Geoinformation for European-wide Integration, Benes (ed.)

Yield and quality estimation of malting barley based on remote sensing and GIS

K. Schelling, C. Weissteiner, K. Hünting & W. Kühbauch
Department of Crop Science and Plant Breeding, University of Bonn, Germany

Keywords: malting barley, yield forecast, yield model, quality

ABSTRACT: Timely information about the production of malting barley is of great significance for the malting industry. The supply situation depends on the area cultivated to spring barley, yield, protein content and the graded proportion, which is suitable for malting. Yield and quality are strongly influenced by environmental factors – weather being the most important. In the current project it was investigated how optical-near infrared satellite images and non-remotely sensed data can be utilized in an operational GIS-based system for the estimation of barley yields and quality.

Phenological observations turned out to be a valuable source of information. The duration of the grain filling period derived from these data proved to be temperature dependent. While the most meaningful parameter for yield estimations is the mean daily temperature between ear emergence and yellow ripeness, the relative humidity proved to be an informative parameter to assess the protein content. Knowledge about the length of this developmental period allowed assessments of the yield level, by means of which quality parameters could be reasonably estimated. In addition the analysis of monthly NDVI-maximum value composites from coarse resolution NOAA-AVHRR data, starting with the onset of grain filling, indicated a good relationship to the regional average yields.

By means of multiple linear regression approaches, using phenological and meteorological data, all three target factors – regional average yields, protein contents and screening percentage – were predicted with deviation to the observed values of well below 5 % at the time of yellow ripeness. For an operational application it is recommended to integrate the various information layers remote sensing and weather data into a Geographical Information System.

1 INTRODUCTION

To supply the need of German breweries nearly 2 million tons of malt per year are required. Out of that approximately 115 millions hectoliters of beer are made. As raw material mostly spring barley is used. The cyclical yield variation is considerable: in good years the oversupply is up to 0.6 million tons, whereas in other years up to 1 million tons of malt have to be imported. Since preliminary contracts between suppliers and breweries are often fixed in advance (1 to 1.5 years), shortages in raw materials are most likely. These shortages can lead to substantial losses for the brewing and malting industry.

The supply of appropriate spring barley is depending on acreage, yield, the protein content and the graded proportion which is suitable for malting. The protein content, being the most important parameter for high quality malting barley, should range between 9.5 and 11.5 % (w/w). Variations of yield and quality are leading to considerable price fluctuations. Therefore malting industry is interested in getting an overview of the development of barley, the expected yield and quality at an early stage of growth.

Yield and quality depend on a multitude of factors with complex interactions among each other. Satellite data can provide the needed input information for a yield forecast model, although additional earth borne data will be necessary. The synergetic use of both earth borne and remote sensing data could lead to a significant improvement of yield and quality models.

For the description of the development of plant stocks many data layers can be used, whereby all data, except remote sensing, are also mentioned as conventional or ancillary data. All data is spatially related and it is therefore opportune to keep and handle data in a Geographic Information System (GIS).

The aim of the present work was to define the requirements for a future malting barley information system, to investigate and detect the main components and analyze their impact on a yield and quality forecast model. Additionally, it should be pointed out how various data layers could be handled and embedded in a GIS-environment.

2 MATERIAL AND METHODS

Cultivations of malting barley are often concentrated in certain regions, according to favourable climatic conditions (see figure 1) . Two regions in Germany were selected, where the required data was expected to be available or where satellite data was already existing.

Regions:

- Northern Rheinland-Pfalz, Germany (abbreviated as RLP), approx. center: 50° N, 7° E, approx. area: 7,000 km².
- Lechfeld (abbreviated as L) in Bavaria, Germany, approx. center: 48°12’ N, 10°45’ E.

Remote sensing data:

Low resolution remote sensing data:

The database for the region RLP were composites of NDVI of NOAA-AVHRR, downloaded from the German Remote Sensing Data Center (DFD). NDVI-composites are neither atmosphere corrected nor corrected for topographic impacts. However, operational availability of corrected NOAA-AVHRR NDVI-composites is expected in the near future (Ebertseder et al., 1999). After implementing the respective stereographic projection NOAA-AVHRR images were georeferenced and integrated into a GIS-system. By means of agricultural statistics and expert knowledge regions of interest for malting barley cropping have been identified and digitized. These vectors were used to determine mean NDVI-values for specified cropping regions.

Due to haze and clouds in the NOAA-AVHRR images (8 bit coded) and the resulting falsification and lack of data, the images were examined visually and the respective area mean values corrected where necessary. Up to an estimated partial cloudiness of 15 % of the area one grey value for each percent cloudiness was subtracted. If an area was affected by haze, an increment of up to 5 grey values, depending on haze occurrence, was performed. If cloud occurrence was too high, a manual interpolation was made between utilizable capture dates. Obviously this subjective method must lead to a considerable variance of the output results. However, a similar approach to this is performed by the USDA (Bethel, 1998).

Figure 1. NOAA-NDVI monthly composit of July 1994 with the most important german cropping regions of malting barley, two of them being the test sites (Nördliches Rheinland-Pfalz and Lechfeld).

High resolution remote sensing data:

Three Landsat TM quarter scences of 21.4.96, 7.5.96 and 8.6.1996 of the region Lechfeld (L) and one SPOT XS scene of 22.7.96 were available as high resolution images. One TM scene and the SPOT scene were georeferenced by topographic maps (scale: 1 : 25,000) and projected into the Gauß-Krüger-projection (4^{th} meridian stripe system), the other two TM scenes were co-registered on the first TM scene. RMS-error was below 1 pixel. These scenes were corrected atmospherically with data of the surrounding meteorological stations.

Ancillary data:

Topographic data:

Topographic maps of a scale of 1 : 25,000 (TK 25) were used for georeferencing, for localizing and for digitizing the test sites in the region Lechfeld (L). Additionally, topographic maps of a scale of 1 : 5,000 (DGK 5), and soil maps of a scale of 1 : 10,000 were used. Furthermore, administrative boundaries were implemented in the existing GIS. A digital elevation model (DEM) in a resolution of 30 seconds (GLOBE) was integrated into the GIS as a base for the existing NOAA-AVHRR data. Moreover, digital land cover data was used (CORINE and regional maps) for the useful feature “arable land”.

Soil data:
Out of a regional digital soil map (1 : 50,000) of Lechfeld a thematic layer "water capacity" was generated.

Agrostatistical data:
These data are recorded on the level of counties (NUTS III level), sometimes on a lower level (municipalities). By means of agrostatistics it was possible to determine main cropping regions and yield per area for each region. Furthermore, time series could be generated for spring barley production.

Phenological data:
These data is reported by the German meteorological service, Deutscher Wetterdienst (DWD), for various phenological stages: seedling, emergence, stem elongation; ear emergence, yellow ripeness and yield. These stages are related to homogeneous natural areas, which obviously differ from the administrative areas. Since spring barley is reported only up to 1990, data after this year had to be extracted out of a correlation between spring barley and oat. Former results showed this to be an acceptable way (Schelling, 2000).

Since agrostatistical data was just available on an annual basis, meteorological data had to be aggregated for stages per year for comparison purposes. These stages are listed in table 1.

Table 1: phenological periods

Period	Name	acronym
Beginning of year to seedling	Preseedling period	PS
Seedling to emergence	emergence	EM
emergence to stem elongation	Youth stage	YS
Stem elongation to ear emergence	Mass growing	MG
Ear emergence to yellow ripeness	Grain filling period	GF

Meteorological data:
The data provided by the DWD was extracted from weather and climate stations within the area of interest. Data was then processed and aggregated by county and by the desired phenological phase. The parameters used are listed in table 2:

Table 2: meteorological parameters

parameter	acronym
Daily average temperature	DT
Daily maximum temperature	MaxT
Temperature sum	SumDT
Growing degree days (base 3 °C)	GDD
Precipitation sum	PrS
Daily average air humidity	AH
Sum of Crop Water Stress Index (CWSI)	CWSI
Additionally the precipitation sum from beginning of year until yellow ripeness was calculated (P_tot)	

Yield and quality data:
In addition to agrostatistical data, which do not contain any quality data of malting barley, for the region RLP results from field trails conducted by a malting company on up to 10 different locations were used. For the years 1974 to 1997, with exception of the year 1985, average yield, protein content and screening percentage were available.

For the region L these data was recorded by a questionnaire.

Model*:*
In the present investigation an empirical-statistical model was applied, based on statistical data between 1974 and 1996. The application of this model is limited to a regional extent, .i. e. to the regions mentioned above (RLP and L). The relationships were extracted by simple and multiple linear regressions. For statistical analyses the meteorological data was accumulated by phenology phases.

3 RESULTS AND DISCUSSION

The analysed parameters and factors were tested for their significance in the forecast model with respect to their impact on yield and quality.

High resolution images:
High resolution images, i. e. spatial resolutions of <30 m yield results on a field level. Although any classification was done at this level, these resolution is necessary to detect and to differentiate various crops. However, Landsat TM and SPOT XS data were used to perform investigations on the relationship of NDVI (normalized and differentiated vegetation index) and statistical yield data on field level. The best results were obtained by comparing yield or protein concentration and accumulated NDVI sums of the latest two images (8.6.96 and 22.7.96) by a simple linear correlation. Depending on soil water capacity and on whether images were atmospheric corrected or not, the correlation coefficient r ranged between 0.43 and 0.85.

Low resolution images:

Low resolution images like NOAA-AVHRR (spatial resolution 1.1 km x 1.1 km) supply general information about the vegetation vigor. By means of NDVI values generated out of these images a coarse survey of potential yield on a regional, national or global scale is possible. Since NDVI-composites were not available before 1994, statistically reliable hypotheses were not possible, however these data showed promising results for a simple and fast yield forecast. The best results were obtained by a correlation of yearly yield per area and area averaged NDVI values for July. This comparison resulted in a coefficient of determination $r^2 = 0.97$. Correlations with June and May composites resulted in lower correlation coefficients. It has to be accounted for, that these correlations are not statistically reliable, since only 4 –5 pairs could be compared (1994/95 – 1998). Figure 2 shows NDVI and yield correlations for the region RLP.

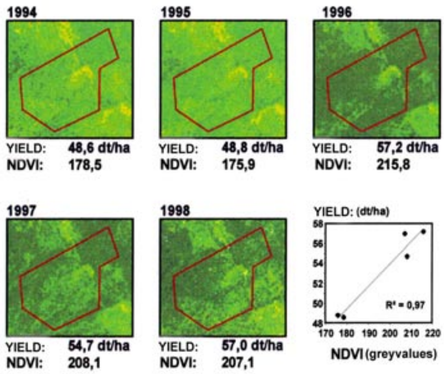

Figure 2: Subsets of the July composite of NOAA-AVHRR NDVI for 1994 to 1998 of the test region RLP. NDVI-values (-0.1 – 0.7) displayed in 8 bit grey values (bright = low NDVI, dark = high NDVI).
Correlation of average malting barley yield and NDVI composite values of July for 5 years (1994 – 1998).

Topographical data:

Masking techniques are useful for limiting data to a certain extend. By means of the CORINE land cover dataset (original scale 1 : 100,000) a coarse mask for arable land can be generated in principle, although datasets in a lower scale, like the one in a regional scale available for the region Lechfeld, would be desirable, especially for classification purposes.

A DEM can be used to stratisfy images for spectral analyses. The impact of elevation on vegetation development and spectral properties has been previously investigated for the region RLP in a previous work (Kühbauch et. al., 1998). A generation of plant development, based on Growing Degree Days (GDD) and a DEM is possible. Unsuitable areas, if due to their altitude, can be excluded from further processing.

Soil data:

Since the nutrient demand of spring barley is relatively low, soil data analysis focused on water supply and water capacity rather than on nitrate. The soil map for the region Lechfeld (L), scale 1 : 50,000, was sufficient to differentiate between different soil elements by the criterion water capacity.

Agricultural statistics:

Since the cultivation of malting barley is not distributed homogeneously over a specific area, agricultural statistics are valuable for determining the main cropping areas. Knowing about the principal cropping regions, data supply can be restricted and costs can be saved. Furthermore, by time series, a trend analysis of yield can be conducted, and thus the partition accounting for “not environmental” variance, but variance due to technical progress, especially breeding, being excluded (Hanus & Aimiller, 1978). The following figure gives an idea about trend corrected and not trend corrected yields.

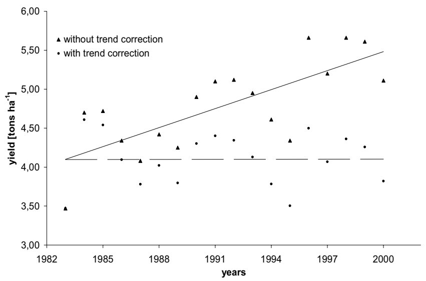

Figure 3: Not trend corrected and trend corrected spring barley yield per area for the years 1983 – 2000. Data is result from current investigations on yield in the region RLP, county “Südliche Weinstrasse”, Germany.

Correlation between yield and quality:

Literature review and the found results showed a significant correlation between yield quantity and quality. With rising temperatures the grain yield of barley is diminishing (Eagles et al. 1995, Chmielweski, 1998), whereas protein content is rising (Savin et al., 1996). This is mainly due to shortening of the grain filling period, even though under certain extent high temperature is leading to higher cell division, assimilation and translocation rates (Wardlaw, 1992). Results showed a correlation of $r = -0.57$ between protein content and absolute yield and $r = -0.75$ between protein content and trend corrected yield. Correspondingly, there was a negative correlation ($r = -0.72$) between protein content and the graded proportion which is suitable for malting (= screening percentage).

Phenological data:

Data was found to be a valuable information source. Above all, the length of the period of grain filling is indispensable. Since phenological phases for spring barley were not reported for the period after 1990, this data had to be calculated by simple regression analysis, based on the data for oat. Correlations between the grain filling period and (trend corrected) yield, protein content and screening percentage resulted in r = 0.72, -0.60 and 0.54, respectively.

Meteorological data:

The impact of weather data to yield is considered to be the most important. The most significant coefficients were found for correlations of temperature (r = -0.76) and air humidity (r = 0.73) during the grain filling period. For the two quality parameters protein content and screening percentage, the humidity proved to be the most valuable indicator (r = -0.67 and r = -0.76, respectively). Precipitation was generally less correlated. Between 1974 and 1996 there were some extreme years from the meteorological point of view. By excluding these extreme values, considering them as outliers, results changed considerably. By taking into account these variations, temperature and air humidity proved to be useful for yield prediction, whereas precipitation and air humidity were found to be the most valuable parameters for protein content prediction. The prediction for the screening percentage was difficult, with the temperature being regarded the best indicator. For the precipitation sum from beginning of the year to yellow ripeness no significant correlation was detected.

Multiple linear regressions:

This analysis was performed over meteorological and phenological data, without including remote sensing data, which was not available for the full time period. Different strategies were used with different numbers of variables. The variables used were aggregated values for meteorological parameters as the daily average temperature (DT), precipitation (PrS), and the daily average air humidity according to the phenological phases described above (PS, EM, YS, MG, GF). Additionally, the crop water stress index (CWSI) for phase GF and the precipitation sum (PrS) of the time span from the beginning of the year to yellow ripeness was integrated. Evaluation of the obtained results was carried out as well as prognosis. The latter was performed by predicting yield or quality of a specified year, whose data was not included in the statistical data-set. Threshold setting variables were selected or excluded from stepwise forward selection of independent variables of the multiple regression model. Evaluation and prognosis were performed for spring barley as well as for oat as the phenologically monitored crop.

Three different strategies were applied:

Strategy 1: stepwise selection from all 22 factors as independent variables.

Strategy 2: stepwise selection from all factors recorded before ear emergence as independent variables.

Strategy 3: utilization of just 5 factors recorded in the grain filling period as independent variables.

Yield:

For spring barley as phenologically monitored crop (1974 – 1990) 8 from 22 variables remained within thresholds, 99 % of the original variance could be explained, whereby the average error of the modeled yield was 0.8 %, the maximum error 3.1 %. The prognosis explained 92 %, 2.5 % was the average error of yield and the maximum 8.9 %. Extending the convergence thresholds resulted in 14 independent variables and $r^2 = 0.99$. Restricting the independent variables to be all recorded before ear emergence (strategy 2), or to be recorded in the grain filling period (strategy 3), r^2 resulted as 0.49 and 0.64, respectively.

Worse results were obtained for oat as phenologically monitored crop. Probably the 22-year period is too long for reliable forecasts. Predictions were ameliorated by dividing the 22-year period into two separate 11-year periods. Following this approach, 99 % of the original variance could be explained (6 factors each). 96 % explanation was reached by prognosis with a maximum error of 5.1 %. By restriction to factors to be recorded before ear emergence (strategy 2) the maximum error between observed and predicted yield was 5.9 %. Restriction to factors recorded in the grain filling period (strategy 3) resulted in $r^2 = 0.28$.

Protein content:

Approaching strategy 1, the prognosis was executed with $r^2 = 0.91$ with a maximum error between observed and predicted protein content of 0.7 %. Strategy 2 and 3 led to lower explanation rates, i. e. $r^2 = 0.69$ and 0.44, respectively. Inclusion of predicted yield as an independent variable did not affect results, since it was excluded in the stepwise selection anyway.

For oat as a phenological monitoring crop, protein concentration was best predictable, dividing into two 11-year periods and without utilization of predicted yield as independent variable ($r^2 = 0.98$ for strategy 1). For strategy 2, modeling in 2 partitions, r^2 was 0.67 and for strategy 3 r^2 resulted in 0.2.

Screening percentage

A considerable improvement of the prognosis for screening percentages was achieved by extending the number of factors (independent variables). A rise from 6 to 12 or from 6 to 11 factors improved the resulting r^2 from 0.47 to 0.99 or from 0.12 to 0.86 for strategies 1 and 2, respectively. Maximum error was 4.3 % for strategy 1.

Oat as phenological monitoring crop led to the best results by dividing into two 11-year periods again, resulting in determination coefficients of 0.87 for strategy 2 and 0.97 for strategy 1. r^2 for strategy 3, considering only those factors recorded after ear emergence, was low (0.23).

4 CONCLUSION

High resolution images are a necessary input for classification purposes on a field level to calculate cropping acreage and localize single fields. This is of interest to calculate regional yields on a quantity basis. However, classification was not performed within the framework of this project, nevertheless this imagery allows a weak differentiation related to yield. The best results have been achieved by composing sums of consecutive NDVI values and comparing them to yield. Although different soil types and atmospheric corrections have been taken into account, the relationship was low (best $r^2 = 0.42$).

The importance of low resolution images for crop monitoring and yield and quality forecasts is described in the literature (Illera et. al., 1998). Even though the number of observations was not sufficient for a statistically significant statement, the use of NOAA-AVHRR data was proved to be applicable and convenient. The advantages of NOAA-AVHRR images are the high repetition rate (hence high availability), the large footprint and the low costs, whereas the low resolution, hence mixed pixel problem, is disadvantageous.

The necessity to integrate ancillary data (earth borne data) is mainly due to the fact that the part of the plants the study focuses on is the generative part, whereas remote sensing methods are mainly recording vegetative features.

Phenological data is not only necessary for comparison purposes, since vegetation progress is varying from year to year, but also for the identification of the most significant phases for yield and quality formation. Furthermore, with the knowledge of the grain filling period length, coarse predictions for yield can be derived.

The most important factor among meteorological data was found to be temperature. Temperature is mainly determining the length of the grain filling period, where the range between 14 – 18 °C was figured out to be optimal. Thus the highest simple correlation was found between spring barley yield and temperature (r = -0.76). Quality was detected to be more susceptible to water balance parameters, such as precipitation and air humidity as to temperature. As a result it can be ascertained that for an operational system only few parameters of ancillary data are valuable: temperature, air humidity and precipitation parameters to be the most important.

The selection and the number of input factors for multiple regression analysis are mainly responsible for a high quality forecast model. Moreover, too large intervals might result in fatal errors for prediction, it is therefore advisable to reduce the sampling years to 10 – 15.

With models based on multiple regression the best results at yellow ripeness stage were obtained in the order protein content, followed by screening percentage and yield. Protein content and screening percentage can better be predicted as yield in early stages, like as before ear emergence. Thus the explanation partition of the original variance for the former two parameters is estimated to be 65 % and 85 %, respectively.

As an additional input factor to the model described above, also spectral data (i. e. NDVI) could be integrated. Approaches have been described and results are promising. This can considerably reduce costs for data procurement and nonetheless lead to accurate results.

REFERENCES

Ebertseder, T., Tungalagsaikhan, P., Bittner, M., Meisner, R., Schroedter, M., Dech, S., 1999: Towards an operational atmospheric correction for AVHRR land surface products, in: Proc. IEEE 1999 Int. Geoscience and Remote Sensing Conf., Hamburg, 28 June – 2 July 1999.

Bethel, G.; 1998 (employee at USDA): personal communications.

Schelling K., 2000; Universität Bonn, Dissertation: Aufbau eines operationellen fernerkundungs- und GIS-gestützten Systems zur Zustandsbeschreibung, Ertrags- und Qualitätsabschätzung von Braugerstenbeständen, Shaker Verlag, Aachen.

Kühbauch W., Dockter K., Müller C., Sarx H. G., Schelling K., 1998: Abschlußbericht zum Forschungsvorhaben „Integration von MOMS-2P-Aufnahmen in ein Landwirtschaftliches Geographisches Informationssystem zur Zustandsbeschreibung, Ertrags- und Qualitätsabschätzung von Braugerstenbeständen“, Friedrich Weissheimer Malzfabrik und Lehrstuhl f. Allg. Pflanzenbau der Universität Bonn; Andernach.

Hanus H., Aimiller O., 1978: Ertragsvorhersage aus Witterungsdaten. Fortschritte im Acker- und Pflanzenbau, Heft 5. Beiheft zur Zeitschrift für Acker- und Pflanzenbau, Parey Verlag, Berlin und Hamburg, 1978.

Eagles, H. A., Bedggood, A. G., Panozzo, J. F., Martin, P. J., 1995: Cultivar and environmental effects on malting quality in barley, Aust. J. Agric. Res. 46 (5), 831 – 844.

Chmielewski, F.-M., 1998: Der Einfluß der Witterung auf die Ertragsbildung von Sommergerste, Mitt. Ges. Pflanzenbauwiss. 11, 97 – 98.

Savin, R., Stone P. J., Nicolas, M. E.:, 1996: Responses of grain growth and malting quality of barley to short periods of high temperature in field studies using portable chambers, Aust. J. Agric. Res. 47 (3), 465 – 477.

Wardlow, I. F., 1992: Temperature effects on source-sink relationships: A review, in: Adaptation of food crops to temperature and water stress, Proceedings of an international symposium, Kuo, C. G. (ed.), 148 – 159, Taiwan.

Illera P., Delgado, J. A., Manso Fernandez A., Fernandez Unzueta, A., 1998: Integration of Remote Sensing and meteorological data for vegetation monitoring in Spain, in: Proc. First Int. Conf. Geospation Inform. Agric. Forestry, Volume 2, Lake Buena Vista, 55 – 62.