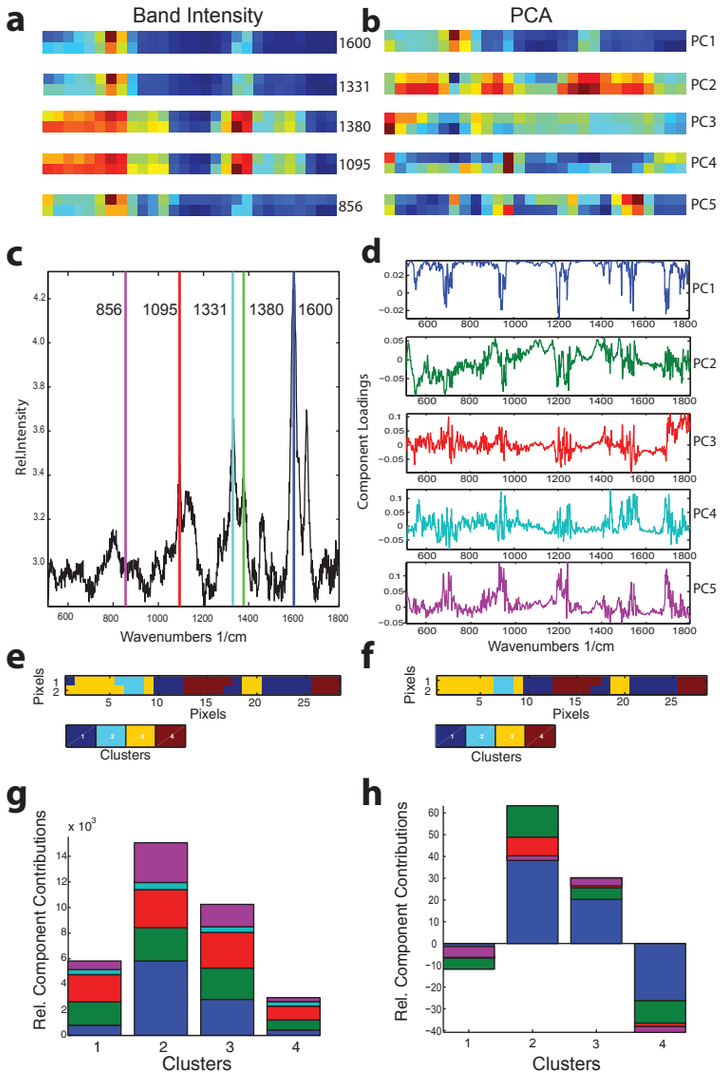

Comparing the results of band intensity heat maps (left column) and principal component analysis (PCA, right column), followed by k-means clustering, to MCR-ALS.

The same, non-normalized input data was used as in Supplementary Figure 2, with the same pre-processing, to allow direct comparison to MCR-ALS. Bands were chosen to represent lignin (1600 and 1331 cm-1), cellulose (1380 and 1095 cm-1) and pectin (856 cm-1), according to N. Gierlinger et al.14 PCA analysis was carried out using five principal components, which described 97.9% percent of the variation. The number of principal components were chosen to facilitate direct comparison to MCR-ALS (number of components) and to band intensities (number of bands).
a and b) Distribution maps based on band intensities and scores, respectively. Note that the maps based on band intensities can be correlated to certain cell wall zones (influenced by total intensity variations) and the bands assigned to the same compounds provide similar maps. On the other hand, maps based on scores (b) are harder to interpret, except for the first principal component (PC1).
c) An example spectrum (taken from pixel 7), marking the band positions used in the band intensity analysis. Colors match the ones used in the centroid profile (part g). Note the extreme low intensity of the 856 cm-1 band, assigned to pectin, in this pixel.
d) Loadings of principal components 1-5 (PC1-5). Note that these are abstract representations of the spectral profiles, and thus spectral band identification is difficult, except for PC1. In addition, all loadings (including that of PC1) contain negative values, which makes the interpretation of the centroid profiles (part h) and reference spectrum matching extremely difficult.
e and f) Segmentation maps, following k-means clustering with 4 clusters, based on the distribution maps in a and b, respectively. Note that the identified zones are matching very closely in both cases, and even when compared to the MCR-ALS results (Supplementary Figure 2e), with the exception of a few borderline pixels.
g and h) The corresponding centroid profiles, showing the contribution of different components to each cluster. Interpreting the results (i.e. identifying the contribution of any particular chemical compound, such as lignin or cellulose to any given cluster) is practically impossible in the case of PCA (h). Interpretation of the band intensity results (g) is considerably easier, assuming diagnostic bands (dark and light blue: lignin, red and green: cellulose, purple: pectin). While interpretation is easy, the results are erroneous. The pattern in the varying lignin amounts is reasonable, due to the fact that the 1600 cm-1 band (dark blue, originating from aromatic skeletal vibrations), is not shifting or overlapping and is very diagnostic. However, the contribution of the other lignin assigned peak (1331 cm-1, light blue) is much less significant, despite the fact that this is a very intense band (see c) and follows the same pattern as the 1600 cm-1 band (see a). Conversely, the contribution of the extremely small (virtually non-existent) pectin band (purple) is overestimated. Comparing this to the contribution of the red (“noise”) component in the MCR-ALS model in Supplementary Figure 2g reveals the strength of MCR-ALS in filtering off noise and using pure spectral profiles. Most critically, the estimation of cellulose contribution by band intensities results in errors. According to the centroid profiles, the amount of cellulose is practically constant in all cell wall layers (see green and red contributions in clusters 1-3). Thus falsely suggests that the lignin to cellulose ratio is (correctly predicted to be significantly different between the middle lamella and the g-layer) is only different because the lignin amount varies. In contrast, the results obtained by MCR-ALS (Supplementary Figure 2g) clearly show increased amounts of cellulose (green) in the g-layer (cluster 1) as compared to the middle lamella (cluster 2), as well as correctly predicting the dramatic differences in lignin amounts (blue) in these zones. This again underlines the importance of using full spectral profiles instead of single band intensities and illustrates how MCR-ALS combines the strengths of the two methods: full spectral profiles are used as in the case of PCA, while interpretation is as easy as in the case of band intensities.