Papers in Natural Resources   Natural Resources, School of

2004

Forecasting Vegetation Greenness With Satellite and Climate Data

Lei Ji
Colorado State University

Albert J. Peters
University of Nebraska-Lincoln, apeters@calmit.unl.edu

Ji, Lei and Peters, Albert J., "Forecasting Vegetation Greenness With Satellite and Climate Data" (2004). *Papers in Natural Resources*. 290.
http://digitalcommons.unl.edu/natrespapers/290

Forecasting Vegetation Greenness With Satellite and Climate Data

Lei Ji and Albert J. Peters

Abstract—A new and unique vegetation greenness forecast (VGF) model was designed to predict future vegetation conditions to three months through the use of current and historical climate data and satellite imagery. The VGF model is implemented through a seasonality-adjusted autoregressive distributed-lag function, based on our finding that the normalized difference vegetation index is highly correlated with lagged precipitation and temperature. Accurate forecasts were obtained from the VGF model in Nebraska grassland and cropland. The regression R^2 values range from 0.97–0.80 for 2–12 week forecasts, with higher R^2 associated with a shorter prediction. An important application would be to produce real-time forecasts of greenness images.

Index Terms—Advanced Very High Resolution Radiometer (AVHRR), forecasting, normalized difference vegetation index (NDVI), regression model, vegetation greenness.

I. INTRODUCTION

SATELLITE sensors have played an important role in the investigation of vegetation condition at regional and continental scales. Images from the Advanced Very High Resolution Radiometer (AVHRR) onboard National Oceanic and Atmospheric Administration (NOAA) satellites have had a key role in this activity because of their frequent observation (daily), large area coverage (continental) and historical record (>20 years). Vegetation condition is typically indicated using the normalized difference vegetation index (NDVI) derived from the reflectance in the visible (VIS) and near-infrared (NIR) bands [1], i.e., $NDVI = (R_{NIR} - R_{VIS})/(R_{NIR} + R_{VIS})$, where R_{NIR} and R_{VIS} are reflectance of the two bands. Numerous studies have shown that NDVI is proportional to the fraction of absorbed photosynthetically active radiation, leaf area, vegetation fraction, and net primary productivity; therefore, the index has been successfully applied to measure and monitor vegetation greenness [2]–[6].

Some NDVI-based indexes such as the vegetation condition index [7], relative greenness [8], and the standardized vegetation index [9] were designed to indicate NDVI variation relative to normal. Currently, real-time NDVI images and their
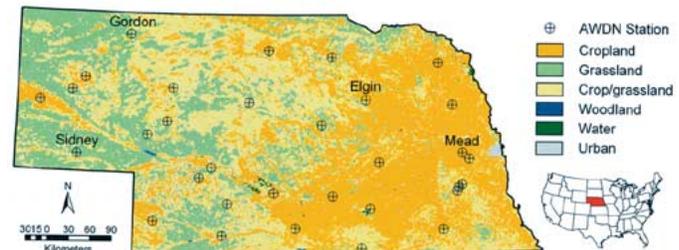

Fig. 1. Land cover and the AWDN stations in Nebraska. Four stations with a high percent area of cropland or grassland within the 10-km buffer were used for displaying detailed results: Elgin (cropland), Gordon (grassland), Mead (cropland), and Sidney (grassland).

derivatives are routinely produced and published on the World Wide Web [e.g., NOAA/NESDIS (http://orbit-net.nesdis.noaa.gov/crad/), U.S. Department of Agriculture Forest Service (http://www.fs.fed.us/land/wfas/), University of Kansas (http://mapster.kgs.ku.edu/kars/)], providing valuable information for academic research, natural resource management, and agribusiness.

Forecasts of vegetation status for the next few months would be of great value to decision makers and managers of agricultural and natural vegetation resources. No forecasting technique, however, has been presented to date because it is a much more difficult proposition than measuring current vegetation status. Forecasting requires a profound understanding of vegetation dynamics and vegetation interaction with the atmosphere, combined with proper data analysis techniques. Previous research has provided evidence that the NDVI is highly correlated to antecedent climate conditions [10]–[15]. According to previous research and our recent investigation of NDVI response to climate, we found that it is possible to forecast vegetation vigor through statistical manipulation of present and historical records of AVHRR NDVI, precipitation and temperature data. The objective of this research is to design a new and unique vegetation greenness forecast (VGF) model that would be capable of forecasting vegetation status for cropland and grassland in advance.

II. DATA AND METHODS

A. Study Area and Data

The study area is the state of Nebraska in the north/central United States Great Plains (Fig. 1). The major cover types in this area are crops (corn, soybeans, sorghum, winter wheat, and alfalfa) and grasses. Based on the 50-year average, annual precipitation ranges from 781 mm in the east to 432 mm in the

west, while average January/July air temperatures range from $-6.0/23.1\ °C$ in the north to $-4.8/24.9\ °C$ in the south.

We used the conterminous U.S. 1-km AVHRR NDVI dataset (1989–2000) produced at the Earth Resources Observation Systems (EROS) Data Center, U.S. Geological Survey. Each NDVI image is a biweekly maximum value composite, with radiometric calibration for sensor degradation and geometric registration [16]. The maximum value composite retains the highest NDVI value for each pixel during a 14-day period and produces spatially continuous cloud-free images over a large area [17]. The 1 km Conterminous U.S. Land Cover Characteristic Data Set [18] produced at the EROS Data Center was used to delineate land cover. Daily precipitation and maximum and minimum air temperatures (1988–2000) were obtained from the Automated Weather Data Network (AWDN) operated by the High Plains Regional Climate Center (http://hprcc.unl.edu/index.html) (Fig. 1).

B. Methods

We developed biweekly NDVI time series (1989–2000) for all AWDN weather stations in Nebraska. For a weather station, the NDVI of each biweekly period was obtained by averaging the NDVI of all grassland or nonirrigated cropland pixels within a 10 km buffer around each weather station. We used only the AWDN stations that have an >8-year climate record and have $>10\%$ area of grassland or cropland within the buffer. Because the biweekly periods of the NDVI composites do not have consistent calendar dates between years, a linear interpolation method was used to rescale them to comparable time periods. Total precipitation and average temperature over 14-day intervals, corresponding to the biweekly NDVI data, were derived from daily climate data for each AWND station. Thus, biweekly time series of NDVI for the growing season (middle April–early November), and precipitation and temperature from all the 31 AWDN stations were used for model development and validation. Because the statistical distribution of biweekly precipitation is skewed due to the presence of zero values and extreme outliers, the center weighted moving average technique was applied to transform precipitation into a normal distribution $P_t = (P_{t-2}^* + 2P_{t-1}^* + 3P_t^* + 2P_{t+1}^* + P_{t+2}^*)/9$, where P_t is transformed precipitation at time t; P_t^* is raw precipitation at t. Biweekly NDVI and temperature are normally or near normally distributed.

Relationships between the NDVI and lagged precipitation were determined with the Pearson correlation for each weather station. Because this relationship varies throughout the growing season, the correlation coefficients were calculated separately for five equal and contiguous intervals throughout the growing season: weeks 16–21 (April 9–May 20), 22–27 (May 21–July 1), 28–33 (July 2–August 12), 34–39 (August 13–September 23), and 40–45 (September 24–November 4). The same cross correlation was performed for the NDVI and lagged temperature at each weather station. Autocorrelation of the NDVI time series at various lags was identified with the Pearson correlation coefficient.

The autoregressive distributed-lag model [19], a special type of regression model, was designed to predict future NDVI, con-

ditioned on current and historical records of NDVI, precipitation, and temperature. The model is expressed as

$$V_{t+s} = \alpha + \sum_{i=0}^{l} \beta_i V_{t-i} + \sum_{j=0}^{m} \gamma_j P_{t-j} + \sum_{k=0}^{n} \delta_k T_{t-k} + \varepsilon_t \quad (1)$$

where t is concurrent time; V_{t+s} is forecast NDVI at s weeks ahead $(s = 1, 2, \ldots)$; V_{t-i} is NDVI measured at lag $i (i = 0, 1, \ldots, l)$; P_{t-j} is precipitation measured at lag $j (j = 0, 1, \ldots, m)$; T_{t-k} is temperature measured at lag k $(k = 0, 1, \ldots, n)$; l, m, and n are the lag lengths for V_{t-i}, V_{t-j}, and V_{t-k}; α, β_i, γ_j, and δ_k are the regression coefficients; and ε_t is random error. The model is based upon two assumptions: 1) NDVI is linearly related to precipitation and temperature; and 2) there is no interaction between precipitation and temperature.

Because the impacts of precipitation and temperature on NDVI are very different between the seasonal periods, an adjustment for the seasonal effect is important to make the model suitable for all time periods within the growing season. To accomplish this, seasonal dummy variables [19] were added to the model (1). These dummy variables are a set of five levels assigned to each corresponding seasonal period (weeks 16–21, 22–27, 28–33, 34–38, and 39–45). When dummy variables and their interaction items with V_{t-i}, P_{t-j}, and T_{t-k} are integrated into the autoregressive distributed-lag model, the model accounts for the difference in intercept and slope between the regression models of the different seasonal periods. With the seasonal dummy variable added to the model, the linearity assumption still holds within each of the five seasonal periods.

We used the data from 2000 to validate the VGF model. Specifically, we compared predicted and observed NDVI during the growing season at selected weather stations. The correlation coefficient and the mean absolute percentage error (MAPE) were applied to measure the goodness-of-fit of the model, and demonstrate the accuracy of the forecast. The MAPE is defined as $\text{MAPE} = (1/n)\Sigma 100|V - V^p|/V$, where V and V^p are observed and predicted NDVI, n is number of observations [20].

III. RESULTS AND DISCUSSION

A. Relationships of NDVI to Lagged Precipitation, Temperature, and NDVI

The relationships between NDVI and lagged precipitation, temperature and NDVI, as described below, provide the foundation of the VGF model.

1) Current Greenness Is Affected by Antecedent Precipitation of the Past Few Months: The NDVI-precipitation relationship at a weather station is demonstrated with a plot of correlation coefficients versus lag [Fig. 2(a)]. It is noted that correlations change with lag and are positive at lags ≤ 14 weeks in most cases; higher correlations tend to occur between 2- and 12-week lags. Also, the correlation-lag pattern, especially the peak-lag (lag with the highest correlation), varies depending on the time of the growing season. Previous research indicates that the longest peak lag is in the middle of the growing season
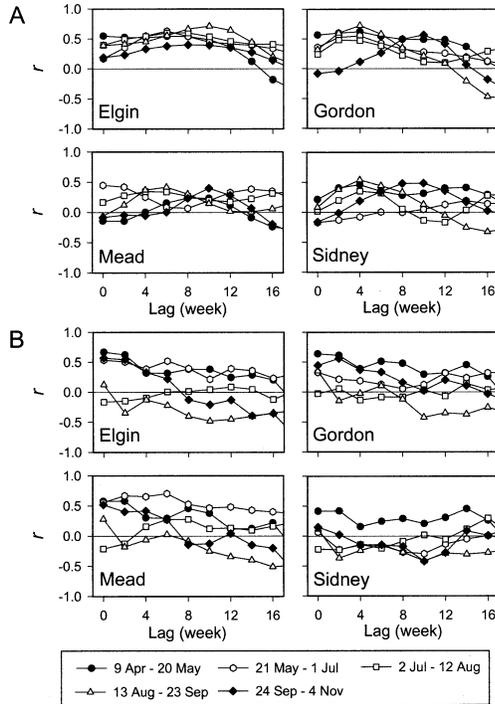
Fig. 2. (A) Pearson correlation coefficients (r) of NDVI versus lagged precipitation and (B) NDVI versus lagged temperature at four weather stations.

in the Nebraska Sandhills grassland [11]. Generally, higher correlations occur in the middle to late growing season, as seen at the Elgin and Mead Stations [Fig. 2(a)]. Agronomic studies suggest that the "moisture-sensitive period" of plant growth occurs at flowering through early fruiting stages [21], [22]. We also noticed a difference in the correlation-lag pattern between geographic locations. Western and southern Nebraska contain cool-season grasses and crops like winter wheat (planted in September and harvested in June) and have two growing season peaks resulting from a dual-cropping pattern. Seasonal changes in the correlation-lag patterns are more complicated in these areas due to the bimodal green peak.

2) Current Greenness Is Affected by Antecedent Temperature of the Past Few Months: The optimum temperature range for a plant varies widely and is species specific [23]. Plant adaptation and resistance to changes in the thermal environment causes a delayed response. Correlation-lag plots indicate the relationship of NDVI versus temperature at various lags [Fig. 2(b)] This correlation-lag pattern also differs among the five time periods during the growing season. For example, correlations are mostly positive at lags of 0–14 weeks in the spring and early summer (April 9–July 1) when plant growth is promoted by higher temperature at all lags. They are negative or near zero from early summer to early fall (July 2–September 23), implying that hot summer temperatures inhibit plant activity. In the fall (September 24–November 5), correlations are positive at shorter time lags (<8 weeks), indicating that recent higher temperature is beneficial to growth. Seasonal effects on NDVI–temperature relationships also depend on geographic location and are related to the optimal range of individual species. For example, at Sidney (located in the drier west),

most correlations are negative after May 21, indicating the impact of high temperature, evapotranspiration, and resulting low water availability.

3) Current Greenness Is Highly Affected by Antecedent Greenness of the Past Few Months: Changes in vegetation vigor have a low-frequency pattern as compared with atmospheric phenomena. This was confirmed by autocorrelation analysis of the NDVI time series. At most weather stations, positive autocorrelation was detected at lag times of up to 12 weeks, but decreased with increasing lag length. Correlation coefficients are usually greater than 0.8 at lags of 1–5 weeks and then decrease to 0.5–0.8 at lags of 6–8 weeks.

B. Performance of the VGF Model

In the autoregressive distributed-lag model (1), we chose lag length $l = 12$, $m = 16$, and $n = 16$ weeks for V_{t-i}, P_{t-j}, and T_{t-k}, respectively, based on our understanding of correlation-lag relationships illustrated in Section III-A and Fig. 2. That is, V_{t-i}, P_{t-j}, and T_{t-k} do not further contribute to NDVI variation when lags exceed these values. The different patterns of NDVI response to precipitation and temperature, among the seasonal periods, were accounted for by incorporating dummy variables into the model. The VGF model was applied to forecast NDVI for 2–12 weeks at all AWDN weather stations in Nebraska, using all NDVI (1989–2000) and climate (1988–2000) datasets. The average R^2 and MAPE of the regression models for the 31 weather stations are as follows:

two-week forecast $R^2 = 0.966$ MAPE $= 5.0\%$;
four-week forecast $R^2 = 0.931$ MAPE $= 7.4\%$;
eight-week forecast $R^2 = 0.854$ MAPE $= 11.1\%$;
twelve-week forecast $R^2 = 0.798$ MAPE $= 12.9\%$.

The R^2 states the proportion of NDVI variation explained by past NDVI, precipitation, and temperature, including the seasonal effects. MAPE indicates the prediction error relative to the true observation. At all weather stations, prediction accuracy is very high for the 2–4 week forecast and then gradually decreases with extended forecasts.

To validate the VGF model, predicted NDVI was compared to observed NDVI during the 2000 growing season (Fig. 3). When NDVI for a week during the 2000-growing season was predicted, the observed NDVI and climate variables for that week and all following weeks were excluded from the dataset. A very close fit between predicted versus observed NDVI was indicated by the correlation coefficients and MAPE. In fact, NDVI predicted for 2–12 weeks even mimics subtle changes in observed NDVI. With longer forecast periods, prediction accuracy decreases, and the 95% confidence interval gets wider. Validation of the VGF model was achieved by conducting a forecast of a deseasonalized NDVI anomaly. The NDVI anomaly is defined as the difference between NDVI of the current week and the average NDVI of the historical record for that same week. A positive anomaly indicates vegetation condition better than normal, while a negative value manifests relatively poor greenness conditions. Results show that the 2- to 12-week forecast of the NDVI anomaly also has satisfactory accuracy (Fig. 4).
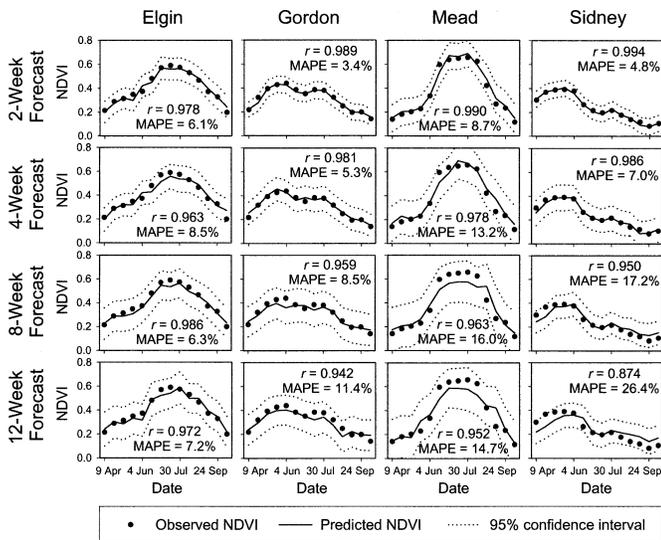
Fig. 3. Comparison of 2-, 4-, 8-, and 12-week forecasted NDVI and observed NDVI for the 2000 growing season. r is the correlation coefficient of forecasted and observed NDVI. The calendar date on the x axis indicates the first date of each biweekly period.
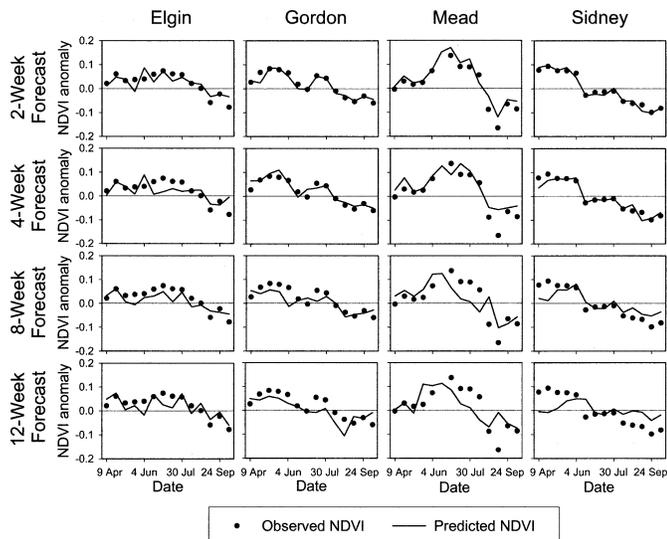

Fig. 4. Comparison of 2-, 4-, 8-, and 12-week forecasted anomaly and observed anomaly of NDVI for the 2000 growing season. The calendar date on the x axis indicates the first date of the biweekly period.

IV. CONCLUSION

This letter demonstrates that current vegetation greenness is strongly affected by antecedent precipitation, temperature, and vegetation greenness of the past few months. Based on this knowledge, we designed the VGF model constituted by an autoregressive distributed-lag function, with seasonal adjustment. The VGF model can reliably forecast short-term greenness up to 12 weeks using integrated historical climate and satellite data. The model can be used for production of real-time VGF maps over large geographic areas through the integration of pixel-based climate parameters and NDVI data. In a two-dimensional space, the predicted NDVI for each pixel location can be calculated using climatic parameters and NDVI time series for that location. Potential usefulness of the VGF model includes assessment of crop status, drought monitoring, and wild fire warning systems.

REFERENCES

[1] P. J. Sellers, "Canopy reflectance, photosynthesis and transpiration," *Int. J. Remote Sens.*, vol. 6, no. 8, pp. 1335–1372, 1985.

[2] C. J. Tucker, C. L. Vanpraet, M. J. Sharman, and G. Van Ittersum, "Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel: 1980–1984," *Remote Sens. Environ.*, vol. 17, pp. 233–249, 1985.

[3] J. A. Gamon, C. B. Field, M. L. Goulden, K. L. Griffin, A. E. Hartley, G. Joel, J. Peñuelas, and R. Valentini, "Relationships between NDVI, canopy structure, and photosynthetic activity in three Californian vegetation types," *Ecol. Appl.*, vol. 5, pp. 28–41, 1995.

[4] T. N. Carlson and D. A. J. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index," *Remote Sens. Environ.*, vol. 62, pp. 241–252, 1997.

[5] M. S. Rasmussen, "Developing simple, operational, consistent NDVI-vegetation models by applying environmental and climatic information: Part I. Assessment of net primary production," *Int. J. Remote Sens.*, vol. 19, no. 1, pp. 97–117, 1998.

[6] R. B. Myneni, F. G. Hall, P. J. Sellers, and A. L. Marshak, "The interpretation of spectral vegetation indexes," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 481–496, Mar. 1995.

[7] F. N. Kogan, "Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data," *Bull. Amer. Meteorol. Soc.*, vol. 76, no. 5, pp. 655–668, 1995.

[8] R. E. Burgan, R. A. Hartford, and J. C. Eidenshink, "Using NDVI to assess departure from average greenness and its relation to fire business," USDA, Forest Service, Washington, DC, General Tech. Rep. INT-GTR-333, 1996.

[9] A. J. Peters, E. A. Walter-Shea, L. Ji, A. Viña, M. Hayes, and M. D. Svoboda, "Drought monitoring with NDVI-based standardized vegetation index," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 1, pp. 71–75, 2002.

[10] S. E. Nicholson, M. L. Davenport, and A. R. Malo, "A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR," *Clim. Change*, vol. 17, pp. 209–241, 1990.

[11] L. Di, D. C. Rundquist, and L. Han, "Modeling relationships between NDVI and precipitation during vegetation growth cycles," *Int. J. Remote Sens.*, vol. 15, no. 10, pp. 2121–2136, 1994.

[12] Y. Richard and I. Poccard, "A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in South Africa," *Int. J. Remote Sens.*, vol. 19, no. 15, pp. 2907–2920, 1998.

[13] L. Ji and A. J. Peters, "Assessing vegetation response to drought in the Northern Great Plains using vegetation and drought indices," *Remote Sens. Environ.*, vol. 87, pp. 85–98, 2003.

[14] J. Wang, P. M. Rich, and K. P. Price, "Temporal responses of NDVI to precipitation and temperature in the Central Great Plains, USA," *Int. J. Remote Sens.*, vol. 24, no. 11, pp. 2345–2364, 2003.

[15] A. Lotsch, M. A. Friedl, B. T. Anderson, and C. J. Tucker, "Coupled vegetation-precipitation variation observed from satellite and climate records," *Geophys. Res. Lett.*, vol. 30, no. 14, pp. 1774–1777, 2003.

[16] J. C. Eidenshink, "The 1990 conterminous U.S. AVHRR data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 6, pp. 809–813, 1992.

[17] B. N. Holben, "Characteristics of maximum-value composite images from temporal AVHRR data," *Int. J. Remote Sens.*, vol. 7, pp. 1417–1434, 1986.

[18] T. R. Loveland, J. W. Merchant, D. O. Ohlen, and J. F. Brown, "Development of a land-cover characteristics database for the conterminous United States," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 11, pp. 1453–1463, 1991.

[19] D. N. Gujarati, *Basic Econometrics*, 2nd ed. New York: McGraw-Hill, 1988, pp. 431–466.

[20] R. Ramanathan, *Introductory Econometrics With Applications*, 3rd ed. Fort Worth, TX: Harcourt Brace, 1995, pp. 118–121.

[21] P. J. Salter and J. E. Goode, *Crop Responses to Water at Different Stages of Growth*. Farnham Royal, U.K.: Commonwealth Agricultural Bureaux, 1967, pp. 3–60.

[22] C. Y. Sullivan and J. D. Eastin, "Plant physiological responses to water stress," *Agr. Meteorol.*, vol. 14, pp. 113–127, 1974.

[23] A. H. Fitter and R. K. M. Hay, *Environmental Physiology of Plants*, 2nd ed. New York: Academic, 1987, pp. 187–224.