APPENDIX 1: EXPLANATIONS ON THE STATISTICAL TREATMENTS

The decomposition of the signal is done through the following steps.

First, we use a *PieceWiseRegressor* from Python library *mlinsight* to simultaneously split the time axis into several periods and perform a linear regression on each part. The decision tree is parameterized with at most 4 leaves (which means it allows a maximum of 4 periods) and at least 36 items per leaf (which means a given period contains at least 36 months in a row). In the end, the resulting signal (the *trend*) is the piecewise linear model that minimizes the least square distance to the data.

$$\min_{a,b,T}\left(\sum_{k=1}^{4}\sum_{i=T_k-1}^{T_k}(q_i - a_k t_i - b_k)^2\right)$$

Second, for each month we compute the average difference (over the full dataset) between the actual signal and the trend, that is the *seasonality*:

$$\sigma_m = \sum_{k=1}^{4}\sum_{i\in[T_{k-1},T_k],i\%12=m}(q_i - a_k t_i - b_k)$$

Finally, we compute the *residuals*, which are the pointwise difference between the signal, the trend and the seasonality:

$$r_i = q_i - a_{k(i)}t_i - b_{k(i)} - \sigma_{i\%12}$$

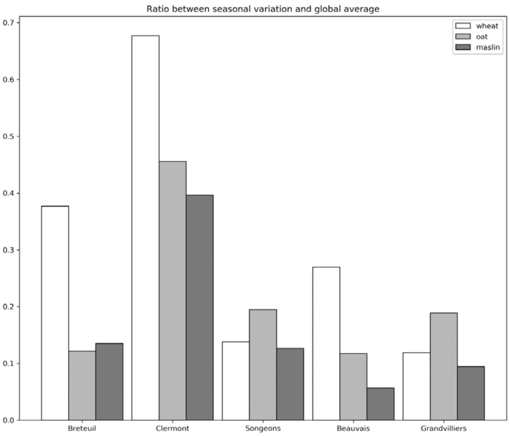
These residuals are studied through a variety of indicators, in order to decide if they can be reasonably considered as a Gaussian noise, namely: autocorrelation, histogram, n vs n+1 plot and QQplot[29].

28. See, for instance, Wetherill (1981).

APPENDIX 2: RELATIVE IMPORTANCE OF SEASONAL VARIATION

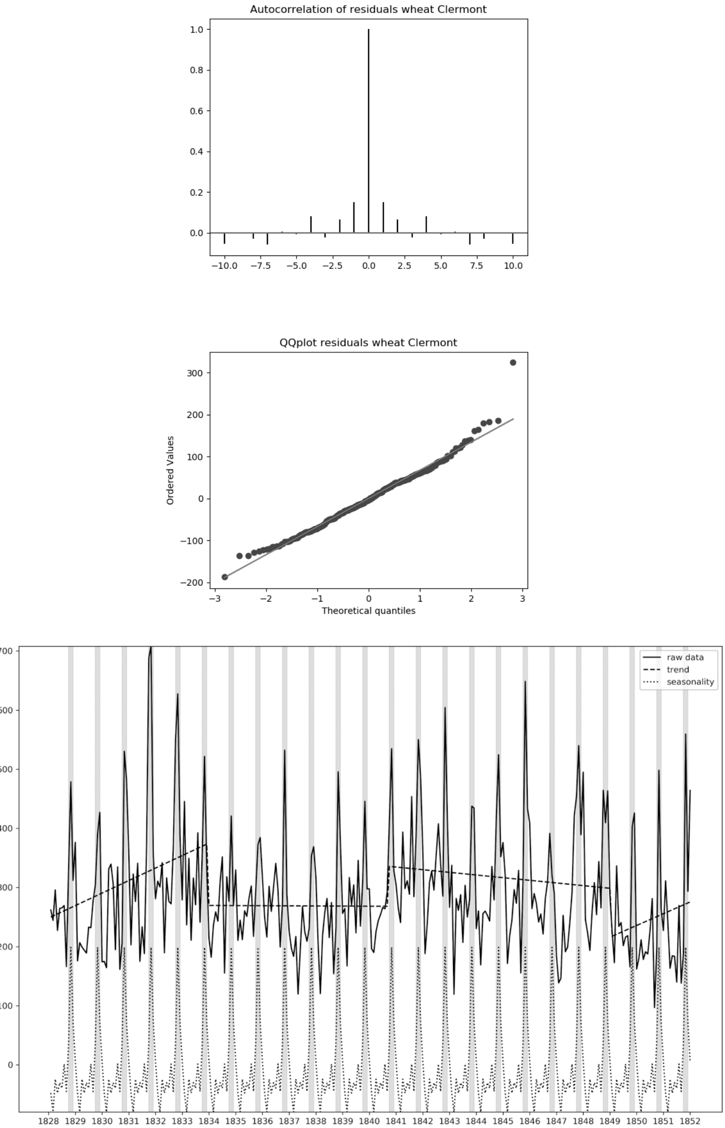
Although this is not a tool as precise as the seasonality signal, the ratio $\frac{\max \sigma_m}{\bar{t}}$ between maximum seasonality on one hand, and the average trend on the other hand allows to catch the importance of seasonal variations. Here we display the results for the five biggest markets, since we have of a full sequence over 25 years for those. For instance, we see that in Clermont, the average seasonal peak (in October) is 68% higher than the global average for wheat and almost 40% for maslin.

GRAPH A-1
Ratio between seasonal variation and global average



Source: see table 1.

FIGURE A-1
Example of residues (sales of wheat in Clermont between 1828 and 1852)



Note: the auto-correlation is very weak and the quantiles are close to those of a Gaussian distribution so we can think that we have identified a trend (flat) and a seasonality (very marked peak around October) which capture the essential information. Below the two residuals analyses, the raw signal, the linear model of the trend and the seasonality estimator are represented on the same graph.

Source: see Table 1.