Consumer-Oriented New-Product Development in Fruit Flavour Breeding-A Bayesian approach

Lebeyesus M. Tesfaye[1], Ivo A. van der Lans[2], Marco C.A.M. Bink[3], Bart G.J. Gremmen[1], Hans C.M. van Trijp[2]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 40

[1] Wageningen University, Methodical Ethics and Technology Assessment, Wageningen, The Netherlands
Lebeyesus.Tesfaye@wur.nl
Bart.Gremmen@wur.nl
[2] Wageningen University, Marketing and Consumer Behaviour Group, Wageningen, The Netherlands
Ivo.vanderLans@wur.nl
Hans.vanTrijp@wur.nl
[3] Wageningen University and Research Centre, Biometris, Wageningen, The Netherlands
Marco.Bink@wur.nl

Abstract

Taking consumer quality perceptions into account is very important for new-fruit product development in todays competitive food market. To this end, consumer-oriented quality improvement models like the Quality Guidance Model (QGM) have been proposed. Implementing such models in the agro industry is challenging. We propose the use of Bayesian Structure Equation Modelling (SEM) for parameterizing the Quality Guidance Model, allowing for the integration of elicited expert knowledge. Such casual modelling would furnish important insights for determining the optimal fruit product in terms of consumer flavour-quality perceptions. In the context of tomato breeding, where we have data about metabolites, sensory-panel judgments, and consumer flavour-quality perceptions, we estimated a benchmark Bayesian SEM using non-informative priors, starting from an initial causal model derived from the data with a score-based Bayesian Network (BN) learning algorithm. The results so far have given some indication of the importance of accounting for consumer heterogeneity in the modeling process.

Keywords: Bayesian Structural Equation Modeling; Quality Guidance Model; Elicitation; Fruit Flavour

1 Introduction

Improving flavour quality traits in fruit breeding, calls for innovative consumer-oriented product development models. However, the wide gap in the agro sector between consumer or marketing data on the one-hand side, and metabolite and genomics data on the other-hand side poses a great challenge to implement such modelling. In our research, we aim to link consumer flavour-quality perceptions, trained sensory-panel judgments and various flavour-affecting metabolites in the context of tomato breeding. To address the challenge we have proposed[3] to use food quality improvement models like the Quality Guidance model (QGM) of Steenkamp and van Trijp[6] and to parameterize it using Bayesian SEM[5][7]. This would enable us to integrate elicited expert knowledge on the degree of causal associations of the metabolite and different flavour-quality perceptions in the model to obtain more valid and robust estimates of the strength of the different causal relations in the model. This could help flavour researchers to pin down the optimum concentration of flavour-affecting metabolites, which further can be used as phenotypes for marker-association studies. The QGM adapted for tomato flavour quality is shown in Figure 1. The quality cues and attributes as well as the quality expectation and quality experience constitute the consumer data, while in the left we have the metabolite and the trained sensory panel data.

2 Material and Methods

So far we have conducted a benchmark Bayesian SEM analysis with non-informative priors. We would conduct an elicitation from experts in the next stage. In our analysis, we included non-averaged consumer ratings of 54 round tomato cultivars on the 1-7 Likert scale. The ratings were on various flavor quality cues and attributes consisting of color, aroma, taste as well as quality expectation and quality experience indicators. For the trained sensory panel traits that have been measured on a 1-100 line scale, we used averaged ratings on various sensory characteristics on the same set of round tomato cultivars consisting of scent, taste and aftertaste. From the literature and an expert feedback, we got a shortlist of 31 metabolites (consisting of acids, sugars, volatiles and some carotenes) that are suspected to affect tomato flavour quality, and we selected these metabolites from a metabolite profile database on the same set of tomatoes. For a more complete description of the nature of the data and how it was generated, see Van Den Heuvel et al.[8]. We standardized the scores for all variables. Before conducting the Bayesian SEM analysis we needed to have an initial, estimable causal model. As the available findings in the literature were not sufficient for constructing such an initial model, we resorted to the use of score-based Bayesian Networks algorithm (Hill Climbing with AIC network score, implemented in the R package bnlearn [4]) to derive it from our data. For the BN learning we only used the cues and attributes from the consumer data and later on we included the quality expectation and quality experience latent constructs in the SEM model according to the QGM. During the BN structure learning, we allowed only causal relations in the directions of the smaller arrows shown in Figure 1. Section 3 gives some details of the results and procedures for the subsequently estimated Bayesian SEM.

3 Results

After learning a BN structure from the data using a score-based algorithm, we needed a way to still further reduce the relations to obtain an initial, estimable Bayesian SEM. Hence we first fitted the learned network with MLE estimation provided in the package and deleted some relations on the basis of those results. An upper (0.5) and a lower (0.25) threshold value were used to select the to-be-included relations on the basis of the estimated regression weights. Both threshold values yielded a rather large SEM models having up to 400 path coefficients for the most extensive model (larger model). WinBugs/R2WinBugs were used in the analysis and the models were run up to 20K draws with adequate burn-in for two MCMC chains. Non-informative priors were used using the recommendation in Lee[7]. Convergence was checked for representative parameters (out of thousands of SEM model parameters including path coefficients, error terms, latent score estimates etc.) that represent the different parts of the model by both observing the trace plots as well as bgr diagnostics. Deviance Information Criterion (DIC)[2] values were used for model comparison and the more comprehensive model (DIC= -50607) was selected over the more restricted model (DIC=-50580). Model adequacy for the selected model was also checked using posterior predictive check and most of the data falls within the 95 percent credible interval of the posterior predictive distribution.

The results show many significant estimates of higher magnitude from metabolites to sensory-panel data, but also within the consumer data (i.e.from cues/attributes to quality expectation and quality experience). We also have significant estimates of higher magnitude from the sensory panel data towards the consumer data. However, we observe non-significant and very small estimates of the paths towards the consumer quality cue and quality attribute perceptions (See Table 1). As presenting all the output is not practical, Table 1 shows only a small portion of the WinBugs output taking sweet taste as representative i.e. sweet taste from consumer data (sweet-tasteC) as well as sweet taste from sensory-panel data (taste-sweet). The table also shows all the estimates of the path coefficients from consumer cues and attributes towards the quality expectation (QexpctC) and quality experience (QexprnC). Besides the mean and median, we have also included the standrad deviation(s.d.), Monte carlo error(MC Err.)and the 95 percent credible interval if there is interest to see more details of the inference and the sampling. In the table, to distinguish the consumer data from the sensory panel data, the variables from the consumer data end with the letter 'C'. The color and smell features associated with 'cut' refers to evaluations made after cutting the fruit. The metabolites are given in their full names. The underlined values are non-significant judged by looking whether the credible interval includes zero[1]. Furthermore, adjusted R^2s (not shown in the table) are higher for both the sensory-panel variables and the consumer quality expectation and quality experience, while for the consumer cues and attributes they are very small. We postulated that heterogeneity among consumers was a major cause for the non-significant path coefficients towards the middle consumer cue and attributes of the QGM. This was supported by an additional Bayesian SEM analysis using an averaged consumer data that showed an increase in the values of these path coefficients.

4 Concluding remarks

Based on the results, in a future research we aim to account for consumer heterogeneity using a finite mixture Bayesian SEM[7]. Once this yields a suitable benchmark model, we will start to specify informative priors on the basis of elicited expert knowledge.

References

[1] B. Muthen,T. Asparouhov. **Bayesian Structural Equation Modeling:A more flexible representation of substantive theory**. *Psychological Methods*; 2012; 17(3); pp. 313-335.

[2] D.J.Spiegelhalter, N.G.Best, B.P.Carlin, A. van der Linde. **Bayesian measures of model complexity and fit**. *Journal of Royal Statistical Society*; 2002; 64(4); pp. 583-639.

[3] L.M. Tesfaye, M.C.A.M. Bink, I.A. van der Lans, B. Gremmen, H.C.M. van Trijp. **Bringing the voice of consumers into plant breeding with Bayesian Modeling**. *Euphytica*; 2013; 189(3); pp. 365-378.

[4] M. Scutari **Learning Bayesian Networks with the bnlearn R package**. *Journal of Statistical Software*; 2010; 35(3)

[5] J. Palomo, D.B. Dubson, K. Bollen. **Bayesian Structural Equation Modeling**. *Handbook of Computing and Statistics with Applications; North-Holland, The Netherlands*; 2007; 1 ; pp. 163-188.

[6] J-B.E.M. Steenkamp; H.C.M. van Trijp. **Quality Guidance: A consumerbased approach to food quality improvement using PLS**. *European Review of Agricultural Economics*; 1996; 23; pp. 195-215.

[7] S.Y. Lee. **Structural equation modeling: A Bayesian approach**. *Wiley, Chichester*; 2007

[8] T. Van Den Heuvel , H.C.M.van Trijp, C. van Woerkum, R. Jan Renes, B. Gremmen. **Linking product offering to consumer needs; inclusion of credence attributes and the influences of product features**. *Food Quality and Preference*; 2007; 18(2); pp. 296-304

Table 1: WinBugs output showing the path coefficient estimates for consumer and sensory panel sweet taste feature and the path coefficient estimates within the consumer data

From	To	Mean	s.d.	MC Err.	2.5 p	Median	97.5 p
scent-sweet	taste-sweet	0.664	0.043	0.001	0.559	0.645	0.727
scent-smoky	taste-sweet	-0.471	0.081	0.003	-0.628	-0.473	-0.309
2-methylbutanal	taste-sweet	0.053	0.087	0.003	-0.121	0.055	0.222
1-penten-3-one	taste-sweet	0.827	0.131	0.006	0.558	0.828	1.074
3-methylbutanol	taste-sweet	-0.318	0.104	0.004	-0.524	-0.316	-0.117
2-methylbutanol	taste-sweet	-0.387	0.102	0.004	0.189	0.386	0.592
cis-3-hexenal	taste-sweet	-0.808	0.089	0.004	-0.984	-0.808	-0.634
hexanal	taste-sweet	0.479	0.069	0.002	0.339	0.479	0.614
trans-2-heptenal	taste-sweet	-0.242	0.091	0.003	-0.420	-0.243	-0.059
methylsalicylate	taste-sweet	0.619	0.082	0.003	0.455	0.623	0.775
1-penten-3-ol	taste-sweet	-0.267	0.081	0.003	-0.424	-0.269	-0.107
beta-ionone	taste-sweet	-0.115	0.059	0.001	-0.227	-0.115	0.001
Hexanol	taste-sweet	0.142	0.055	0.001	0.035	0.142	0.254
scent-smoky	sweet-tasteC	0.010	0.151	0.006	-0.298	0.012	0.297
taste-sweet	sweet-tasteC	-0.166	0.081	0.002	-0.328	-0.165	-0.010
aftertaste-salt	sweet-tasteC	0.091	0.096	0.003	-0.099	0.090	0.277
taste-tomato	sweet-tasteC	0.171	0.103	0.004	-0.031	0.171	0.374
aftertaste-chemical	sweet-tasteC	0.108	0.079	0.001	-0.047	0.107	0.266
pleasant-smell(cut)	sweet-tasteC	0.144	0.056	0.000	0.033	0.144	0.257
2-methylbutanal	sweet-tasteC	0.206	0.129	0.005	-0.051	0.206	0.451
1-penten-3-one	sweet-tasteC	-0.002	0.161	0.007	-0.323	-0.016	0.310
cis-3-hexenol	sweet-tasteC	-0.009	0.148	0.006	-0.289	-0.011	0.281
2-izobutylthiazol	sweet-tasteC	-0.123	0.095	0.003	-0.308	-0.124	0.064
phenylethanol	sweet-tasteC	0.160	0.095	0.004	-0.024	0.160	0.352
methylsalicylate	sweet-tasteC	0.015	0.133	0.005	-0.238	0.012	0.288
beta-damascenone	sweet-tasteC	-0.057	0.117	0.005	-0.285	-0.060	0.178
3-methylbutanal	sweet-tasteC	-0.144	0.146	0.007	-0.434	-0.144	0.140
1-penten-3-ol	sweet-tasteC	0.011	0.131	0.005	-0.241	0.011	0.268
hexanol	sweet-tasteC	0.030	0.133	0.005	-0.237	0.030	0.290
aspartic-acid	sweet-tasteC	0.227	0.141	0.006	-0.055	0.227	0.500
glutamate	sweet-tasteC	-0.274	0.161	0.008	-0.582	-0.275	0.051
glucose1	sweet-tasteC	-0.078	0.104	0.004	-0.283	-0.078	0.128
citric-acid	sweet-tasteC	-0.079	0.092	0.003	-0.258	-0.079	0.102
myo-Inositol	sweet-tasteC	0.204	0.095	0.002	0.021	0.204	0.392
sucrose	sweet-tasteC	-0.043	0.084	0.002	-0.205	-0.043	0.121
pleasant Smell	QexpctC	0.809	0.077	0.001	0.661	0.809	0.964
colorC	QexpctC	0.283	0.043	0.000	0.198	0.284	0.367
QexpctC	QexprncC	0.369	0.050	0.000	0.269	0.37	0.469
pleasant-smelC(cut)	Qexprnc	0.057	0.045	0.000	-0.029	0.056	0.146
bitter-tasteC	QexprncC	-0.011	0.046	0.000	-0.099	-0.011	0.081
sour-tasteC	QexprncC	-0.039	0.047	0.000	-0.130	-0.039	0.052
watery-tasteC	QexprncC	-0.032	0.041	0.000	-0.113	-0.032	0.047
fresh-tasteC	QexprncC	0.309	0.044	0.000	0.220	0.309	0.395
sweet-tasteC	QexprncC	0.201	0.040	0.000	0.123	0.200	0.282
colorC(cut)	QexprncC	0.174	0.045	0.000	0.085	0.175	0.263

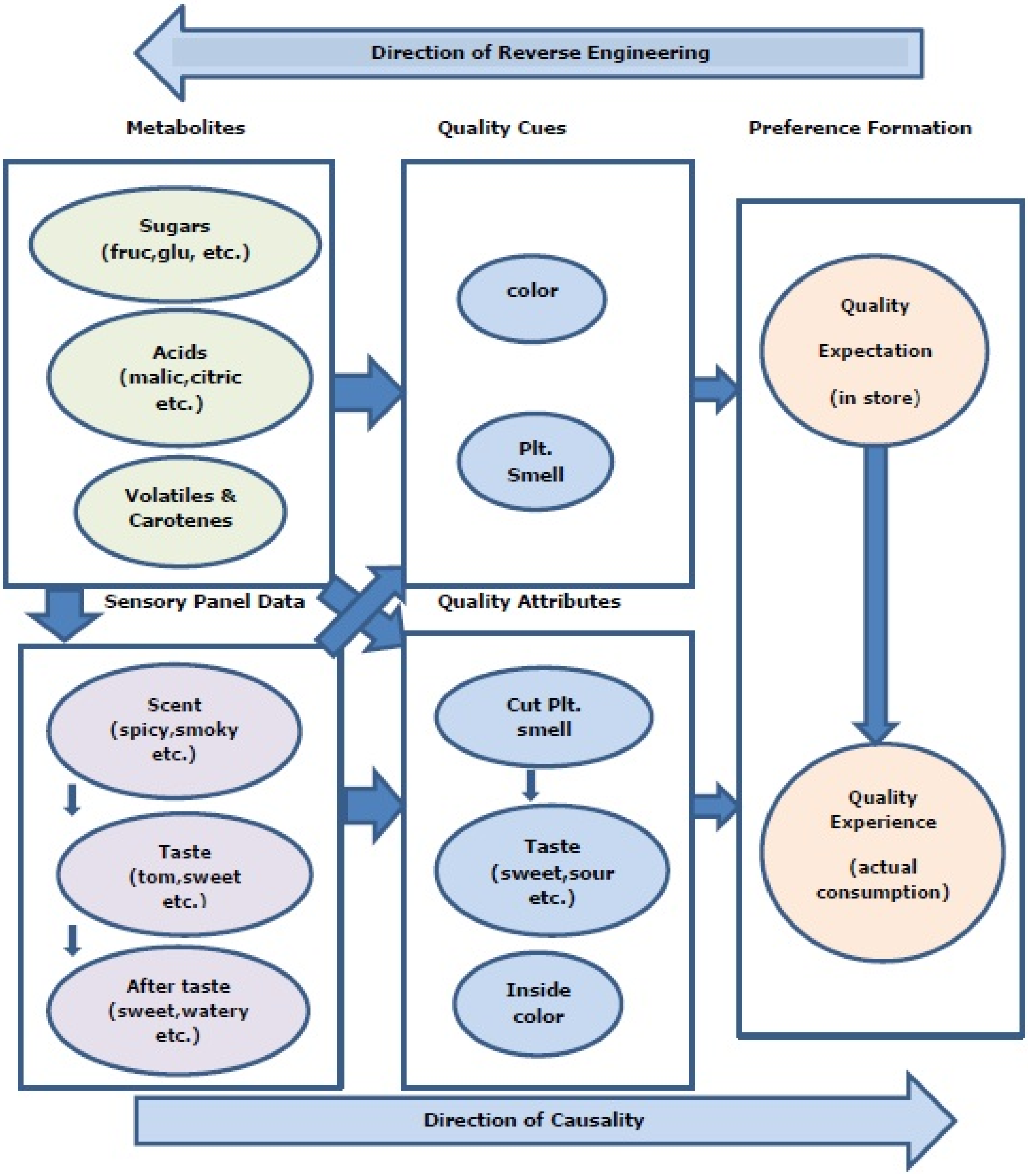

Figure 1: The QGM for tomato-flavour improvement