

Web-based Supplementary Materials for Extraction
of food consumption systems by non-negative
matrix factorization (NMF) for the assessment of
food choices by Mélanie Zetlaoui, Max Feinberg,
Philippe Verger, Stephan Clémentçon

Web Appendix A

NMF gradient descent algorithm : technical details

Let consider the following constrained optimization problem:

$$\begin{aligned} D_K(V, (W, H)) &= \|V - WH\|^2 \\ &= \sum_{n=1}^N \sum_{f=1}^F \left(v_{fn} - \sum_{k=1}^K W_{fk} H_{kn} \right)^2 \end{aligned}$$

subject to the constraints that $W_{fk} \geq 0$, $H_{kn} \geq 0$ and $\sum_{f=1}^F W_{fk} = 1$.

The standard gradient descent approach consists at each step in assigning to each component W_{fk} of W the value

$$W_{fk} \leftarrow W_{fk} - \eta_{fk} \frac{\partial D_K(V, (W, H))}{\partial W_{fk}}$$

that is

$$W_{fk} \leftarrow W_{fk} + 2\eta_{fk} \left[\sum_{n=1}^N V_{fn} H_{kn} - \sum_{n=1}^N \sum_{k'=1}^K W_{fk'} H_{k'n} H_{kn} \right]$$

with η_{fk} a small positive real number. To assure the positivity of W_{fk} let put

$$\eta_{fk} = W_{fk} [VH^t]_{fk}$$

that implies the multiplicative update rule for W_{fk}

$$W_{fk} \leftarrow W_{fk} \frac{[VH^t]_{fk}}{[WHH^t]_{fk}}$$

Notice that the choice of η_{fk} is not necessarily small that not assures directly the algorithm convergence.

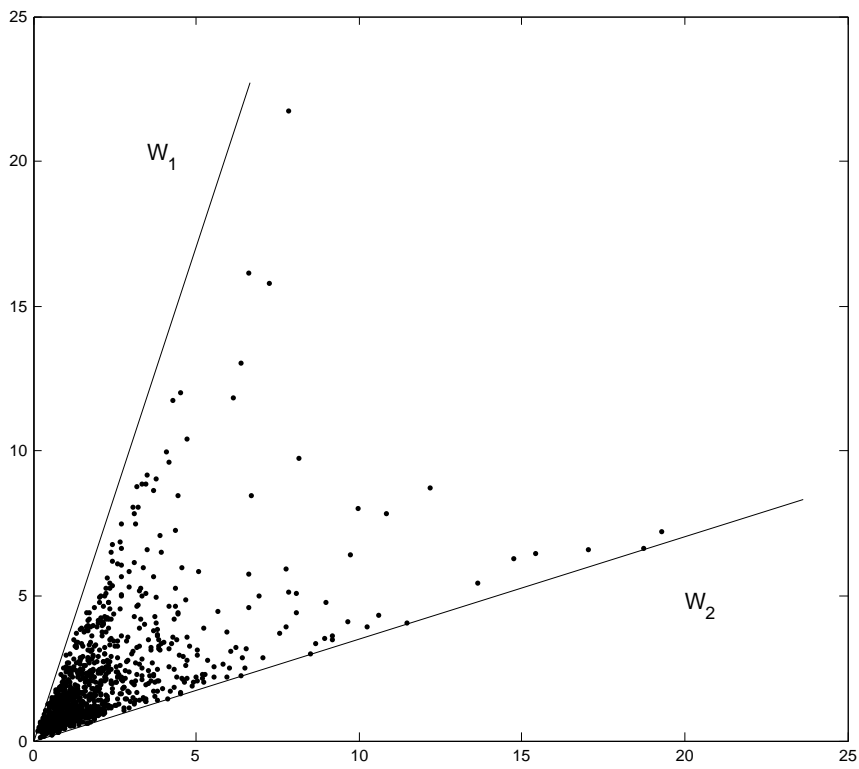


Figure 1: Representation of the predicted data in the map containing \widehat{W}_1 and \widehat{W}_2 .

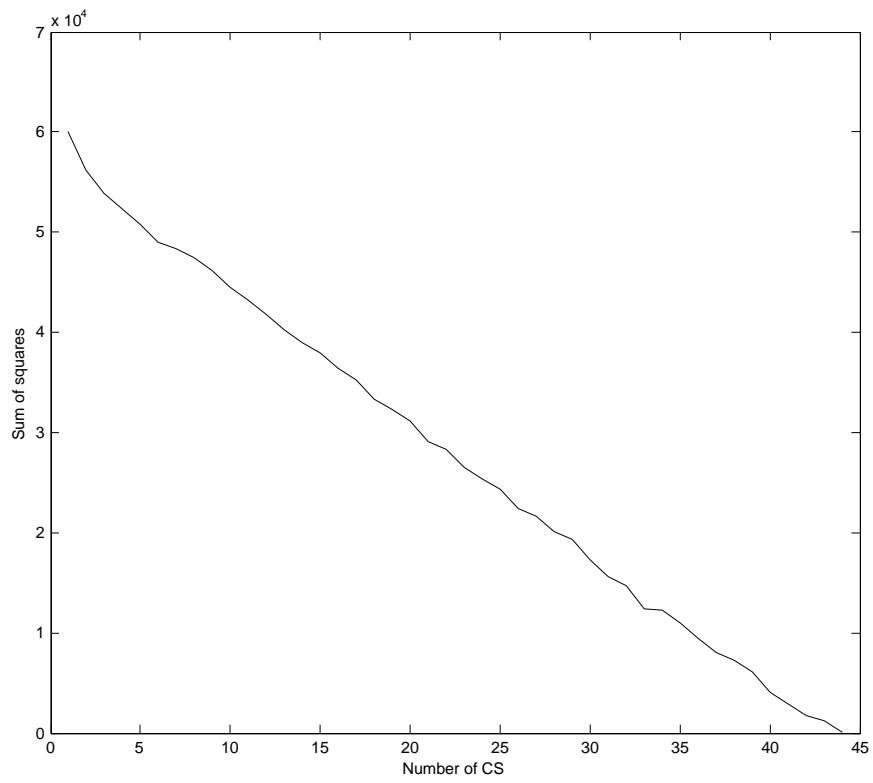


Figure 2: Residual sum of squares for K running from 1 to 44.

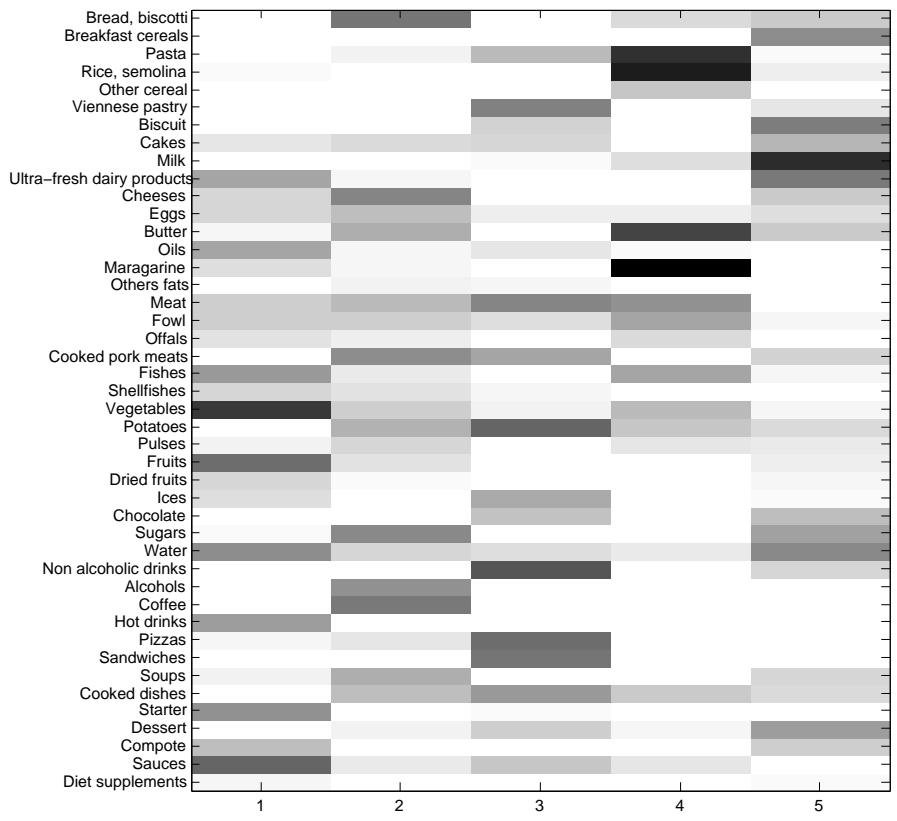


Figure 3: Graphical representation of the W matrix which describes the loadings of each of the 44 food groups to each of the K consumption systems for $K = 5$.

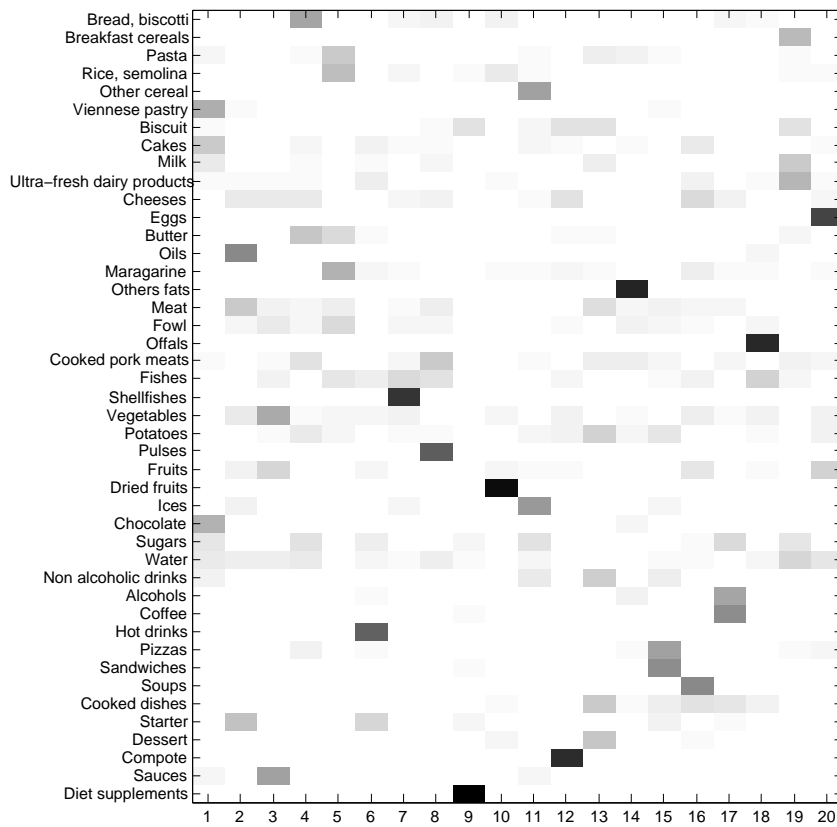


Figure 4: Graphical representation of the W matrix which describes the loadings of each of the 44 food groups to each of the K consumption systems for $K = 20$.

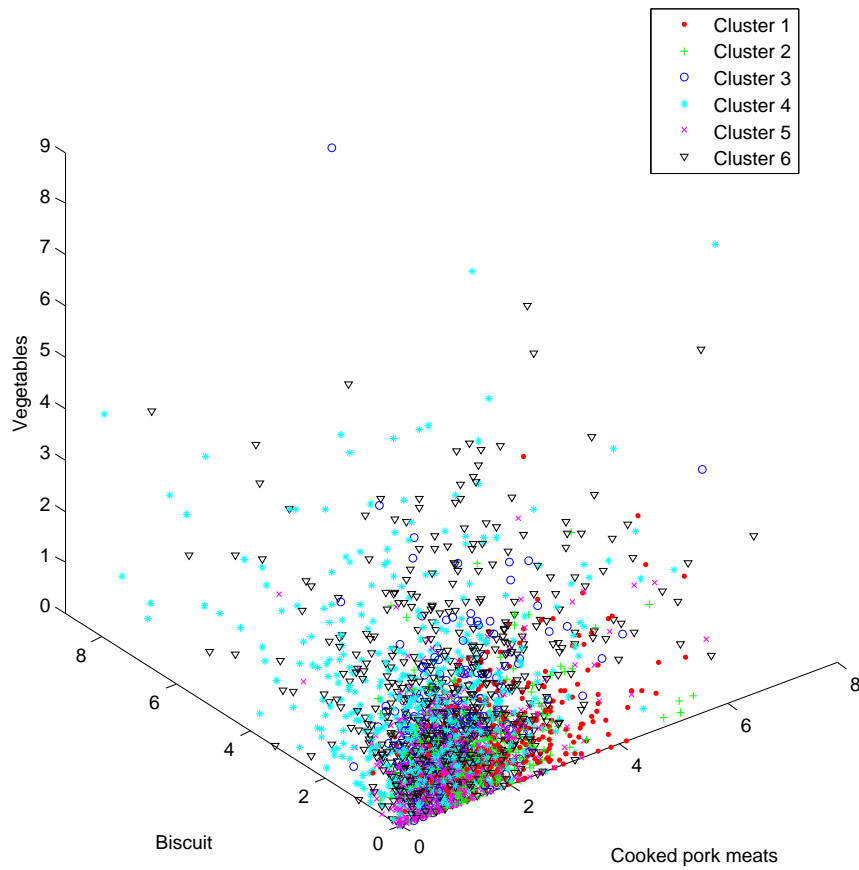


Figure 5: Representation of the six clusters obtained by k -means clustering in the base of cooked pork meats and vegetables.

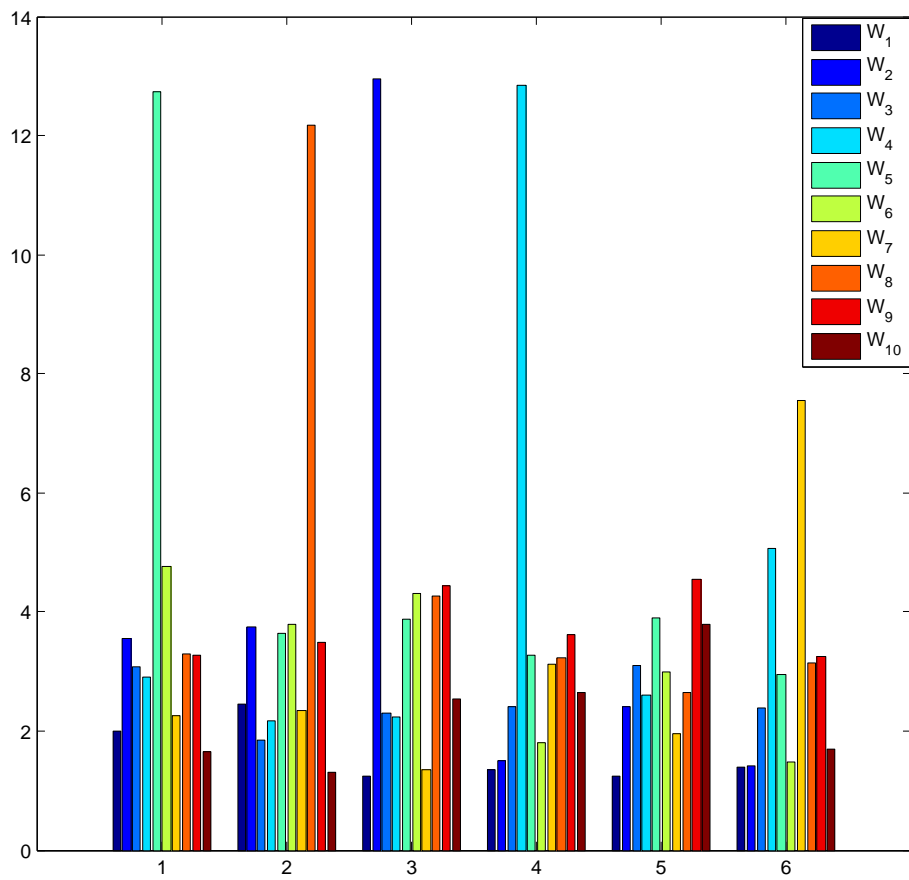


Figure 6: Representation of the coordinates in consumption system base of the 6 cluster centroids obtained by k -means clustering.