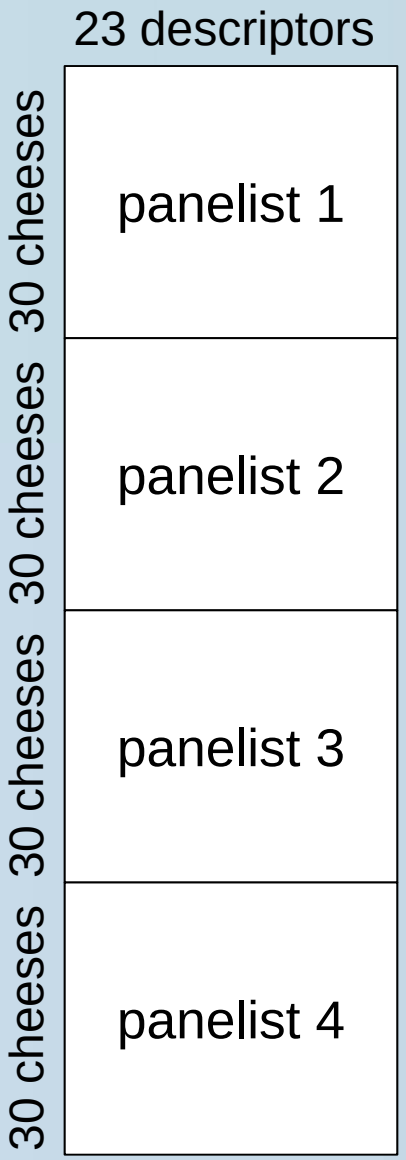
KATHOLIEKE UNIVERSITEIT
LEUVEN

The CHull procedure for selecting among multilevel component solutions

Eva Ceulemans, K.U.Leuven
Marieke E. Timmerman, R.U.Groningen
Henk A.L. Kiers, R.U.Groningen

1. Class of multilevel component models

- two-level multivariate data
  - example: sensory profiling study
    - 8 panelists were asked to rate samples of 30 cream cheeses on 23 descriptors



1. Class of multilevel component models

- similar to ANOVA, data are split up in two parts:

DATA (**X**) = BETWEEN PART ($\mathbf{X}^b$) + WITHIN PART ($\mathbf{X}^w$)

mean values of each panelist

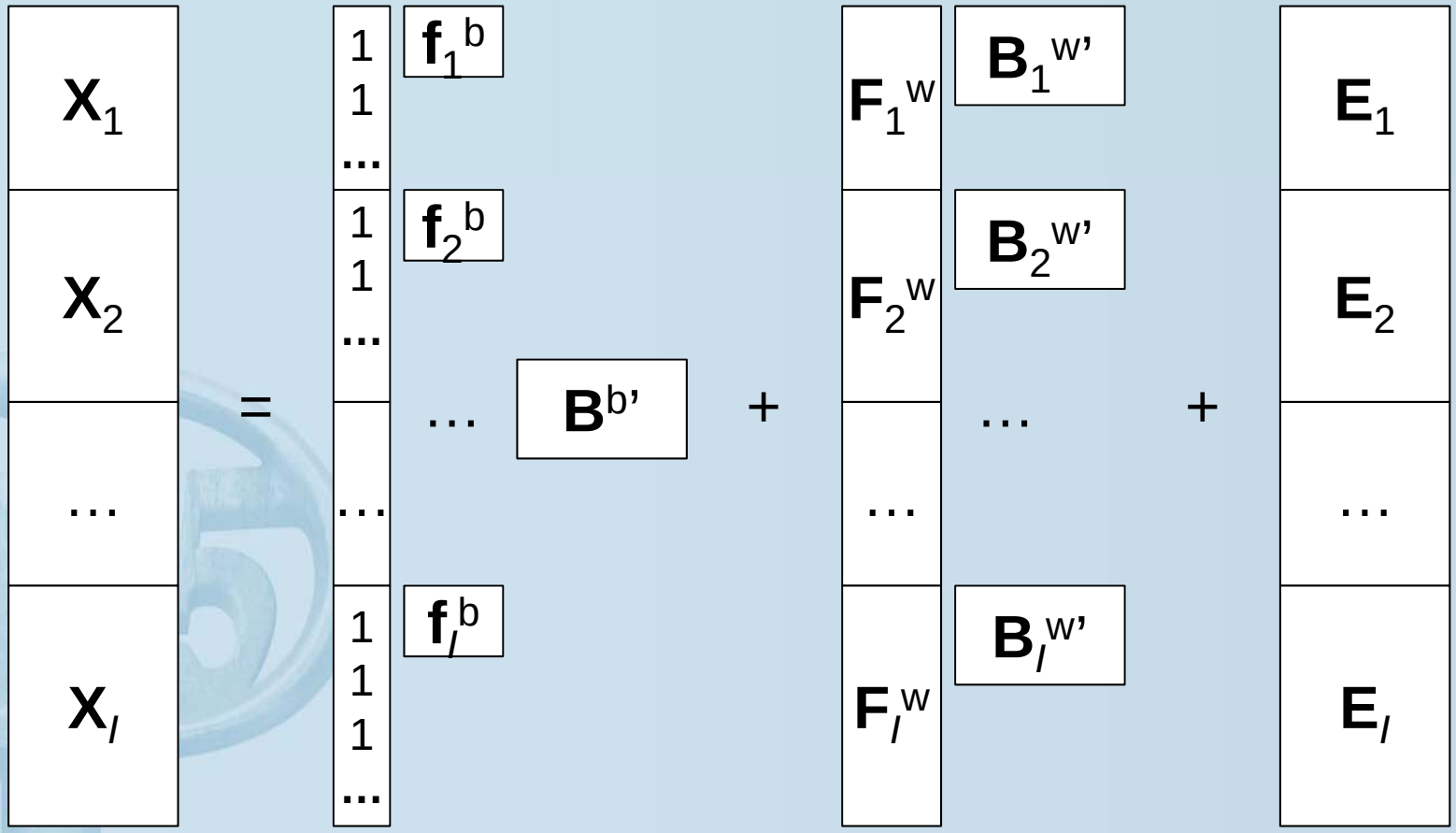
differences between-panelists

deviations from mean values per panelist

differences within-panelists

1. Class of multilevel component models

$$\mathbf{X}_i = \mathbf{X}_i^b + \mathbf{X}_i^w$$
$$= 1_{K_i}\mathbf{f}_i^b\mathbf{B}^{b\prime} + \mathbf{F}_i^w\mathbf{B}_i^{w\prime} + \mathbf{E}_i$$

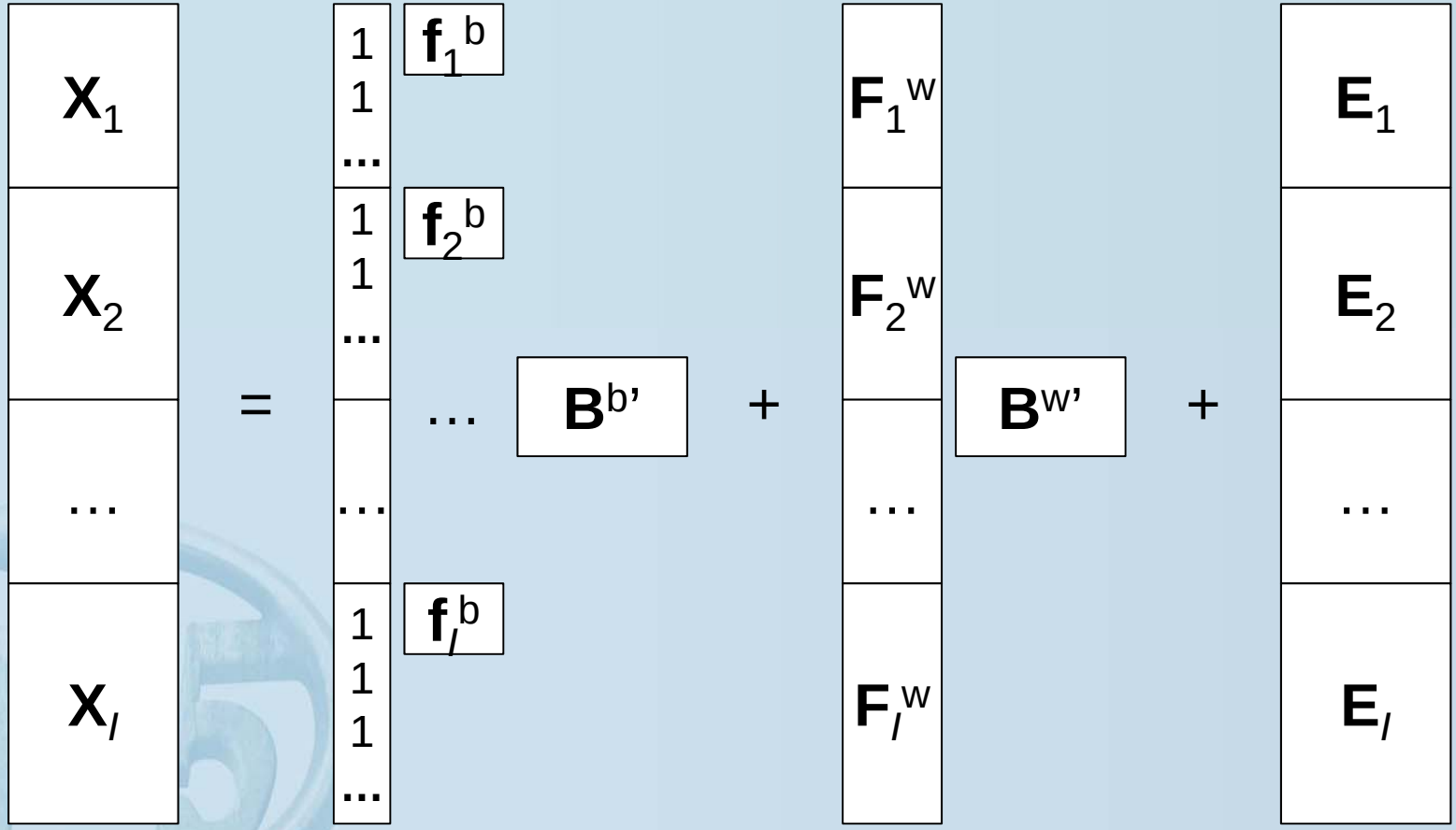


1. Class of multilevel component models

$$\mathbf{X}_i = 1_{K_i}\mathbf{f}_i^b\mathbf{B}^{b\prime} + \mathbf{F}_i^w\mathbf{B}_i^{w\prime} + \mathbf{E}_i$$

variant	Within-Loadings $\mathbf{B}_i^w$	Correlations $\mathbf{F}_i^w$	Variances $\mathbf{F}_i^w$
MLCA	Free	-	-
MLSCA-P	Equal for all *i*	Free	Free
MLSCA-PF2	Equal for all *i*	Equal for all *i*	Free
MLSCA-IND	Equal for all *i*	Equal to 0	Free
MLSCA-ECP	Equal for all *i*	Equal for all *i*	Equal for all *i*

1. Class of multilevel component models



2. CHull heuristic

- between-model selection problem
  - number of between-components?
- within-model selection problem
  - variant? number of within-components?
- formal rule which assesses complexity of different solutions by considering number of free parameters (Ceulemans & Kiers, 2006)

2. CHull heuristic: within-part

$$\mathbf{X}_i^w \approx \mathbf{F}_i^w \mathbf{B}_i^{w\prime}$$

component scores + # loadings – Q_w^2 - Q_w

transformation freedom

mean within-component score of each panelist = 0

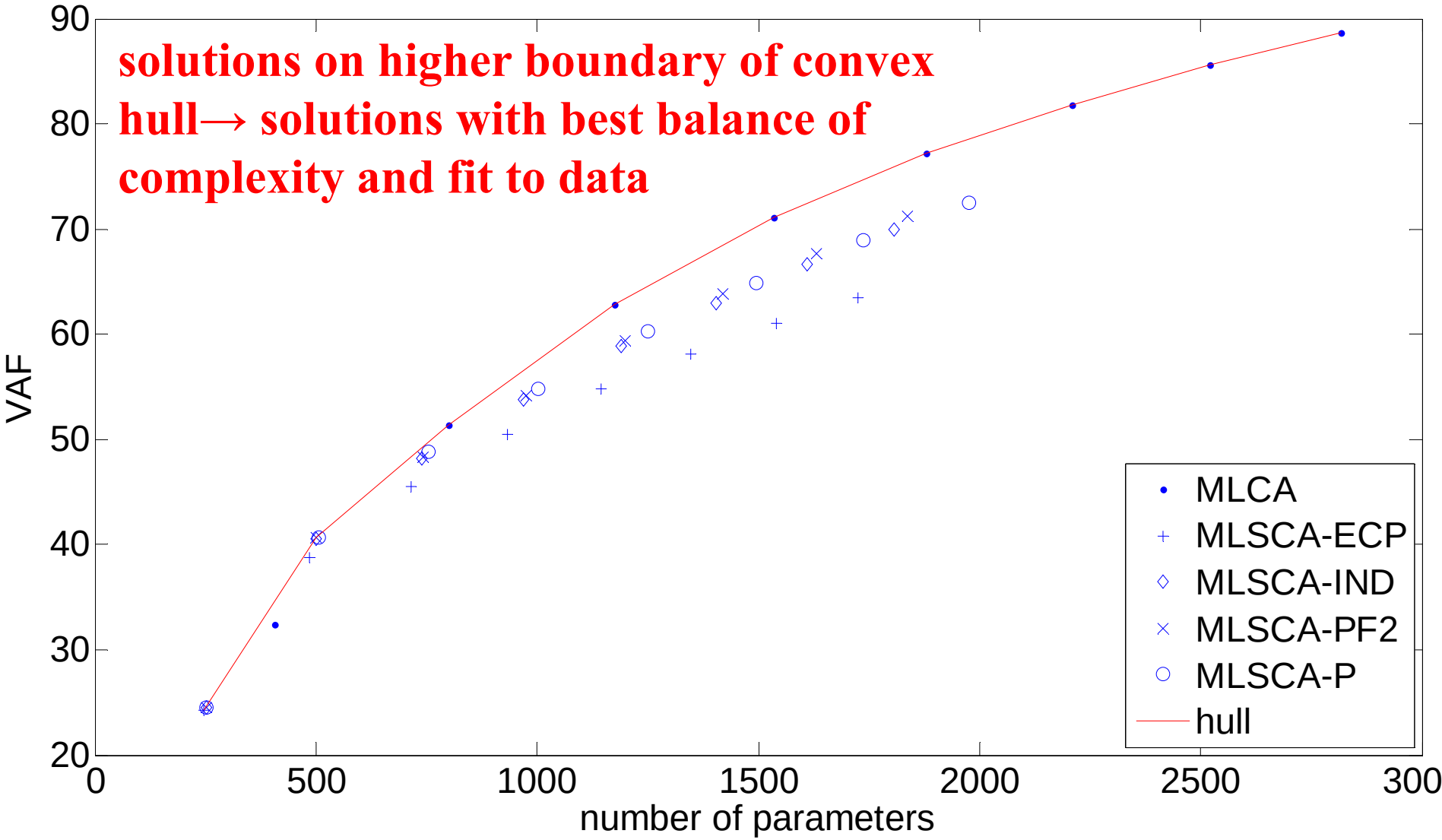
2. CHull heuristic: within-part

$$\mathbf{X}_i^w \approx \mathbf{F}_i^w \mathbf{B}_i^{w\prime}$$

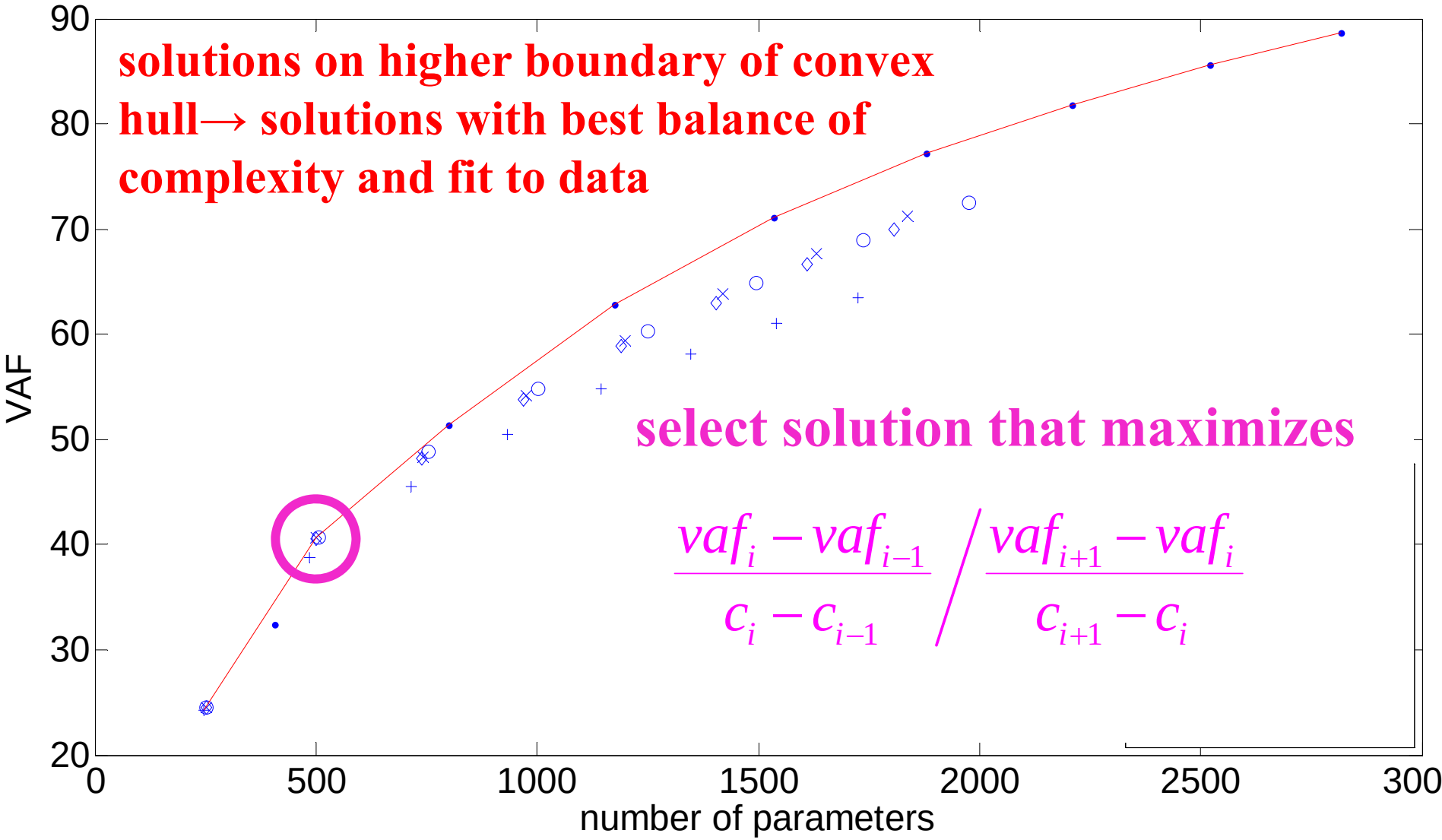
component scores + # loadings – Q_w^2 - Q_w

- #cheeses*Q_w : if #cheeses increases, term becomes too large
- min(#cheeses,ln(#cheeses)*#variables)*Q_w: mitigates influence of additional cheeses -> works well in simulation study!

2. CHull heuristic: within-part



2. CHull heuristic: within-part



3. Simulation study: 84240 data sets

- assessing the number of between-components: easy (98.8%)
- determining the number of within-components: easy (91.4%)
- tracing the underlying within-model variant (60.71%):
  - differences in within-loadings: easy
  - differences in variances of within-components: easy
  - differences in correlational structure of within-components: difficult (procedure often indicates that correlations differ, whereas they do not)

4. Discussion

- CHull heuristic is a useful tool
- more fundamental problem remains: how to determine number of free parameters in component analysis?

References

- Ceulemans, E., & Kiers, H.A.L. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology*, *59*, 133-150.
- Ceulemans, E., Timmerman, M.E., & Kiers, H.A.L. (in press). The CHULL procedure for selecting among multilevel component solutions. *Chemometrics and Intelligent Laboratory Systems.*
- Timmerman, M.E. (2006). Multilevel component analysis. *British Journal of Mathematical and Statistical Psychology, 59,* 301–320.