low and high salt level considered separately**

Since salt level is the main factor of importance here, we decided to do a PCA for low and high salt level separately as suggested in Section 2. The low salt results are presented in Figure 3a and as can be seen almost all variation is explained by one component only (89% explained). This one component is basically a contrast between low and high protein+carbohydrates level. As can also be seen, for low salt, firmness increases with increased protein+carbohydrate level and bitterness and “other salt” decrease. The main effect for A is about 4 times as large as for factor B. This corresponds well to the results above, except that the variable “other salt” now plays a stronger role as compared to the PCA based on all data. Note that in this case, significance testing is difficult since there will only be one degree of freedom (DF) for a model with main effects and two-factor interactions. The number of effects is also too low for obtaining a really useful Q-Q plot. The same will be the case for the other results based on splitting below. Interpretations must therefore be based on visual inspection, as is usually done in most applications of PCA. The CV clearly supports the results.

For high salt level (Figure 3b), two dimensions are needed (the first is clearly significant according to CV, the second one has only a low explained CV variance). The first component is again related to firmness, vs. bitterness and “other salt” plus now also the attribute liver paste taste shows up as being more important. The other variables lie on the same side of the 0-axis as firmness. Now the main effect for A is only 3 times as large as for B. Along the second axis, the main contrast is between samples 7 on one side vs. samples 5 and 13 on the other. The differences between these are the fat level. The fat (B) effect is along this axis more than 2 times as large as the effect of A. There also seems to be a tendency of interaction between fat and protein+carbohydrates along component 2 which can be more easily spotted when looking at the interaction plot (not shown). Also baking temperature has an effect here comparable to the effect of fat content. When only low salt was considered, this effect of fat content is not visible as strongly as it is here. In other words, splitting of the data set can highlight certain aspects and improve interpretation.

The replicates are again superimposed. For high salt content, the replicates are close to each other. The same is true for the first axis for low salt, but not for the second component. Since the second axis in this case has a small variation, and is thus very influenced by noise, this observation is as expected.

When high salt level observations are orthogonalized (see Section 2) with respect to the one-dimensional solution for low salt, the two first components explain 59% and 21% of the variation (Figure 4) respectively. Note that this plot gives explicit information about the additional correlation structure that comes in when salt is added, i.e. it highlights the effect of salt explicitly on the correlation structure. The main component now distinguishes between sample 7 and samples 5 and 13 indicating that it is a fat related component. This means that when adding salt, the main new variation added to the sensory properties is related to changes in fat content and not to protein+carbohydrate variation. Note that this does not mean that protein+carbohydrate does not have any effect, which is clearly proven above, it only tell us that the new and additional information obtained is more strongly related to fat variation. An interesting observation is that the firmness and bitterness are on the same side in the plot; when fat is at high level the bitterness and firmness levels are low. These findings were less obvious from the analyses above. Note that the first axis in Figure 4 has some similarities with the second axis in Figure 3b.

The same exercise based on splitting into two data sets was then done for protein+carbohydrate. The low protein+carbohydrate results were comparable to the overall view from the full PCA. The high protein+carbohydrate results are presented in Figure 5. In this case the firmness and bitterness are now closer together than above and most closely related to the salt content factor.

**4.4. Orthogonalisation with respect to the protein+carbohydrate factor**

The results from this method are very simple to interpret. The PCA gave one very dominating component (74%) which was for the most related to changes in salt content. For the loadings, bitterness was associated with low salt while all the other response variables with high salt content. There was also a small effect of fat along the same axis. For the second axis (11%) it was difficult to see any clear tendency. This analysis does not in this case seem to add much additional insight as compared to the results above.

**4.5. PLS regression**

In this case we tested both a model with main effects only and one with main effects and two-factor interactions incorporated as dummy variables. The results are present in Figure 6 for the main effects case. The results are more or less the same as for PC-ANOVA with the use of passive variables. The correlation loadings show the position of both the design variables and the response variables and the results correspond well to the interpretation of the PCA above. The most important design variables are protein+carbohydrate content and salt content with a small contribution for fat.

Also in this case, the replicates are projected onto the model and the results are as above quite good with respect to reproducibility of the experiment. The results are almost the same as for PCA. The two first components were clearly significant according to CV for explaining the variation in Y.

In this cases, testing the effect of the design variables on the PLS axes is not possible with the established tools. The jack-knife (see e.g. Martens and Martens (2001)) can, however, be used to test the significance of each of the design variables on each of the response variables (in a similar way as the ANOVA referred to in Table 2), but this aspect is not pursued here.

**4.6 Splitting up the PCA according to design factors (ASCA modifications)**

This method is useful since it considers the multivariate structure only for the design variables considered and not for all of them at the same time (see Section 2.3). In this sense it may possibly be more sensitive to details than a full PCA. In this case we decided to consider ASCA for three variables at a time, two quite important ones (fat and salt) and one less important (cooking temperature) determined by the ANOVA above. The main idea is to eliminate the effect of the very dominating protein+carbohydrate factor in order to possibly see some more details associated with the other three design variables. Before doing this analysis, we tested different combinations of input variables and it is clear (not shown here) that in this case the multivariate response structure is different for the different design variables.

The PCA solution obtained according to the simple averaging procedure described above is presented in Figure 7. The two components account for 87 and 7% of the variation respectively. The interpretation of the two axes is clear, the first axis corresponds to salt level and the second axis to fat level and to a certain extent also cooking temperature. There is a slight tendency of interaction between salt and fat as can be seen by the fact that the effect of salt is a bit larger for the low fat level. Again, eliminating dominating effects can be used for highlighting the minor ones. For the loadings, the salt axis is primarily related to bitterness vs. the rest. For the second axis one can see that firmness and sweetness are the most important. Although the multivariate correlation structure was somewhat different for the different design variables, it seems that the approach based on eliminating factor B in this case adds little to what is already obtained above.

**5. Discussion and conclusions**

This paper is primarily a comparison of alternative methods for analyzing designed experiments with multivariate output. Both one step procedures like the PLS which uses both the design and output variables in the same analysis as well as two step procedures such as PC-ANOVA and ASCA are tested. In fact a slight modification of ASCA for two-level experiments containing both main effects and interactions was proposed. A special emphasis is also given to splitting the data set in two and thus obtaining more detailed information. A new method based on splitting and orthogonalisation is also proposed and illustrated. Standard ANOVA of each variable separately was used as a benchmark and found to be useful for the overall interpretation.

**Comparison of methods**

The main information obtained by the plots of PLS and PC-ANOVA was very similar. Both the scores plots and the correlation loadings plots in two dimensions gave almost identical results and thus the same overall interpretation. The splitting of the data set before the use of PCA improved the interpretation of the data set. This procedure accompanied with orthogonalisation gave additional insight into the multivariate correlation changes when going from low to high salt level. Although PCA gave the most important information, the ANOVA used for the scores improved the interpretation and provided additional confidence in the results.

The results obtained by the modified ASCA are comparable to those obtained by the PC-ANOVA for those aspects that were compared. The ASCA showed that the multivariate correlation structure for the different design variables was different, but this did not have any strong effect on the overall interpretations. A drawback with the ASCA method is that it does not provide tests of significance within the plot. Residuals can be superimposed as was discussed above which can give some indication of variability, but this is not a real test of significance.

**Replicates**

The replicates were in this paper primarily used for checking the stability of the experimentation. As could be seen, both for the PLS and PCA the projections of the replicates were very close to the original measurements. Compared to the line segments in the PCA plot, used for indicating the error level, the replicates fall well within their range. It is also possible to use replicates to improve precision, but if the number of replicates is limited as it is here, this gives little additional precision and only complicates analysis because of imbalance. In such cases, using the projections as presented here is in most cases the best choice.

**Limitations**

The methods based on splitting in two data sets have the limitation that it cannot be used for all possible data sets. The design must be such that when split in two according to one of the variables, the experiments considered in each of the two blocks must be the same. For fractional data one will therefore have to eliminate one or some of the less important factors factors (as judged by significance tests) before further analysis. For full factorial designs it is always possible to use.

**Acknowledgments.** The work is financed by the project SALTO supported by Norwegian industry and the Research Council of Norway.

**References**

Baardseth, P., Næs, T., Mielnik, J. Skrede, G., Hølland, S. and Eide, O. (1992). Dairy ingredients effects on sausage sensory properties studied by principal component analysis. J. Food Science, 57, 4., 822-828.

Ellekjær, M.R., Ilseng, M.A. and Næs, T. (1996). A case study of the use of experimental design and multivariate analysis in product improvement. Food Quality and Preference, 7, 1, 29-36.

Endrizzi, E, Gasperi, F. Rødbotten, M. and Næs. T. (2013). Interpretation, validation and segmentation of preference mapping models. Submitted.

Harrington, P. de. B., Vieira, N.E., Espinoza, J., Nien, J.K., Romero, R., Yergey, A.L. (2005). Analysis of variance – principal component analysis: a soft tool for proteomic discovery. Analytica Chimica Acta, 544, 118-127.

Jansen, J., Hoefsloot, J. van der Greef, M., Timmerman, E., Westerhuis, J. and Smilde, A.K. (2005). “ASCA : analysis of multivariate data obtained from an experimemtal design.” J. Chemometrics. 19(9), 469-481.

Jansen, J., Bro, R., Huub C. Hoefsloot, J. van den Berg, F.W.J., Westerhuis, J and Smilde, A. K. (2008). “PARAFASCA: ASCA combined with PARAFAC for the analysis of metabolic fingerprinting data.” J. Chemometrics, vol. 22, 114-121.

Langsrud, Ø. (2002). 50-50 multivariate analysis of variance for collinear repsonses. J. Roy. Stat. Soc. Ser D. (The Statistician). 51, 305-317

Li, B., Martin, E. and Morris, J. (2008). “Process performance monitoring in the presence of confounding variation.” Chemometrics and intelligent laboratory systems. 94, 104-111.

Luciano, G. and Næs, T. (2009). Interpreting sensory data by combining principal component analysis and analysis of variance. Food Quality and Preference, 20, 3, 167-175.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Multivariate analysis. Academic Press, London, UK.

Martens, H. and Næs, T. (1989). Multivariate calibration. John. Wiley and sons, Chichester, UK.

Martens, H. and Martens, M. (2001). Multivariate analysis of quality. An introduction. John Wiley and Sons. Ltd. Chichester, UK.

Myers, R.H. and Montgomery, D.C. (1995). Response surface methodology. John Wily and sons, NY.

Næs, T., Brockhoff, P.B. and Tomic, O. (2010). Statistics for sensory and consumer science. John Wiley and Sons, Chichester, UK.

Zwanenburg, G., Hoefsloot, H.C.J., Westerhuis, J.A., Jansen, J.J. and Smilde, A.K. (2011).

ANOVA – principal components analysis and ANOVA - simultaneous component analysis: A comparison. J. Chemometrics. 25, 561-567.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  Sample |  A  Protein+carb | BFat | CSalt |   DTemp |
| 1 | -1 | -1 | -1 | -1 |
| 2 | 1 | -1 | -1 | -1 |
| 3 | -1 | 1 | -1 | -1 |
| 4 | 1 | 1 | -1 | -1 |
| 5 | -1 | -1 | 1 | -1 |
| 6 | 1 | -1 | 1 | -1 |
| 7 | -1 | 1 | 1 | -1 |
| 8 | 1 | 1 | 1 | -1 |
| 9 | -1 | -1 | -1 | 1 |
| 10 | 1 | -1 | -1 | 1 |
| 11 | -1 | 1 | -1 | 1 |
| 12 | 1 | 1 | -1 | 1 |
| 13 | -1 | -1 | 1 | 1 |
| 14 | 1 | -1 | 1 | 1 |
| 15 | -1 | 1 | 1 | 1 |
| 16 | 1 | 1 | 1 | 1 |

**Table 1.** Experimental design of the study. The symbol -1 corresponds to low level and the value 1 corresponds to high level of a factor. Samples no. 2, 8, 10 and 16 were repeated.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Firmness | Ordinary salt | Bitter taste | Other salt | Sweet taste | Spicy/onion | Liver paste taste |
| A-prot+carb | **0.00 (3.3)** | 0.39 | **0.00****(-0.5)** | **0.01****(-0.4)** | 0.32 | 0.23 | 0.11 |
| B-fat | **0.04****(0.6)** | 0.87 | 0.56 | 0.87 | 0.27 | 0.59 | 0.20 |
| C-salt | **0.05****(0.6)** | **0.00****(2.4)** | **0.00****(-0.4)** | 0.21 | 0.15 | **0.05****(0.9)** | **0.00** |
| D-Temp | 0.66 | 0.68 | 0.56 | 0.18 | 0.94 | 0.87 | 0.65 |
| AC | **0.06** |  | **0.01** |  |  |  |  |
| BC |  |  | **0.03** |  |  |  |  |
| BD |  |  | **0.03** |  |  |  |  |

**Table 2.** p-values for all main effects and for the significant interactions (p=0.05). The sensory reponse variables are presented along the horizontal axis. The actual effect sizes are given in parentheses. The significant factors (or close to, i.e 0.06) are highlighted with boldface.



 

 

**Figure 1.** Main effect plots (top and in the middle) and interactions between A and C (at the bottom) for firmness and bitterness. Firmness on the top, bitterness in the middle. For the interactions plots, firmness is to the left and bitterness to the right. The two levels -1 and 1 along the horizontal axes for the interaction plots correspond to low and high level of salt. The inner panel of the two interactions plots indicate that the solid lines correspond to low level og A and the dotted line to high level of A.







**Figure 2.** Regular PCA plots for all 16 samples with projected replicates (joined by lines). Correlation loadings, scores with replicates projected (joined by lines) and correlation loadings with passive design variables (light grey).

|  |  |  |
| --- | --- | --- |
| Factor | PC1 | PC2 |
| A | **0.00****(3.2)** | **0.01****(-1.2)** |
| B | **0.02****(0.67)** | 0.58 |
| C  | **0.00****(1.78)** | **0.00****(2.3)** |
| D | 0.90 | 0.44 |
| A\*B | 0.23 | 0.19 |
| A\*C | 0.09 | 0.20 |
| A\*D | 0.92 | 0.60 |
| B\*C | 0.43 | 0.22 |
| B\*D | 0.77 | 0.70 |
| C\*D | 0.44 | 0.18 |

**Table 3.** P-values for the input factors and their interactions for principal component 1 (PC1) and principal component 2 (PC2). The effect sizes are given in parenthesis.





**Figure 3.** PCA scores and correlation loadings for low and high salt content seperately. The two plot at the top (a) represent low salt, while the two at the bottom (b) represent high salt level. Replicates are joined by lines.





**Figure 4.** PCA scores and loadings for high salt samples when orthognalised with respect to first component from low salt.



**Figure 5.** PCA scores and correlation loadings plots for high protein+carbohydrate samples only.





**Figure 6**. Correlation loadings and scores plots of PLS. The replicates are obtained by projections as describd above (joined by lines). Design variables in grey in the loadings plot. The design variable C is partly overlapped with attribute salt taste.



**Figure 7**. Projections of averages over the levels of factor A. The design factors involved are B, C and D. The symbols -1 and 1 have the same meaning as above (low level and high level).