Satellite based data mining to support Egypt's agriculture

Wilco Terink[1], Peter Droogers[1], Jos van Dam[2], Gijs Simons[3], Maurits Voogt[3], and Amor Ines[4]

[1] FutureWater, Wageningen, The Netherlands
w.terink@futurewater.nl
WWW home page: www.futurewater.nl
[2] Wageningen University, Wageningen, The Netherlands
[3] eLEAF, Wageningen, The Netherlands
[4] Columbia University, New York, United States

Abstract. Agriculture in developing countries might benefit from advanced satellite based data mining approaches. The objective of the current study was to evaluate the added value of high- above coarse-resolution satellite imagery in crop yield forecasting. This study focused on a coarse-resolution pixel in the Nile Delta in Egypt. Within this coarse-resolution pixel, 256 high-resolution (15 m) ASTER pixels are present, with wheat and berseem being the main crops. A crop-water model was used to simulate crop yields for the coarse-resolution pixel and for each of the high-resolution pixels. The model was driven by remotely sensed LAIs; one time-series for the coarse-resolution run, and 256 time-series for the high-resolution runs. The model was calibrated with SEBAL retrieved ET_a. It was concluded that with the use of coarse-resolution remote sensing, yields were overestimated between 9-26%, while high-resolution remote sensing resulted in errors below 3%.

Keywords: Crop yield forecasting, SWAP, SEBAL, remote sensing, ASTER, MODIS

1 Introduction

Crop growth models play a major role in sustaining the world-wide food security. In developing countries, like e.g. Egypt, such models are hardly used. These models are used to simulate crop growth during the growing season, given the farmers' management practices. Crop growth models provide the farmer with the option to simulate certain farm management measures (e.g. irrigation depth), in order to evaluate the effect of these measures on the final crop yield. If these crop growth models are sufficiently accurate, then a farmer is able to optimize his management practices to obtain a higher crop yield. At a more strategic level, these crop growth models can play an important role to decision makers to take timely decisions regarding food import and/or export strategies.

If the uncertainty in the spatial variation of soil properties, initial soil conditions, crop parameters, and meteorological forcing is small, then crop growth models are capable of simulating crop yields quite accurately [9]. When crop yield forecasting applications are applied over large areas that rely on a spatially distributed crop growth model, the uncertainty in the spatial variation of the input data increases [6].

Nowadays, in data mining remote sensing images are often used in crop growth models to improve the simulation of these processes [2], [8], [10], [11]. Data that can be derived from remote sensing are e.g. the Leaf-Area-Index (LAI) [5], crop yield and biomass [12]. Remote sensing images are available in numerous spatial resolutions, where coarse-resolution images often contain a mixed signal (e.g. different Leaf-Area-Indexes (LAIs)) of the crops present in the area of interest. Therefore, it is expected that high-resolution satellite images result in improved crop yield forecasts in areas where the distribution in crop types is heterogeneous. This situation is very likely if the focus is on small-scale farming, where many different crops are grown next to each other. This leads to an increased uncertainty in the model input parameters (soil properties, crop parameters, etc.). It is, however, unknown to what extent the use of high- above coarse-resolution satellite imagery is really significant in the crop yield forecast. Therefore, the objective of the current study is to evaluate the added value of high- above coarse-resolution satellite imagery in crop yield forecasting.

2 Methodology

2.1 Pixel selection

In order to evaluate the added value of high-resolution above coarse-resolution remote sensing images in crop yield forecasting, an area of 240 x 240 m in the Meet Yazid command area (Fig. 1) in the Egyptian Nile Delta has been selected for this study. This area is, together with the command areas Mahmoudia and Manaifa, home to 140,000 small-scale farmers, which makes the spatial crop distribution highly heterogeneous and thus increases the uncertainty in crop yield forecasting.

The current study uses NASA's ASTER remote sensing images for the high-resolution (15 m) images. In order to undertake the comparison based on difference in resolution only, it is important that i) the coarse-resolution pixel contains a mixture of major crops and some noise (built-up area and other crops), ii) the pixels are cloud-free, and iii) crop classification has been conducted in the area. A good option for a coarse-resolution image would be to use MODIS [1] imagery (250 m spatial resolution). However, for the current study a synthetic coarse-resolution image (240 m), referred to as MODIS_r (Fig. 1) hereafter, has been created by bilinear interpolation of the high-resolution ASTER images. MODIS_r contains 256 ASTER pixels (Fig. 1, left), with 131 wheat pixels, 86 berseem pixels, 31 built-up area pixels, and 8 other crop pixels. The current study focuses on the winter growing season (1 November 2011 – 14 May 2012), with berseem and wheat being the main crops.

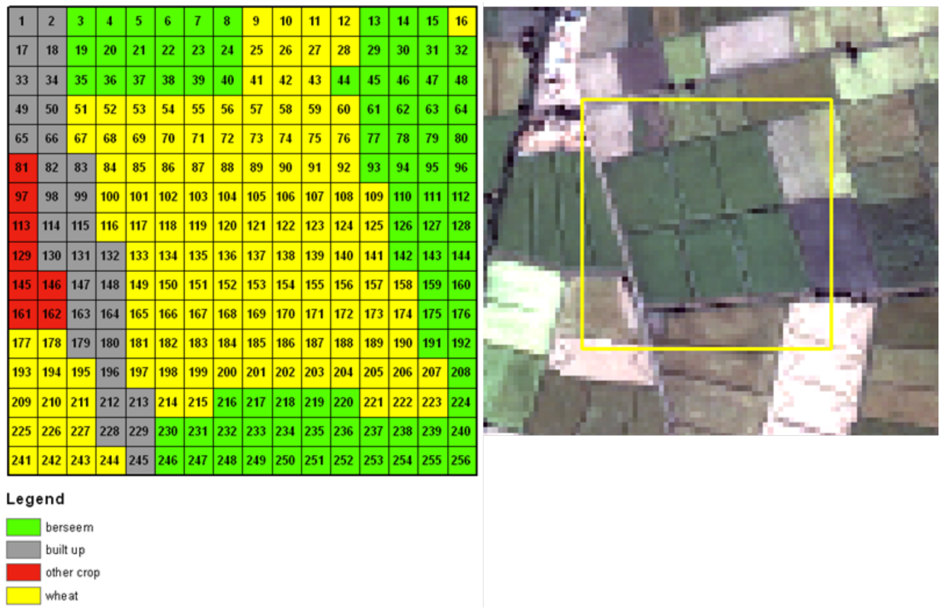

Fig. 1. The left plot shows the high-resolution ASTER pixels with wheat (yellow), berseem (green), build-up area (grey), and another crop (red). The right plot shows a high-resolution 5 m IKONOS image with the selected coarse-resolution image (yellow box) superimposed.

2.2 Remote sensing and SEBAL

The current study uses the Soil-Water-Atmosphere-Plant (SWAP) model [14] to simulate crop yields. SWAP was calibrated using bi-weekly actual evapotranspiration (ET_a) values retrieved from the SEBAL (Surface Energy Balance Algorithm for Land) [4] algorithm. SEBAL was also used to calculate the potential evapotranspiration (ET_p), which was used in combination with the reference evapotranspiration (ET_{ref}) to calculate the periodically crop factor (Kc). For detailed information regarding SEBAL, we refer to [4], [3]. The combination of remote sensing and SEBAL resulted for each of the high-resolution pixels and the coarse-resolution pixel in a bi-weekly time-series of ET_a, ET_p, LAI, and Kc.

2.3 SWAP modeling

Model schematization. SWAP [14] was selected as "the" modeling tool for the current study, because it simulates the soil moisture content in the soil profile, which is a major process in the determination of the actual transpiration and reduction of photosynthesis as a result of drought stress [6]. Since the scope of the current project is to evaluate the added value of high-resolution remote sensing (e.g. LAI, Kc) above coarse-resolution remote sensing, it is a huge advantage that the user can assign remotely sensed LAIs and Kc to SWAP, instead of simulating these variables like more complex crop growth models do (e.g. WOFOST [15]).

We have used the simple crop module in SWAP, where we have specified for each high-resolution pixel, being either berseem or wheat, as function of the development stage the LAI, Kc, rooting depth, irrigation depth, and the stress criterion (T_{stress}) which triggers irrigation applications. The same is done for the coarse-resolution pixel, being only one LAI and Kc time-series, representing a mixture of wheat, berseem, build-up area, and another crop. The yield response factor [7] relates the reduction in transpiration to the reduction in yield according to the FAO33 method (1).

$$1 - \frac{Y_a}{Y_p} = K_y \left(1 - \frac{T_a}{T_p}\right) \tag{1}$$

with Y_a the actual yield (kg/ha), Y_p the potential yield (kg/ha), K_y the yield response factor (-), T_p (cm) and T_a (cm) are the potential and actual transpiration, respectively. The yield response factors are 1.05 for wheat and 1.1 for berseem. SWAP calculates the actual (fresh) yield based on the relative yield and the potential yield. The potential yield for berseem is taken as 90,000 kg/ha and for wheat the potential yield is taken as 7,220 kg/ha. In following-up studies the more detailed WOFOST [15] schematization can be used.

SWAP calibration. SWAP needs to be calibrated in order to give reliable crop yield forecasts. Since the fraction of the actual evapotranspiration (ET_a) over the potential evapotranspiration (ET_p) indicates the relative crop yield, SWAP has been calibrated such that the simulated ET_a matches the measured ET_a from SEBAL, on a bi-weekly basis. For the current study, SWAP has been calibrated separately for a representative wheat pixel and a representative berseem pixel. The calibration of SWAP is done using the model-independent Parameter Estimation Software PEST[5]. SWAP was calibrated by optimizing the soil hydraulic parameters, initial soil moisture conditions, irrigation depths and frequencies, and rooting depths.
The calibration of SWAP for wheat was satisfactory (Fig. 2); the R^2 increased from 0.36 (uncalibrated) to 0.90 (calibrated). Also berseem was calibrated succesfully; the R^2 increased from 0.57 (uncalibrated) to 0.89 (calibrated).

2.4 Crop yield forecasting

To evaluate the added value of high-resolution above coarse-resolution remote sensing images in crop yield forecasting, it is required to have a reference data set which represents the real crop yield situation in the field. The reference data set was created by running SWAP for the entire growing season, using high-resolution ASTER imagery for the entire growing season. This run will be

[5] `http://www.pesthomepage.org/`

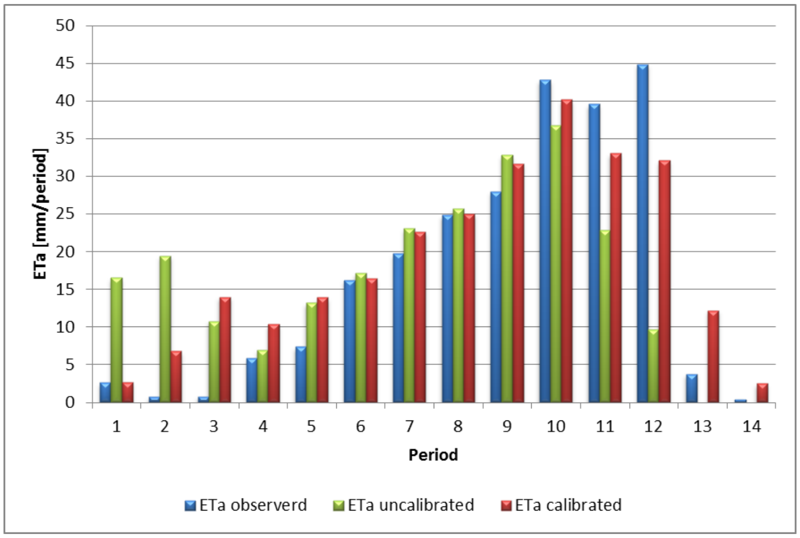

Fig. 2. Observed (SEBAL) and simulated ET_a for wheat. Results are shown for each period (bi-weekly) for the uncalibrated and calibrated model.

referred to as the reference run hereafter.The last four bi-weekly periods (two months) are used as the crop yield forecasting period. For the coarse-resolution crop yield simulation we have used the $MODIS_r$ LAIs and Kc for the first ten bi-weekly periods, whereas for the high-resolution crop yield simulation we have used the ASTER LAIs and Kc for the first ten bi-weekly periods. For the forecasting period, no remote sensing images are available. To simulate crop yields for the entire growing season, SWAP also requires LAI and Kc input for the forecasting period. To use representative LAIs and Kc during the forecasting period, "standard values" can be used for both the high- and coarse-resolution simulations. The current study uses the $MODIS_r$ imagery as "standard values" for the forecasting period for both the coarse- and high-resolution runs. In summary three simulations were compared:

1. high-resolution, no forecasting (reference);
2. high-resolution, 2 months forecast;
3. coarse-resolution, 2 months forecast;

3 Results

Table 1 shows the average wheat and berseem yields forecasts for the reference run, and the high- and coarse-resolution runs. Based on these results it is clear that the use of high-resolution remote sensing in crop yield forecasting gives a very small error when compared with the reference situation.
Fig. 3 shows the spatial variation in relative yields for each of the three runs: for i) the reference run, ii) the high-resolution run, and iii) the coarse-resolution run.

Table 1. Average wheat and berseem yields forecasts for the reference situation, and the high- and coarse-resolution runs. Also the errors with respect to the average reference situation are shown.

Run	Wheat yield [kg/ha]	Error[%]	Berseem yield [kg/ha]	Error [%]
Reference	6,390		67,107	
High-resolution	6,303	-1.4	68,544	2.1
Coarse-resolution	6,955	8.8	84,478	25.9

Also the errors with respect to the reference run are shown for both the high- and coarse-resolution run. The white area represents the area that is covered with build-up area or another crop and is left out of the analysis. It can be seen that the distribution in relative yields is not significantly different for the reference run and the high-resolution run. This is especially true in the central and southwestern part of the pixel where wheat is grown. Differences between the high-resolution run and the reference run are mainly present at the northern and eastern borders of the pixel where mainly berseem is grown. Within this area the relative yield for berseem is higher for the high-resolution run as for the reference run. A clear transition zone is present in the northern and eastern pixels where wheat borders berseem. In this zone the yield is underestimated with respect to the reference run. This holds for both the wheat and the berseem pixels. For the coarse-resolution run the relative yields are overestimated for both wheat and berseem, with the largest overestimations for berseem. The overestimated yield for berseem is mainly present in the northern and eastern part of the pixel, where the overestimation can reach 30%. For wheat the overestimation is less significant, and is almost zero in the central and southwestern part of the pixel.

4 Discussions and implications

4.1 Advantage for farmers

The current study has shown that for farmers, the use of high-resolution remote sensing has several advantages above the use of coarse-resolution remote sensing in crop yield forecasting. First of all, a farmer can obtain a much more accurate yield forecast if high-resolution satellite imagery is used. If coarse-resolution remote sensing would have been used, then for wheat the yield is overestimated with approximately 9%, and for berseem with 26%. If the farmer assumes these numbers for the forecast, then he will be very satisfied for the wrong reasons, because eventually his or her yields turn out to be considerably lower. Additionally, he or she will take the wrong measures: e.g. less irrigation, reduced fertilizer applications etc. If high-resolution satellite imagery will be used, then the forecast accuracy is much better; wheat underestimated with 1.4% and berseem overestimated with 2.1%.

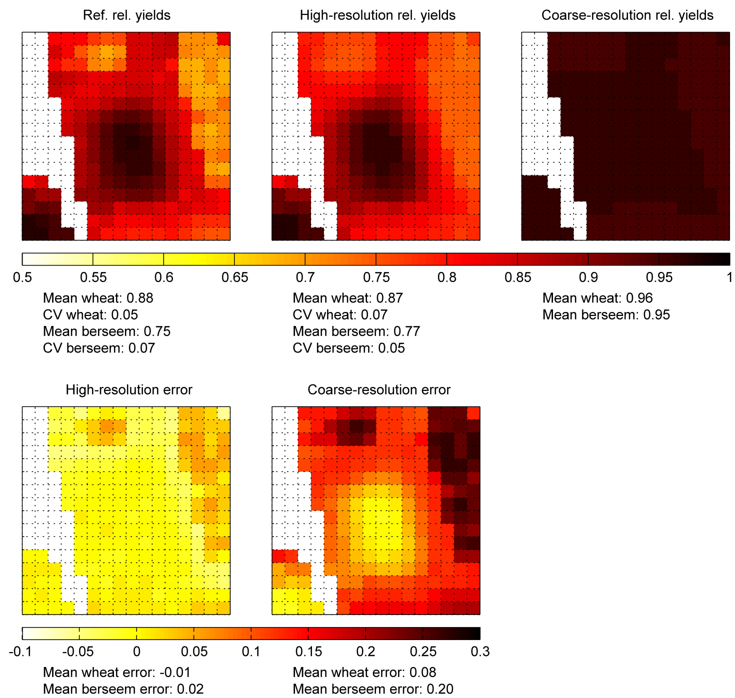

Fig. 3. Spatial variation in relative yields at the end of the growing season for the reference situation (top left), high-resolution run (top middle), and coarse-resolution run (top right). The bottom plots represent the errors of the high- and coarse-resolution runs with respect to the reference run.

4.2 Advantage for decision makers

At a more strategic level, decision makers will have considerable advantages if high-resolution remote sensing is used in crop yield forecasting; they can better take timely decisions regarding food import and/or export strategies. If yields are significantly overestimated in crop yield forecasting models, then decision makers might assume to have sufficient yields to consider export, but in reality they will have considerably lower yields. If yields are underestimated on the other hand, then decision makers will import the required amount of food to feed the country's population. Then it finally turns out that, because of the underestimation and the imported food, there is too much food purchased. In both cases money will be lost by either i) lower than expected food export or ii) too much food import.

Differences in production with respect to the reference situation are shown in Fig. 4. Based on this figure, it is clear that a significant amount of money can be saved if high-resolution remote sensing is used in crop yield forecasting. If coarse-resolution remote sensing is used in crop yield forecasting for wheat, then approximately 660 million US$ is lost through less export due to overestimated wheat yields. High-resolution remote sensing for wheat would lead to underestimated wheat yields, meaning that decision makers import too much wheat for a price of approximately 100 million US$. Losses are even more significant for berseem. Both coarse- and high-resolution remote sensing results in overestimated berseem yields. If coarse-resolution remote sensing will be used for berseem, then 3.2 billion US$ is lost through less exports, whereas high-resolution leads to a loss of approximately 0.3 billion US$. Costs of high-resolution remote sensing in crop yield forecasting are very modest and are calculated at about 0.05 US$/ha. This is negligible small if compared to the benefits; 68 US$/ha for wheat (250 million US$/36,832 km^2), and 869 US$/ha for berseem (3.2 billion US$/36,832 km^2).

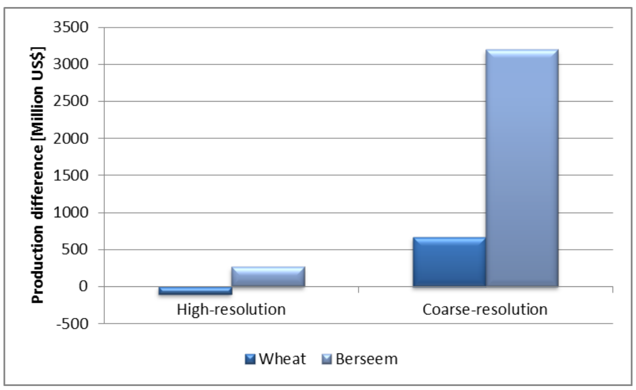

Fig. 4. Production difference with respect to the reference situation if high- vs. coarse-resolution remote sensing is used in crop yield forecasting.

4.3 Future outlook

The current study evaluated the added value of high-resolution above coarse-resolution remote sensing in crop yield forecasting by analyzing one pixel in the Meet Yazid command area in the Nile Delta in Egypt. The results of the current study provide some interesting perspectives for future activities:

1. A more in-depth study over a larger area and multiple years; the results of this study are based on one pixel in the Meet Yazid command area in the Nile Delta in Egypt. Since the land use fractions for this pixel are known, it would be interesting to conduct the study over a larger area, using many coarse-resolution MODIS images and high-resolution ASTER images. This will lead to increased model input uncertainties, since probably more different crops are to be found over a larger area.

2. Actual implementation to support farmers in Egypt.
3. Advisory services to decision makers and agro-industry and trade.
4. Support small-scale rainfed farmers in developing countries.

More detailed information regarding this study can be found in the technical report [13][6].

Acknowledgments. This document is produced by the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), which is a strategic partnership of CGIAR and Future Earth. The views expressed in this document cannot be taken to reflect the official opinions of CGIAR or Future Earth.

References

1. Baccini, A., Friedl, M.A., Woodcock, C.E., Warbington, R. Forest biomass estimation over regional scales using multisource data. Geophysical Research Letters. 31, 1–4, DOI:10.1029/2004GL019782 (2004)
2. Bastiaanssen, W.G.M., Allen, R.G., Droogers, P., DUrso, G., Steduto, P. Twenty-five years modeling irrigated and drained soils: State of the art. Agricultural Water Management. 92(3), 111–125, http://dx.doi.org/10.1016/j.agwat.2007.05.013 (2007)
3. Bastiaanssen, W.G.M., Molden, D.J., Makin, I.W. Remote sensing for irrigated agriculture: examples from research and possible applications. Agricultural Water Management. 46(2), 137–155. http://dx.doi.org/10.1016/S0378-3774(00)00080-9 (2000)
4. Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M. A remote sensing surface energy balance algorithm for land (SEBAL): Part 1 formulation. Journal of Hydrology. 212–213, 198–212, http://dx.doi.org/10.1016/S0022-1694(98)00253-4 (1998)
5. Boegh, E., Thorsen, M., Butts, M.B., Hansen, S., Christiansen, J.S., Abrahamsen, P., Hasager, C.B., Jensen, N.O., van der Keur, P., Refsgaard, J.C., Schelde, K., Soegaard, H., Thomsen, A. Incorporating remote sensing data in physically based distributed agro-hydrological modelling. Journal of Hydrology. 287, 279–299, http://dx.doi.org/10.1016/j.jhydrol.2003.10.018 (2004)
6. De Wit, A.J.W., van Diepen, C.A. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. Agricultural and Forest Meteorology. 146, 38–56, http://dx.doi.org/10.1016/j.agrformet.2007.05.004 (2007)
7. Doorenbos, J., Kassam, A.H. Yield response to water. FAO Irrigation and Drainage Paper no. 33. (1979)
8. Droogers, P., Immerzeel, W.W., Lorite, I. Estimating actual irrigation application by remotely sensed evapotranspiration observations. Agricultural Water Management. 97, 1351–1359, http://dx.doi.org/10.1016/j.agwat.2010.03.017 (2010)

[6] http://www.futurewater.nl/wp-content/uploads/2013/02/Report_CropYieldForecasting_highlow_scales_v6.pdf

9. Hansen, J.W., Jones, J.W. Scaling-up crop models for climate variability applications. Agricultural Systems. 65(1), 43–72, http://dx.doi.org/10.1016/S0308-521X(00)00025-1 (2000)
10. Immerzeel, W.W., Droogers, P. Calibration of a distributed hydrological model based on satellite evapotranspiration. Journal of Hydrology. 349(3-4), 411–424, http://dx.doi.org/10.1016/j.jhydrol.2007.11.017 (2008)
11. Ines, A.V.M., Honda, K., Gupta, A.D., Droogers, P., Clemente, R.S. Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture. Agricultural Water Management. 83(3), 221–232, http://dx.doi.org/10.1016/j.agwat.2005.12.006 (2006)
12. Serrano, L., Filella, I., Peñuelas, J. Remote Sensing of Biomass and Yield of Winter Wheat under Different Nitrogen Supplies. Crop Science. 40(3), 723-731, 10.2135/cropsci2000.403723x (2000)
13. Terink, W., Droogers, P., van Dam, J., Simons, G., Voogt, M. 2012. The added value of high-resolution above coarse-resolution remote sensing images in crop yield forecasting: a case study in the Egyptian Nile Delta. FutureWater report 116.
14. Van Dam, J.C. Field scale water flow and solute transport. SWAP model concepts, parameter estimation and case studies. PhD thesis, Wageningen Universiteit, 167 pp (2000)
15. Van Diepen, C.A., Wolf, J., van Keulen, H. and Rappoldt, C. WOFOST: a simulation model of crop production. Soil Use and Management. 5, 1624, doi:10.1111/j.1475-2743.1989.tb00755.x (1989)
16. Van Genuchten, M.Th. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal. 44, 892–898 (1980)