Analysis of Pre-Germinated Barley using Hyperspectral Image Analysis.

TECHNICAL REPORT v1.0

Morten Arngren

Technical University of Denmark, DTU Informatics, Bldg. 321, Richard Petersens Plads, DK-2800 Lyngby
and FOSS Analytical A/S, Slangerupgade 69, DK-3400 Hillerød
ma@imm.dtu.dk,moa@foss.dk, info@arngren.com

Mikkel N. Schmidt

Technical University of Denmark, DTU Informatics, Bldg. 321, Richard Petersens Plads, DK-2800 Lyngby
mns@imm.dtu.dk

Jan Larsen

Technical University of Denmark, DTU Informatics, Bldg. 321, Richard Petersens Plads, DK-2800 Lyngby
jl@imm.dtu.dk

Per Waaben Hansen

FOSS Analytical A/S, Slangerupgade 69, DK-3400 Hillerød
pwh@foss.dkk

Birger Eriksen

Sejet Planteforædling, Nørremarksvej 67, DK-8700 Horsens
bee@sejet.com

Version History

Version	Date	Change Revision
1.0	Feb. 2011	Initial issue.

Contents

1 Introduction

This technical report acts as an appendix and supplement to the Journal paper entitled 'Analysis of Pre-Germinated Barley using Hyperspectral Image Analysis.' [1]. It is not self-contained as only relevant sections are included in this report.

If you should have any comments or corrections, please forward any inquiry directly to Morten Arngren (ma@imm.dtu.dk or info@arngren.com).

2 Scatter Correction

A common effect of image acquisition systems is the light scattering occurring especially for penetrating NIR wavelengths. The scattering effect are mainly be introduced as a subsurface scattering or due to the physical shape of the subject, in this case barley kernels. The scattering is manifested as undesired components in the acquired spectra. This means the scatter correction method acts both as a scatter removal tool in a pre-processing step and as a feature extractor.

This section describes three scatter correction approaches evaluated in detail, ie. *Standard Normal Variate (SNV)*, 0th and 1st *order polynomial detrending*. Common to these methods is that they operate on the M-dimensional observed spectra $\mathbf{S}^{raw} \in \mathcal{R}^{M \times N}$ separately, where N is the number of spectra observed.

2.1 Standard Normal Variate

The SNV correction is among the simplest methods, as it both mean centers and normalizes each spectra separately. If we denote a single observed spectrum $\mathbf{s}_i^{raw}$ the SNV correction can be expressed by

$$\mathbf{s}_i^{snv} = \frac{\mathbf{s}_i^{raw} - \mu_i}{\sigma_i}, \tag{1}$$

where μ_i is the scalar mean value and σ_i is the associated standard deviation for the spectra $\mathbf{s}_i^{raw}$.

2.2 Detrending

In applying detrending a predefined basis spectrum is subtracted from the observed spectrum $\mathbf{s}_i^{raw}$.

$$\mathbf{s}_i^{detrend} = \mathbf{s}_i^{raw} - \mathbf{s}_i^{basis}, \tag{2}$$

where $\mathbf{s}^{basis}$ denote the basis spectrum, which can be a single background spectrum or consist of a combination of several spectra both dependent and independent of the observed spectrum $\mathbf{s}_i^{raw}$. For certain applications the spectrum of water is known and is removed from all observations if desired. In our case the basis spectrum is made from n'th order polynomials fitted to each observed spectra separately. This means for 0th order detrending the mean constant is subtracted and similarly for 1st order detrending a sloping line is subtracted.

2.3 Applied Scatter Correction

The acquired images are subjected to the three scatter correction techniques as a preprocessing step on the spectral information. Figure 1 illustrates a few raw spectra and their scatter corrected equivalents.
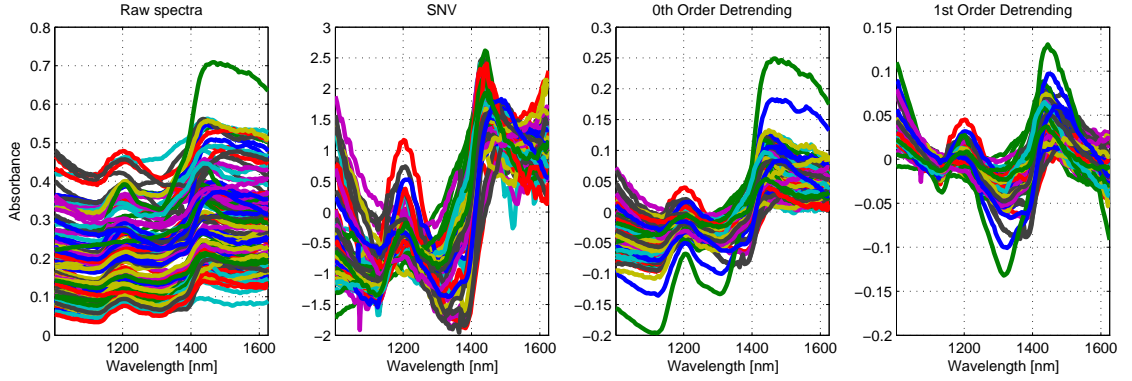

Fig. 1: A subset of raw spectra from an acquired image in the left most illustration showing the massive scatter effect. The remaining depictions show the impact of the different scatter corrections.

To compare the scatter correction methods for our application a set of PC score images for each scatter correction method are evaluated in terms of their ability to capture the chemical changes inside the kernels during the germination process. Tensors from one images for each germination time were concatenated and unfolded to a $132 \times N_{samples}$ matrix $\mathbf{X}_{pca}$ such that each mean centered spectrum is represented as a column, where $N_{samples} = 8$ germ. durations $\times$ 45 kernels per image $\times$ 400 avg. pixels per kernel ≈ 140000 spectra. An unfolded matrix is calculated for each type of scatter corrected data and subjected to PCA providing score images for the evaluation. Figure 2 illustrates the PC score images for the scatter correction methods.

The SNV scatter correction is known to equalize structures in the data due to the scaling and in this case tends to emphasize the edges of the kernels. The only clear captured structure is a strong contrast of the embryo in PC score image 5. The two detrending approaches show similar decompositions. Both has captured the husk of the kernels in the first components as a intensity difference. Activity at the embryo of the kernels is represented in different components. The 0th order pol. detrending extracts this activity in the 3rd and 5th component, whereas the 1st order pol. detrending has captured this feature in the 4th component. Based on the visual evaluation the SNV is used for scatter correction in order to obtain a strong pre-germination feature as separately as possible.

3 Data Compression

The acquired spectra consists of 132 bands and the relevant information for our analysis is expected to span a subspace. Regular PCA is hence used to compress the data to reduce the computational load and to suppress the noise, but at the risks of removing information important for the final classification.

A set of projection loadings $\mathbf{U}_{pca}$ are estimated from meancentered spectra from one hyperspectral image from each germination duration becoming a $132 \times \sim 140000$ matrix. The calculated loadings are then afterwards applied to the rest of the data set. Figure 2 (middle) and 3 show the PCA score plots $1 - 5$ and $7 - 11$ respectively.
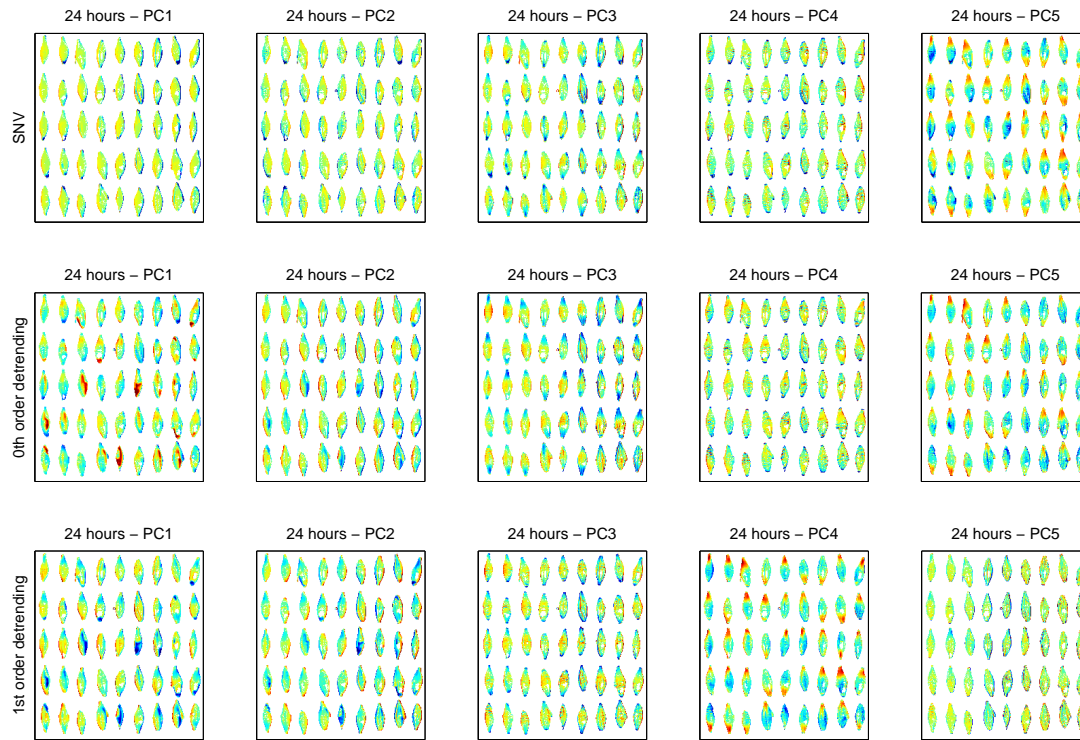
Fig. 2: Individual PC score images for each of the scatter corrections: *SNV* (top), *0th order detrending* (middle) and *1st order detrending* (bottom).
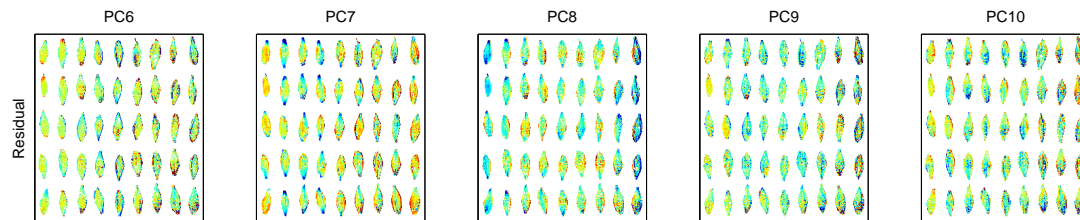

Fig. 3: PC score images from 6-10 represent insignificant structure in the image and hence PC's above 8 are disregarded.

The score plots indicate the majority of the structure is covered by the first 8 principal components. To ensure the relevant information is retained in the data, while achieving a suitable compression, the data is projected onto the first 8 PC's. These components further explain 99% of the variance.

4 Feature Extraction

This section presents the three decompositions, *Principal Components Analysis (PCA)*, *kernel PCA* and *Minimum/Maximum Noise Fraction (MNF)*, used for features extraction to capture the germination progression over time in the acquired hyperspectral images.

4.1 Principal Component Analysis

The first and most simple of the three decomposition methods is the principal component analysis (PCA) applied to the spectra. The PCA decomposition maximizes the variance in the spectra and does not consider any spatial information in the images. The extracted PCA features are described in the main article [1].

4.2 Kernel Principal Component Analysis (kPCA)

Regular linear PCA is based on inner products to estimate the covariance matrix of the spectral data. The kernel trick framework exploits this fact and can be used to transform the image data into a higher dimensional feature space via a non-linear mapping function $\mathbf{x} \mapsto \mathbf{\Phi}(\mathbf{x})$ applied to the inner products $\mathcal{K} \propto \mathbf{\Phi}\mathbf{\Phi}^\top$ [5]. Applying the kernel trick means the mapping function $\mathbf{\Phi}$ is actually unknown, only the kernel $\mathcal{K}$ is are calculated as the inner products, which in our case is given by the Gaussian kernel

$$k(x_i, x_j) = \exp\left(-\tfrac{1}{2\sigma^2} \parallel x_i - x_j \parallel^2 \right), \tag{3}$$

where σ is a predefined scaling parameter. Extracting the score images requires processing in the non-linear space due to the mapping function, ie. $\mathbf{\Phi}\mathbf{U}$, where $\mathbf{U}$ is the loading vector in feature space. Since this is not directly feasible Larsen *et al.* [5] shows how the projection can be calculated by

$$\mathbf{\Phi}\mathbf{U} = \mathbf{K}\mathbf{V}\mathbf{\Lambda}^{-1/2}, \tag{4}$$

where $\mathbf{V}$ holds the loadings for the kernel covariance matrix $\mathcal{K}$ and $\mathbf{\Lambda}$ represent the associated eigenvalues. The kernel PCA decompositon is memory based, which has the disadvantage of requiring all training data to be available, when projection new data. This means the kernel PCA has a higher computational load when applying to our image data.

The spectral loadings are estimated from a subset of the data consisting of one images from each germination duration, ie. $\mathbf{X}_{train}$ becomes a $8\times \sim 140000$ matrix. Figure 4 depicts the score image showing the clearest progression in the germination process, but it unfortunately includes the other components such as the barley husk and hence does not represent the progression as clean as PCA or MNF score images.

4.3 Minimum/Maximum Noise Fraction (MNF)

A different linear decomposition method is the *Minimum/Maximum Noise Fraction (MNF)*. It has received wide attention in the hyperspectral imaging community as it typically represents the structure in the data with fewer components than regular PCA. The MNF approach uses neighbour pixel dependencies and exploits both the spectral and spatial information in the images to estimate more relevant directions for explaining most
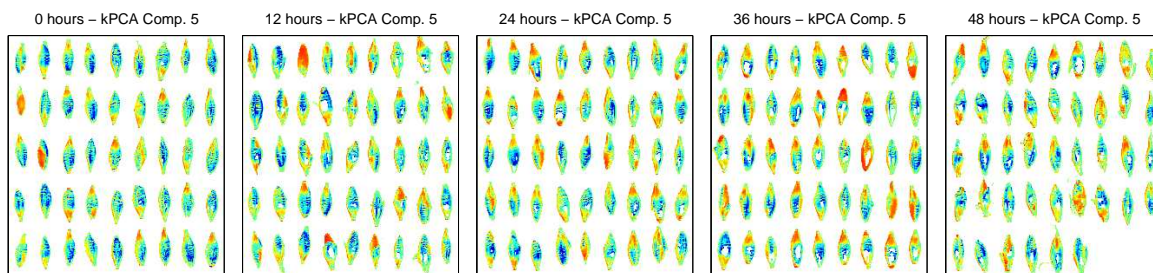
Fig. 4: The 5$^{\text{th}}$ kPCA score images for the germination durations $[0, 12, 24, 36, 48]$ (left to right). The component does not have the same pure representation as in the PCA or MNF score images.

variance in fewer components. The MNF estimates a set of loading vectors $\mathbf{a}_i$ maximizing the signal-to-noise ratio ρ defined as the Rayleigh quotient expressed by

$$\rho = \frac{\mathbf{a}^\top \Sigma \mathbf{a}}{\mathbf{a}^\top \Sigma_\Delta \mathbf{a}}, \tag{5}$$

where Σ_Δ denote the estimated noise covariance matrix and Σ is the sample observation covariance matrix. The noise covariance matrix Σ_Δ is estimated under the assumption that two neighbouring pixels are similar and only differ primarily due to noise. Σ_Δ is then estimated as the neighbour differences for each pixel both vertically and horizontally taking the spatial mask into account and pooled into single covariance matrix [3] [2]. The MNF loading vectors $\mathbf{a}_i$ can then be estimated by simple generalized eigenvalue decomposition [5].

The MNF transform is closely related to the Maximum Autocorrelation Factor (MAF), where the noise covariance matrix is interpreted at the pixel correlation matrix and the estimated loadings point to the maximum autocorrelation.

5 Germination Time Classification

The classification of the germination time can be conducted in many different approaches. In the main article the germination times were classified directly by a multinomial classifier. In this section alternative approaches are shortly investigated.

5.1 Ordinal Classification

The germination time references exhibits a natural ordering, since class 12h are lower than class 18h etc. Regular classifiers do not utilize this information and risk providing error distributions, where misclassifications are more than one class away from the ground truth. For ordinal data the error distribution should instead only allow misclassifications to neighbour classes.

Regression models incorporates a cost function, where predictions far from the ground truth are penalized hard, e.g. least-squares loss function. The ordinality can therefore be exploited by including a regression term prior to the classification.

The ordinal classifier framework includes a prediction step to capture the progression in the references followed by a multinominal regression classifier providing a probability for each class for each barley kernel. The prediction of the germination times are modelled by a regularized non-linear artificial neural network (ANN) [4] [6] [7] [8] [9]. The ANN prediction model is structured as a two layer feed forward network and includes two parameters. Initially the number of hidden units N_{hu} to define the network structure and secondly a regularization parameter λ to prevent overfit and force a more linear performance. The `MATLAB` implementation used is the `ANN:DTU Toolbox` from DTU Informatics[1]. The subsequent classification model used is the same linear maximum likelihood multinominal regression classifier as in the main article [1].

The optimal hyperparameters N_{hu} and λ for the ANN are estimated in a leave-1-time-out cross validation framework using a single training/test set and applying small range of parameters, $N_{hu} \in [5 - 18]$ and $\lambda \in [0.5, 1, 2, 5]$. The lowest least-squares prediction error was found for $N_{hu} = 10$ and $\lambda = 1$. Using the optimal hyperparameters, a different training/test set is extracted 100 times for each validation sample and a corresponding ANN model is inferred for each training/test set. The average germination time predictions illustrated in Figure 6 show how there is a large overlap between the times. During the germination process the starch is broken down at an accelerating rate after approx. 24 hours, corresponding to the positive slope in Figure 5. This means the kernels within the same germination time can exhibit different germination rates leading to wrong associated time labels. This is seen by the relatively high prediction errors. The last germination time 60h in Figure 6 is predicted close to 48h. indicating minor differences in the germination progression.
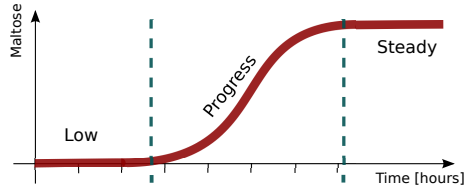

Fig. 5: Conceptual illustration of the progression of germinating barley showing the starch break down. The two thresholds indicate the start and stop of the major chemical changes inside the kernel during germination.
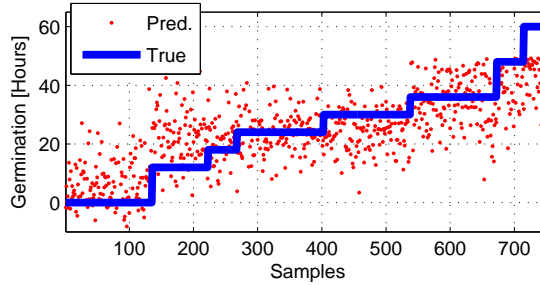

Fig. 6: The predicted and true germination time for each of the 755 kernels. The ANN model has captured the overall structure of the germination time despite the large prediction variance. The similar predictions of the last two classes 48h and 60h suggests the germination process begins to level out after 48h of germination.

An associated classifier is inferred for each of the 100 ANN model using the predicted germination times. The validation classification error for all single kernels in all eight classes become 63% and 33% for the three

---

[1] Available for download at `http://cogsys.imm.dtu.dk/toolbox/ann/index.html`

aggregated classes. This can be compared to the random guess error rate of $7/8 = 87.5\%$ and $2/3 = 66.7\%$ respectively.
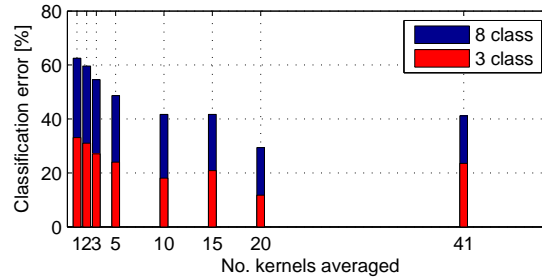

Fig. 7: Classification errors for different amount of averaged kernels. The error clearly decreased when averaging more kernels. Averaging kernels leaves fewer samples to analyse and hence higher sensitivity to our results. When averaging 41 kernels the error is based on only 17 samples.

The class probabilities for each kernel can be averaged to estimate bulk classification results for a small range of combinations as shown in Figure 7. The classification error clearly decreases and is minimum by averaging 20 kernels with a error of 29% and 12% for eight and three classes respectively based on 34 samples. Since we only have 755 kernels averaging them leads to even fewer samples and less accurate results. This is evident for averaging 41 kernels, which has only 17 samples and a higher classification error.

The classification distributions for eight and three classes for 20 averaged kernels are illustrated in Figure 8. The three class validation error of 12% corresponds to 4 misclassifications out of 34 samples and is thus sensitive to the amount of samples.
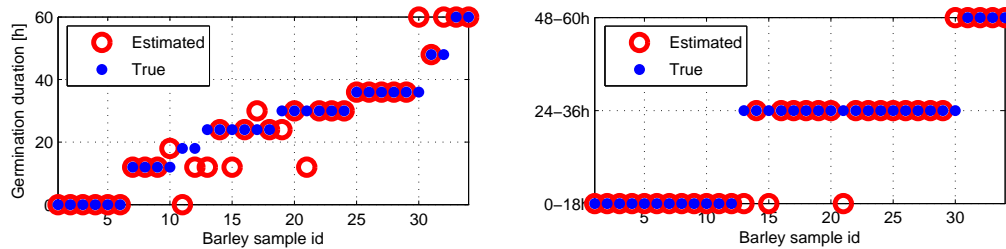

Fig. 8: Error distribution for averaging 20 kernels for all eight classes and the three aggregated classes, left and right respectively.

5.1.1 Linear Prediction

The classification of germination times is modelled using an ordinal classifier in order to capture the progression in the references. Our ordinal classifier framework is implemented by conducting the classification on predicted germination times instead of the features directly. This approach ensures the progression is modelled in the prediction step.

In the main article, an Artificial Neural Network (ANN) is used to predict the germination times. A more simple approach is to include a linear model with a regularization term. Using the same notation as in [1], the weights $\mathbf{W}$ of the linear model can be inferred as

$$\mathbf{Y}^{pred} = (\hat{\mathbf{F}}\hat{\mathbf{F}}^\top + \alpha \mathbf{I})^{-1}\hat{\mathbf{F}}\hat{\mathbf{Y}}, \tag{6}$$

where $\hat{\mathbf{F}}$ and $\hat{\mathbf{Y}}$ denote the features and references subset used for training respectively, $\mathbf{Y}_{pred}$ are the estimated germination times and α represents the degree of regularization. The regularization term limits any poorly estimated directions in calculating the covariance matrix $\hat{\mathbf{F}}\hat{\mathbf{F}}^\top$ due to few samples.

The optimal regularization parameter α is estimated in a leave-1-time-out cross validation framework using 100 different training/test sets and applying small range of parameters, $\alpha \in [0, 1e9, 1e-8, 1e-7\ 1e-6\ 1e-5\ 1e-4\ 1e-3\ 1e-2\ 1e-1\ 0.2\ 0.5\ 1\ 2]$ as in [1]. The lowest least-squares prediction error was found for $\alpha = 0$, ie. no regularization. Using no regularization, a different training/test set is extracted 3000 times for each validation sample and a corresponding linear model is inferred for each training/test set.

The average germination time predictions illustrated in 9 show how linear predictions are comparable with the non-linear ANN model. Using the predictions in the same classification framework as before, we achieve a set of higher classification error compared to the ANN model, cf. Figure 10. The error decreases little as the kernels are averaged and does not reveal a strong minimum. The lowest error at 20 averaged kernels is 59% and 29% for the eight and three classes respectively.

The classification distributions for eight and three classes for 20 averaged kernels are illustrated in Figure 11. The three class validation error of 29% is by far inferior than the classification error achieved using the ANN predictor.
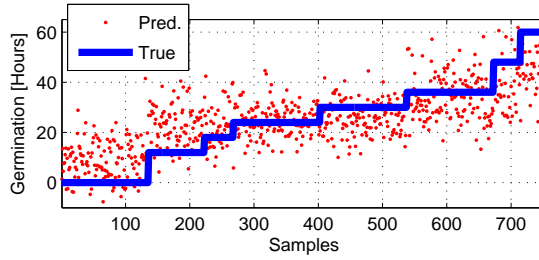

Fig. 9: The predicted and true germination time for each of the 755 kernels using the linear predictor. The linear model has captured the overall structure of the germination time despite the large prediction variance and is in general comparable with the ANN model.
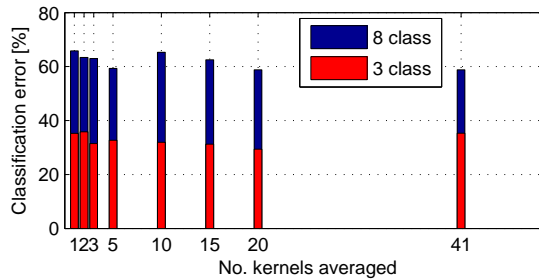

Fig. 10: Classification errors for different amount of averaged kernels based on the using the linear prediction model. The error clearly decreased when averaging more kernels. Averaging kernels leaves fewer samples to analyse and hence higher sensitivity to our results. When averaging 41 kernels the error is based on only 17 samples.
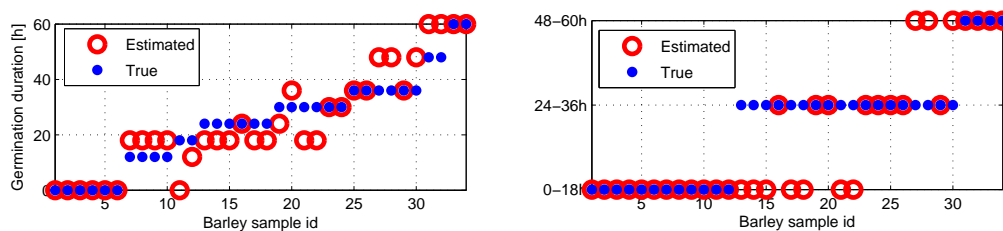
Fig. 11: Error distribution for averaging 20 kernels using the linear predictor for all eight classes and the three aggregated classes, left and right respectively.

6 Discussion

From a macroscopic viewpoint the α-amylase breaks the starch down from the outside and in and hence allows the hyperspectral NIR camera technology to capture these biochemical changes. In contrast if the starch break down would have been initiated from the center of the kernels, it would most likely not have been possible to capture the germination until a later stage in the germination process.

References

1. Arngren, M., Larsen, J., Hansen, P.W., Eriksen, B., Larsen, R.: Analysis of pre-germinated barley using hyperspectral image analysis. Journal of Agricultural and Food Chemistry (2010)
2. Gordon, C.: A generalization of the maximum noise fraction transform. IEEE Transactions on Geoscience and Remote Sensing **38**(1, part 2), 608–610 (2000)
3. Green, A.A., Berman, M., Switzer, P., Craig, M.D.: A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing **26**(1), 65–74 (1988)
4. Larsen, J.: Design of Neural Network Filters. Ph.D. thesis, Electronics Institute, Technical University of Denmark (1993)
5. Larsen, R., Arngren, M., Hansen, P.W., Nielsen, A.A.: Kernel based subspace projection of near infrared hyperspectral images of maize kernels. Lecture Notes in Computer Science **5575/2009**, 560–569 (2009)
6. MacKay, D.J.C.: A practical Bayesian framework for backpropagation networks. Neural Computation **4**, 448–472 (1992)
7. Nielsen, H.: Ucminf - an algorithm for unconstrained, nonlinear optimization. Tech. Rep. IMM-TEC-0019, IMM, Technical University of Denmark (2001)
8. Pedersen, M.: Optimization of Recurrent Neural Networks for Time Series Modeling. Ph.D. thesis, Institute of Mathematical Modeling, Technical University of Denmark (1997)
9. Svarer, C., Hansen, L., Larsen, J.: On Design and Evaluation of Tapped-Delay Neural Network Architectures. In: Proceedings of the 1993 IEEE International Conference on Neural Networks, pp. 46–51. New York, New York (1993)