Daily NDVI relationship to cloud cover

Qiuhong Tang[*]

Institute of Industrial Science, University of Tokyo,

Tokyo 153-8505, Japan

Taikan Oki[†]

Institute of Industrial Science, University of Tokyo,

Tokyo 153-8505, Japan

June 15, 2006

---

[*]Tel: +81-(0)3-5452-6381 Fax: +81-(0)3-5452-6383 Email: tangqh@iis.u-tokyo.ac.jp
[†]Tel: +81-(0)3-5452-6381 Fax: +81-(0)3-5452-6383 Email: taikan@iis.u-tokyo.ac.jp

Abstract

An NDVI Cloud Index (NCI) was derived from Pathfinder AVHRR daily NDVI data and compared to observed cloud amounts and a sunshine duration cloud index (SCI) over an area of diverse land cover. Ground observations from 120 meteorological stations were significantly related to the daily NCI and the SCI, with R^2 values of 0.41 and 0.50, respectively. The daily NCI and interpolated cloud indices derived from ground observations over the 776,900 km^2 study area were compared. The correlation coefficient between the NCI and the observed cloud amount was less than 0.6 for less than 20% of the area. The correlation coefficient between the NCI and the observed sunshine duration index was less than 0.6 for less than 10% of the area and less than 0.7 for 41% of the area. There were strong correlations for high elevations in summer, while correlations for low elevations in winter were weaker. A frozen soil surface or snow cover degrades the NDVI relationship to clouds. The NCI and observed cloud indices had high correlation coefficients in areas with diverse land uses, suggesting that the NCI may be useful in estimating cloudiness over a large region.

1

1. Introduction

The normalized difference vegetation index (NDVI) is a non-linear transformation of the visible (red) and near-infrared bands of remotely sensed imagery. The NDVI is defined as the ratio of (NIR - Red) to (NIR + Red), where NIR is the spectral response in the near-infrared band, and Red is the spectral response in the red band (Tarpley et al. 1984). The NDVI obtained using visible and near infrared data from channels 1 (0.58-0.681 μm) and 2 (0.725-1.11 μm) (i.e., [CH 2 - CH 1]/[CH 2 + CH 1]) of the Advanced Very High Resolution Radiometer (AVHRR) is commonly used to monitor vegetation (Tucker 1979; Jackson et al. 1983; Justice et al. 1985; Tucker et al. 1991). As an index of relative seasonal changes in vegetation rather than the vegetation amount, the NDVI cannot be used directly as a parameter in numerical models. Various studies have correlated the NDVI with physical properties of the vegetation canopy, such as the leaf area index (LAI), fractional vegetation cover, vegetation condition, and biomass (Wiegand et al. 1979; Zhang and Williams 1997; Carlson and Ripley 1997). Previous studies have also related the NDVI to components of the water balance equation, such as soil moisture, precipitation, and evaporation (Choudhury and Golus 1988; Grist et al. 1997; Szilagyi et al. 2000). However, clouds can block satellite observations, and numerous studies have explored methods to yield a "cloud free" NDVI, as would be measured at the ground surface. Maximum-value composite images from temporal satellite data have been used to remove errors caused by clouds (Holben 1986). This compositing process provides the most cloud-free image possible.

Although clouds can be considered obstacles to satellite-derived observations, cloud distortions in satellite data can also allow for cloud estimates. The NOAA National Environmental Satellite, Data and Information Service (NESDIS) developed an experimental satellite-derived cloud dataset to provide cloud parameterization schemes in real time (Stowe et al. 1999). This "cloud from

AVHRR" (CLAVR) dataset is retrieved by a sequential, multispectral, decision-tree threshold algorithm that uses information from NOAA satellites (Stowe 1991; Stowe et al. 1991). Preliminary analysis has compared the CLAVR cloud product to other satellite-derived cloud analyses, such as data from the U.S. Air Force Real-Time Nephanalysis (RTNEPH) and the International Satellite Cloud Climatology Project (ISCCP) (Hou et al. 1993). However, no direct comparisons between CLAVR and a ground-observed cloud index have been made over a large region.

This study used daily NDVI images derived from NOAA AVHRR data. Monthly composite NDVI values were compared to daily NDVI data to create a daily NDVI Cloud Index (NCI). Daily NCI values were related to observed cloud amounts and sunshine duration indices from 120 meteorological stations over a 6-year period. The NCI and CLAVR cloud values were then compared to the observed cloud amounts and sunshine index over a 776,900 km^2 study area with diverse land cover. Relationships between the NCI and CLAVR and ground-observed cloud cover are presented and analyzed.

2. Methodology

Daily time series of NDVI between 1995 and 2000 over the Yellow River Basin in China (Figure 1) were obtained from the NOAA/NASA Pathfinder AVHRR Land Database (PAL; 8 km resolution; available at http://daac.gsfc.nasa.gov/; (Goward et al. 1991)). The study area contains three main landforms: the Qinghai-Tibet Plateau, the Loess Plateau, and an alluvial plain (Yang et al. 2004). Daily NDVI values, which had been corrected for Rayleigh scattering and ozone absorption but not for atmospheric water vapor (James and Kalluri 1994), were the basis of the NDVI dataset. Although the PAL dataset is not sufficiently stable to represent cloud variations within one day, it can represent daily variations in cloud status. More advanced imagers are now in operation, but

the AVHRR remains an important source of remotely sensed data. Only the AVHRR offers a data record longer than 20 years, which is critical for decadal climate studies. Daily CLAVR flag data from the PAL dataset were also used in this study. Land use data were obtained from the U.S. Geological Survey Land Use dataset based on 1-km AVHRR data between April 1992 and March 1993 (Anderson et al. 1976; Loveland et al. 1999). Elevation data were derived from HYDRO1k, a geographic database developed at the U.S. Geological Surveys (USGS) Earth Resources Observation and Science (EROS) Data Center that provides comprehensive and consistent global coverage of topographically derived datasets (HYDRO1k Team 2003).

[Figure 1 about here.]

Numerous methods allow the production of a cloud-free NDVI (Holben 1986; Verhoef et al. 1996; Roerink et al. 2000). The maximum composite method used in this study simplifies calculations. Daily NDVI values are composited at each grid point in the study area based on comparison of NDVI values on consecutive days in each month. The pixel with the highest NDVI value for the month is chosen as the date for inclusion in the composite. NDVI values for each day in the study period are then interpolated from the composite monthly data using a cubic spline method. The estimated daily NDVI is considered a "cloud free" NDVI.

The ratio of the NDVI value directly derived from instantaneous satellite observations to the "cloud free" NDVI value represents the cloud influence. For clear sky, the NDVI derived directly from satellite observations matches the "cloud free" NDVI. For an overcast sky, the NDVI directly derived from satellite observations is the same as the minimum NDVI for the entire study period. A partially cloudy sky is indicated when the NDVI directly derived from satellite observations falls between the "cloud free" NDVI value and the minimum NDVI. The NCI ranges from 0 for overcast to 1 for clear and is defined as (Figure 2)

4

$$NCI = (NDVI_r - NDVI_{min})/(NDVI_c - NDVI_{min}) \tag{1}$$

[Figure 2 about here.]

where $NDVI_r$ is the NDVI value directly derived from instantaneous satellite observations; $NDVI_c$ is the "cloud free" NDVI value, and $NDVI_{min}$ is the minimum value of NDVI. Note that the NCI is bounded so that it is less than 1.0.

The China Meteorological Administration (CMA) provided daily observed cloud amount and actual duration of sunshine data from 120 meteorological stations within or close to the study area (Figure 1). The cloud amount data was in tenths unit system (Sun et al. 2001). Solar radiation, R_s, was calculated with the Angstrom formula that relates solar radiation to extraterrestrial radiation and relative sunshine duration (Allen et al. 1998):

$$R_s = (a_s + b_s \times n/N)R_a \tag{2}$$

where n is the actual duration of sunshine (hour), N is the maximum possible duration of sunshine or daylight hours (hour), a_s is a regression constant that is the fraction of extraterrestrial radiation reaching the Earth on overcast days, $(a_s + b_s)$ is the fraction of extraterrestrial radiation reaching the Earth on clear days, and R_a is extraterrestrial radiation. The ratio of actual sunshine duration to daylight hours n/N describes the influence of clouds on solar radiation and is defined as the Sunshine-Cloud Index (SCI).

Daily NCI and CLAVR values were related to ground observations of cloud amount and the SCI from the 120 meteorological stations. A linear least-squares best fit method was used, and the overall quality of the fit was then parameterized in terms of the correlation coefficient R defined

5

by

$$R = \frac{N \sum x_s x_g - \sum x_s \sum x_g}{\sqrt{[N \sum x_s^2 - (\sum x_s)^2][N \sum x_g^2 - (\sum x_g)^2]}} \tag{3}$$

where N is the total number of time series being compared, x_s represents the satellite-derived cloud indices (i.e., NCI or CLAVR values), and x_g represents ground-observed cloud indices (i.e., the observed cloud amount or SCI). The R^2 value is the square of the correlation coefficient. The linear least-squares fitting method requires that ground-observed cloud indices be predicted by satellite-derived cloud indices as follows:

$$x_g' = a + b x_s \tag{4}$$

where x_g' is the cloud index predicted by the satellite data, $b = (N \sum x_s x_g - \sum x_s \sum x_g)/[N \sum x_s^2 - (\sum x_s)^2]$ and $a = (\sum x_g - b \sum x_s)/N$. The root-mean-square error (RMSE) and mean absolute error (MAE) between the predicted cloud index x_g' and the ground-observed cloud indices are given as

$$RMSE = \sqrt{\frac{1}{N} \sum \left(x_g' - x_g \right)^2} \tag{5}$$

and

$$MAE = \frac{1}{N} \sum \left| x_g' - x_g \right| \tag{6}$$

The NCI values can be estimated over the entire study area, but ground stations provide only point observations. Surface climate data from station observations are thus usually interpolated to gridded data (New et al. 1999). Similarly, the observed cloud amount and SCI values for the stations were interpolated to a grid over the study area using a thin-plate spline algorithm. Daily NCI values over the study area were then compared to the gridded cloud amount and SCI data for

the same period (1995 to 2000). The relationship as a function of land cover was analyzed using U.S. Geological Survey Land Use data. HYDRO1k digital elevation model (DEM) data were used to derive variation with elevation.

Comparisons for each month yielded seasonal variations in the relationship. Daily NCI values of one month were selected and related to the cloud amount and SCI from 1995-2000. The monthly variation of the relationship was demonstrated.

3. Results and Conclusions

Figure 3 shows time series of NCI, SCI, CLAVR, and ground-observed cloud amounts. Values of NCI, SCI, and CLAVR are scaled from [0, 1], while observed cloud amount value is from [1, 0]. Figure 3a compares July and August, and Fig. 3b compares February and March. Both CLAVR and the NCI capture the daily cloud variance in summer. The NCI cloud estimates do not agree with the ground-observed cloud amount or the SCI in winter. In winter, the cloud amount estimated by CLAVR poorly matches the ground observations.

[Figure 3 about here.]

Figure 4 shows the relationships of the cloud amount, NCI, SCI, and CLAVR at the observation stations. Valid data observed during the study period were cataloged into ten categories following the ground observed cloud amount in tenths, and the distributions are presented. The linear fit line for the data points is shown with the squared correlation coefficient. Figure 4a shows NCI values and SCI ($R^2 = 0.506$). Figures 4b, 4c, and 4d show the relationship between ground-observed cloud amount and the CLAVR, NCI, and SCI values, respectively. The CLAVR values and cloud amount have the poorest correlation ($R^2 = 0.169$). CLAVR values range mainly from 1-15, indicating

cloudy and mixed cloudy are predominant classifications from the CLAVR algorithm. The NCI and SCI both had higher correlations with ground-observed cloud amount with R^2 values of 0.407 and 0.572, respectively. The NCI results are more consistent with the ground-observed cloud index. The direct relationship between CLAVR values and the ground-observed cloud amount is not as robust.

[Figure 4 about here.]

Figure 5 shows histograms of cloud distributions at the ground-observation stations during the study period. More than 20% of the points have cloud fractions between 0.55 and 0.65 for the CLAVR cloud estimates because "mixed cloudy" is the predominant classification from the CLAVR algorithm. Less than 1% of the points have cloud fractions ranging from 0.95-1.0 for the NCI clouds because $NDVI_{min}$ was derived from the minimum NDVI value for the entire study period. For the SCI clouds, more than 20% of the points have cloud fractions between 0.05 and 0.25. The ground-observed cloud amount is relatively uniformly distributed in each cloud category, with higher occurrences in the nearly clear and cloudy categories.

[Figure 5 about here.]

Table 1 shows the R^2 value for satellite remotely sensed data and ground observations. The R^2 values for the NCI and cloud amount range from 0.38 to 0.47, with an average value of 0.41 for the 120 stations representing diverse land covers. The averaged standard deviation (SD) of R^2 is 0.07. For the NCI and SCI, R^2 values range from 0.45 to 0.58, with an average value of 0.50. The averaged standard deviation of the R^2 values is 0.07. Results show that the correlation coefficients for NCI and the observed cloud amount or SCI have high values across diverse land cover types. The R^2 associated with CLAVR values and the observed cloud amount or SCI are 0.16 and 0.22,

8

respectively. Correlation coefficients for CLAVR values and the observed cloud amount or SCI show lower values than the correlation coefficients between the NCI values and ground-observed indices.

[Table 1 about here.]

The RMSE associated with the NCI and observed cloud amount ranges from 0.21 to 0.30, with an average value of 0.25. The MAE ranges from 0.17 to 0.25 with an average value of 0.20. Figure 6a shows the R^2 value for the daily NCI values and observed cloud amount in each pixel of the study area. Figure 7a shows the cumulative distribution function of R^2 values in the study area. For less than 20% of the pixels, R^2 is less than 0.36 (i.e., the correlation coefficient equals 0.6). In the study area, 87% of the pixels have R^2 values of less than 0.49 (i.e., the correlation coefficient equals 0.7). The RMSE associated with NCI and SCI ranges from 0.16 to 0.26 with an average value of 0.20, and the MAE ranges from 0.12 to 0.22, with an average value of 0.16. Figure 6b shows the correlation coefficients between daily NCI values and observed SCI for each pixel. Figure 7b shows the distribution of R^2 values in the study area. The R^2 values are less than 0.36 (i.e., the correlation coefficient equals 0.6) at fewer than 10% of the pixels. Pixels where the R^2 value is less than 0.49 (i.e., the correlation coefficient equals 0.7) cover 41% of the study area. For CLAVR data, the RMSE associated with the cloud amount ranges from 0.26 to 0.34, with an average value of 0.30, and the averaged MAE is 0.25. The RMSE associated with the SCI ranges from 0.21 to 0.30, with an average of 0.25 and average MAE of 0.20. The error in CLAVR estimates is larger than that for the NCI. Figures 6c and 6d show the R^2 associated with daily CLAVR values and ground-observed cloud indices. The R^2 values in most of grid boxes are less than those associated with daily NCI values. Pixels where R^2 is less than 0.36 occupy 99.8% and 94.9% of the study area for the cloud amount and SCI, respectively (Figure 7).

9

[Figure 6 about here.]

[Figure 7 about here.]

The NDVI is sensitive to and influenced by land cover. Comparison of Figures 6 and 1 indicates a close relationship between the NCI and observed cloud indices over most of the well vegetated land covers. Table 2 shows the R^2 associated with the cloud indices calculated from satellite-derived data and from ground observations. The R^2 values associated with the NCI and cloud amount range from 0.26 to 0.49, with an average value of 0.42 for all the pixels in the study area. Pixels classified by the USGS to have wooded tundra, mixed shrubland/grassland, and water bodies have the lowest R^2, with average values of 0.26, 0.27, and 0.35, respectively. Standard deviations of R^2 are substantially larger, and the relationships more uncertain, over water bodies and over wooded tundra than over other land types. R^2 values exceeding 0.4 occur for relatively well vegetated land uses including deciduous broadleaf forest, cropland/woodland mosaics, cropland/grassland mosaics, savanna, mixed forest, irrigated cropland and pasture, dryland cropland and pasture, deciduous needleleaf forest, and shrubland. The R^2 associated with the NCI and SCI range from 0.33 to 0.57, with an average value of 0.49 in the study area. The smallest R^2 values again occur over regions classified by the USGS as having wooded tundra. The standard deviations of R^2 values are the largest, indicating uncertainty, over water bodies and wooded tundra. The R^2 values remain higher than 0.5 over the well vegetated land uses mentioned above. The lower R^2 values over wooded tundra suggest that frozen surface soil and snow cover may disturb the relationships between the NCI and observed cloud indices.

[Table 2 about here.]

Figure 6 shows that the weakest relationships for both the cloud amount and SCI occur at the Qinghai-Tibet Plateau, where the land use is classified as wooded tundra. Figure 8 shows R^2 values

10

versus altitude for all the pixels in the study area. Figure 8a provides R^2 values relating the NCI and SCI; these values are approximately 0.45 for altitudes less than 500 m, near 0.40 for altitudes between 500 and 3,500 m, and near 0.20 for altitudes around 4,500 m. Figure 8b shows that the R^2 values for the NCI and cloud amount vary between 0.40 and 0.65 for altitudes below 3,500 m, decreasing sharply to 0.20 at higher altitudes. The relationship at low elevations is more robust than at high elevations. Because frozen surface soil and snow cover are more common at higher elevations, this result further confirms the hypothesis that frozen surface soil and snow cover may disturb the relationship between the NCI and observed cloud indices.

[Figure 8 about here.]

Figure 9 shows seasonal variations in R^2 values in the study area. The mean R^2 value for all the pixels is shown by the curve; standard deviations are shown by error bars. The R^2 values are small in January during boreal winter. Values increase in March, April, and May as spring arrives. Peaks occur in September when snow and frozen surface soil melt on the Qinghai-Tibet Plateau. The R^2 values decrease starting in October as snow begins to cover the Qinghai-Tibet Plateau. Standard deviations in summer are significantly smaller than in winter. The NDVI is a more reliable estimate of cloudiness in summer. Figure 10 gives the difference between monthly maximum NDVI value and the minimum NDVI value. The difference in winter season is general smaller than that in summer season. In winter season, the difference in low elevations is larger than that in high elevations. These indicate the relatively poor relationship occurs in winter because of snow and frozen surface soil, which makes the difference between maximum NDVI value and the minimum NDVI value become smaller.

[Figure 9 about here.]

11

[Figure 10 about here.]

Figure 11 compares the R^2 values in January and September. The R^2 values are large in September for the relationship for both cloud amount and SCI. Smallest values in September occur at high elevations where permanently frozen soil and snow cover exist. The R^2 values are lower in January over the entire study area. Smallest values persist at altitudes above 3,500 m, the elevation of the Qinghai-Tibet Plateau. Values in January at altitudes of 1,000-2,000 m corresponding to the Loess Plateau decrease to the values found at high elevations in September. The R^2 values remain high over the alluvial plain where altitudes are less than 500 m.

[Figure 11 about here.]

The relationships between NCI and observed cloud amount and the SCI are robust over most well-vegetated regions. The strongest relationships occur at low elevations in summer, and the weakest relationships occur at high elevations in winter. Frozen surface soil and snow cover may preclude the NDVI index from correctly determining cloudiness in the latter area. Nevertheless, the daily NDVI index is a useful tool for estimating the influence of clouds and solar radiation over a large area.

4. Discussion

Cloud cover is characterized by large spatial and temporal variations. Only cloud cover estimated at the time of the afternoon satellite overpass was used in this study, so diurnal variations in cloud cover were not considered. However, results show a robust relationship between the daily NDVI cloud index and ground-observed cloud indices; the R^2 values in summer were large, around 0.6

12

(i.e., the correlation coefficient equals 0.8), suggesting that day-to-day cloud variations are captured by the daily NDVI indicator. The NCI is a simple method for estimating clouds and showed good agreement with a ground-observed cloud index. NCI values were more consistent with the ground-observed cloud indices than the CLAVR values were. The agreement between the NDVI index and ground-observed cloudiness suggests that spatial distributions in cloud cover are captured by NDVI values. Correlations between the NCI and ground observations were better in summer than in winter, better over cropland than over wooded tundra, mixed shrubland/grassland, or water bodies, and better over low elevations than over high elevations.

References

Allen, R. G., L. S. Pereira, D. Raes, and M. Smithz, 1998: *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56*. FAO, Rome, Italy, 300 pp.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer, 1976: A land use and land cover classification system for use with remote sensor data. Technical report, U.S. Geological Survey Professional Paper.

Carlson, T. N. and D. A. Ripley, 1997: On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.*, **62**, 241–252.

Choudhury, B. J. and R. E. Golus, 1988: Estimating soil wetness using satellite data. *Int. J. Remote Sens.*, **9**, 1251–1257.

Goward, S. N., B. Markham, D. G. Dye, W. Dulaney, and J. Yang, 1991: Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote Sens. Environ.*, **35**, 257–277.

Grist, J., S. E. Nicholson, and A. Mpolokang, 1997: On the use of NDVI for estimating rainfall fields in the Kalahari of Botswana. *J. Arid Environ.*, **35**, 195–214.

Holben, B. N., 1986: Characteristics of maximum-value composite images for temporal AVHRR data. *Int. J. Remote Sens.*, **7**, 1417–1434.

Hou, Y. T., K. A. Campana, K. E. Mitchell, S. K. Yang , and L. L. Stowe, 1993: Comparison of an experimental NOAA AVHRR cloud dataset with other observed and forecast cloud datasets. *J. Atmos. Ocean. Technol.*, **10**, 833–849.

HYDRO1k Team, 2003: Web page of the HYDRO1k Elevation Derivative Database. Technical report, Available at http://edcdaac.usgs.gov/gtopo30/hydro.

Jackson, R. D., P. N. Slater, and P. J. Pinter, 1983: Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres. *Remote Sens. Environ.*, **13**, 187–208.

James, M. E. and S. N. V. Kalluri, 1994: The Pathfinder AVHRR land data set: An improved coarse resolution data set for terrestrial monitoring. *Int. J. Remote Sens.*, **15**, 3347–3363.

Justice, C. O., J. R. G. Townshend, B. N. Holben, and C. J. Tucker, 1985: Analysis of the phenology of global vegetation using meteorological satellite data. *Int. J. Remote Sens.*, **6**, 1271–1318.

Loveland, T. R., Z. Zhu, D. O. Ohlen, J. F. Brown, B. C. Reed, and L. Yang, 1999: An analysis of the IGBP global land-cover characterization process. *Photogramm. Eng. Remote Sens.*, **65**, 1021–1032.

New, M., M. Hulme, and P. Jones., 1999: Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology. *J. Climate*, **12**, 829–856.

Roerink, G. J., M. Menenti, and W. Verhoef, 2000: Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *Int. J. Remote Sens.*, **21**, 1911C1917.

Stowe, L. L., 1991: Cloud and aerosol products at NOAA/NESDIS. *Glob. Planet. Change*, **4**, 25–32.

Stowe, L. L., P. A. Davis, and E. P. McClain, 1999: Scientific basis and initial evaluation of the

CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer. *J. Atmos. Ocean. Technol.*, **16**, 656–681.

Stowe, L. L., E. P. McClain, R. Carey, P. Pellegrino, G. G. Gutman, P. Davis, C. Long, and S. Hart, 1991: Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Adv. Space Res.*, **11**, 51–54.

Sun, B., P. Y. Groisman, and I. I. Mokhov, 2001: Recent changes in cloud-type frequency and inferred increases in convection over the United States and the Former USSR. *J. Climate*, **14**, 1864–1880.

Szilagyi, J., D. C. Rundquist, D. C. Gosselin, and M. B. Parlange, 2000: NDVI relationship to monthly evaporation. *Geophys. Res. Lett.*, **25**, 1753–1756.

Tarpley, J. D., S. R. Schneider, and R. L. Money, 1984: Global vegetation indices from the NOAA-7 meteorological satellite. *J. Appl. Meteor.*, **23**, 491–494.

Tucker, C. J., 1979: Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.*, **8**, 127–150.

Tucker, C. J., H. E. Dregne, and W. W. Newcomb, 1991: Expansion and contraction of the Sahara Desert from 1980 to 1990. *Science*, **253**, 299–301.

Verhoef, W., M. Menenti, and S. Azzali, 1996: A colour composite of NOAA-AVHRR-NDVI based on time series analysis (1981-1992). *Int. J. Remote Sens.*, **17**, 231–235.

Wiegand, C. L., A. J. Richardson, and E. T. Kanemasu, 1979: Leaf area index estimates for wheat from LANDSAT and their implications for evapotranspiration and crop modeling. *Agron. J.*, **71**, 336–342.

Yang, D., C. Li, H. Hu, Z. Lei, S. Yang, T. Kusuda, T. Koike, and K. Musiake, 2004: Analysis of water resources variability in the Yellow River of China during the last half century using historical data. *Water Resour. Res.*, **40**.

Zhang, S. Y. and T. H. L. Williams, 1997: Obtaining spatial and temporal vegetation data from Landsat MSS and AVHRR/NOAA satellite images for a hydrologic model. *Photogramm. Eng. Rem. S.*, **63**, 69–77.

List of Figures

Figure 1: Study area

$$NCI = (NDVI_r - NDVI_{min}) / (NDVI_c - NDVI_{min}) \qquad [0, 1]$$

Figure 2: NDVI Cloud Index (NCI)
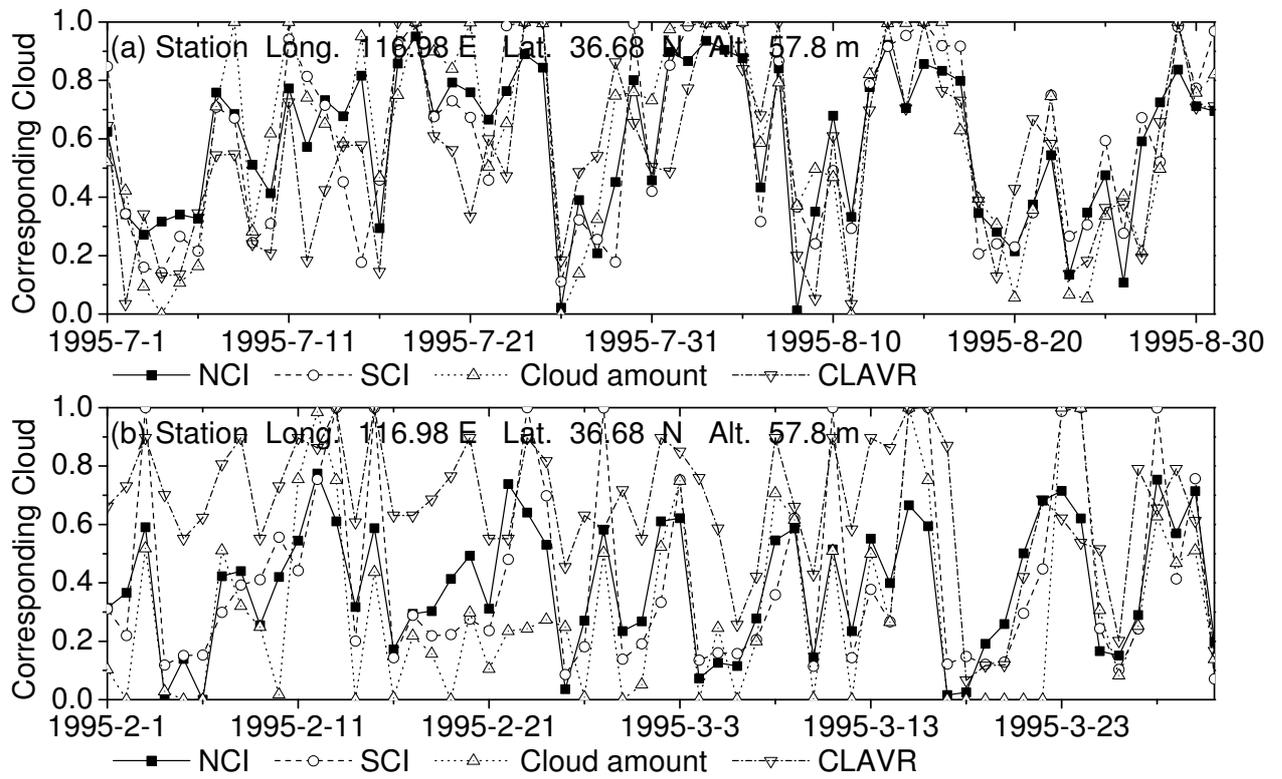
Figure 3: Time series of NCI, SCI, CLAVR, and ground-observed cloud amounts
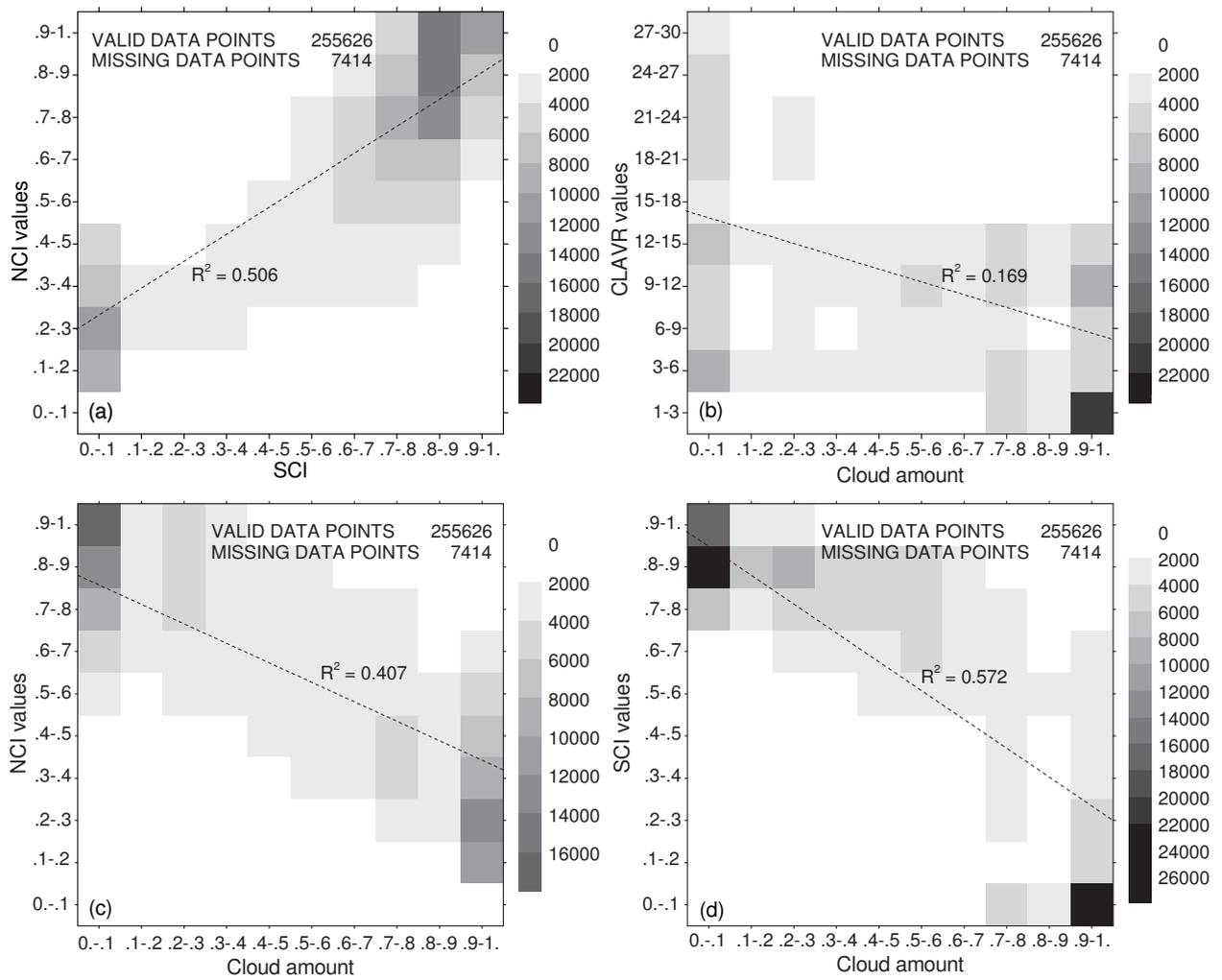
Figure 4: Cloud amount, NCI, SCI, and CLAVR values. (a) NCI values versus SCI; (b) CLAVR versus cloud amount; (c) NCI versus cloud amount; (d) SCI versus cloud amount.
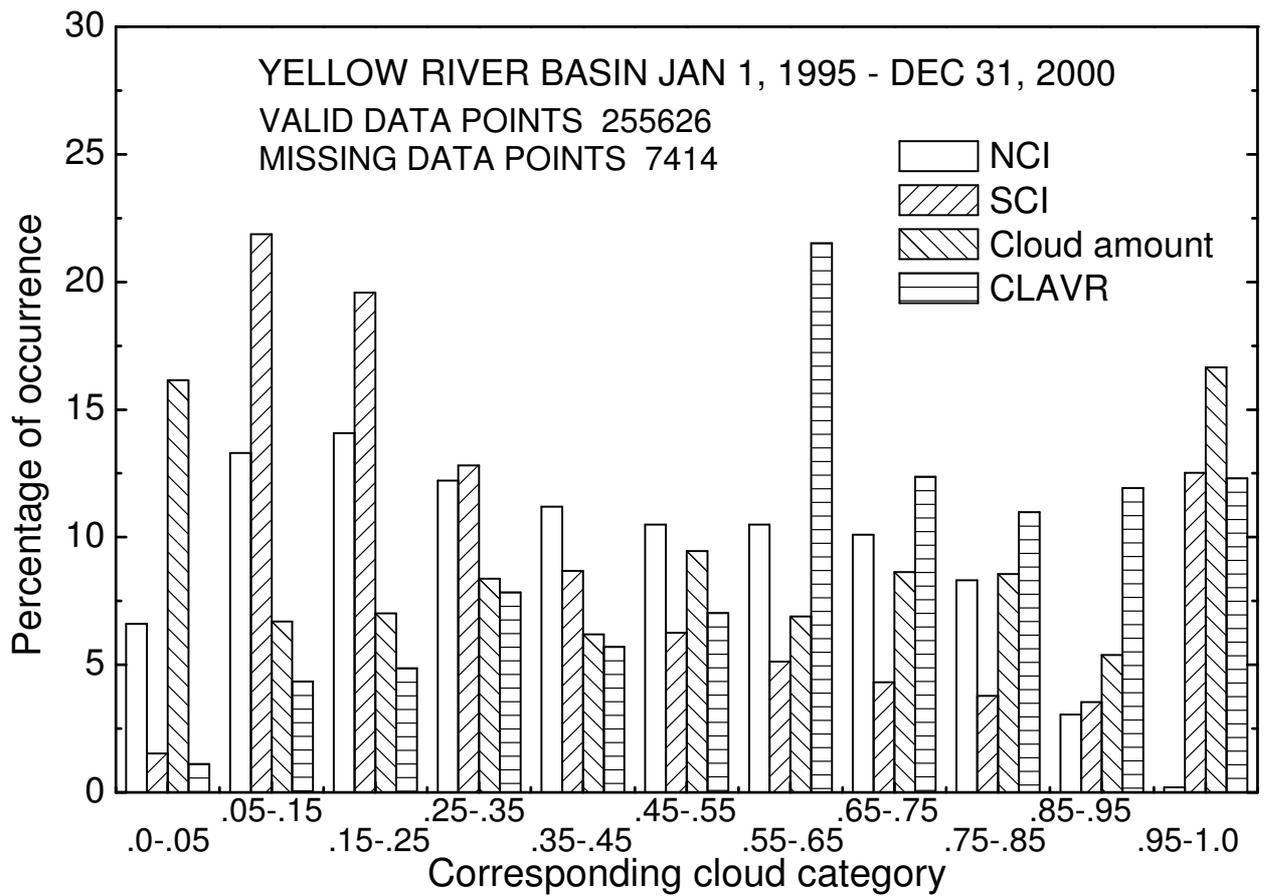
Figure 5: Histogram of cloud distributions at the ground-observation stations.
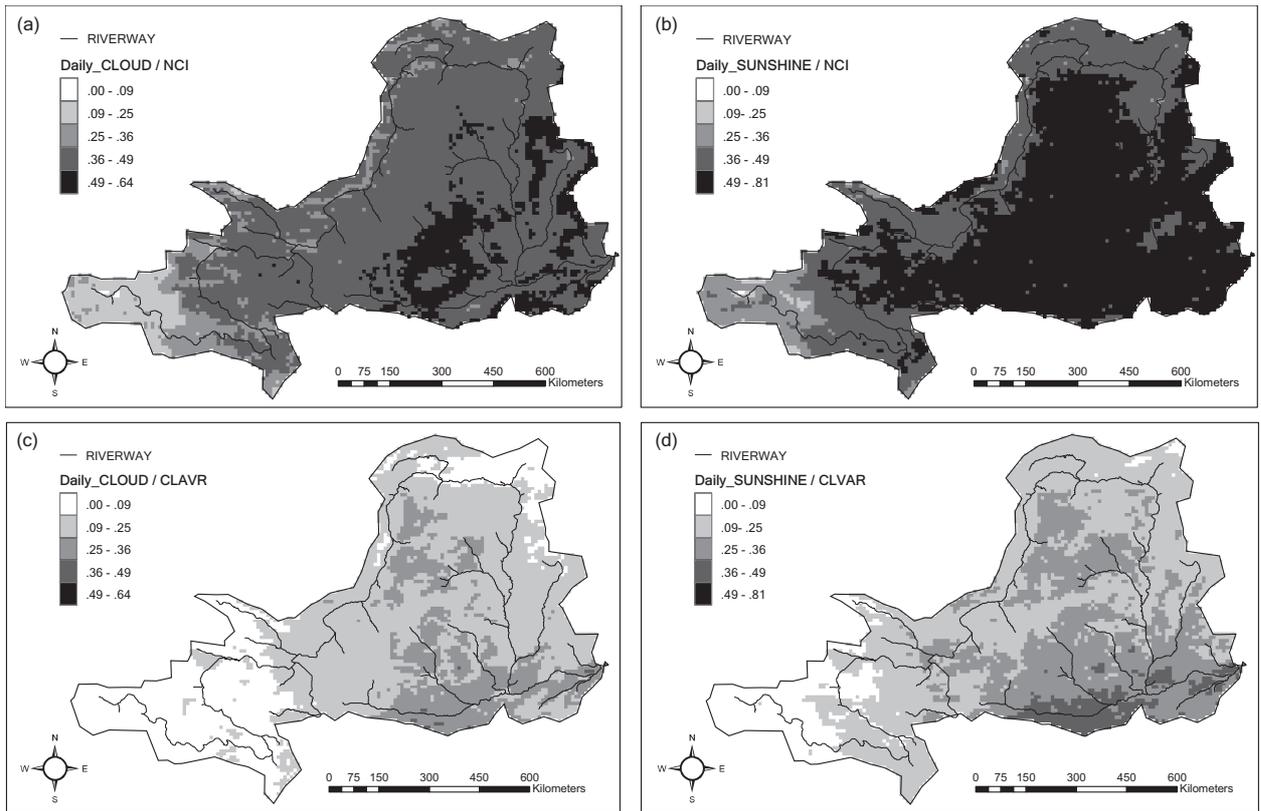
Figure 6: (a) R^2 for daily NCI values and the observed cloud amount; (b) R^2 for daily NCI values and the observed SCI; (c) R^2 for CLAVR values and the observed cloud amount; (d) R^2 for CLAVR values and the observed SCI.

24
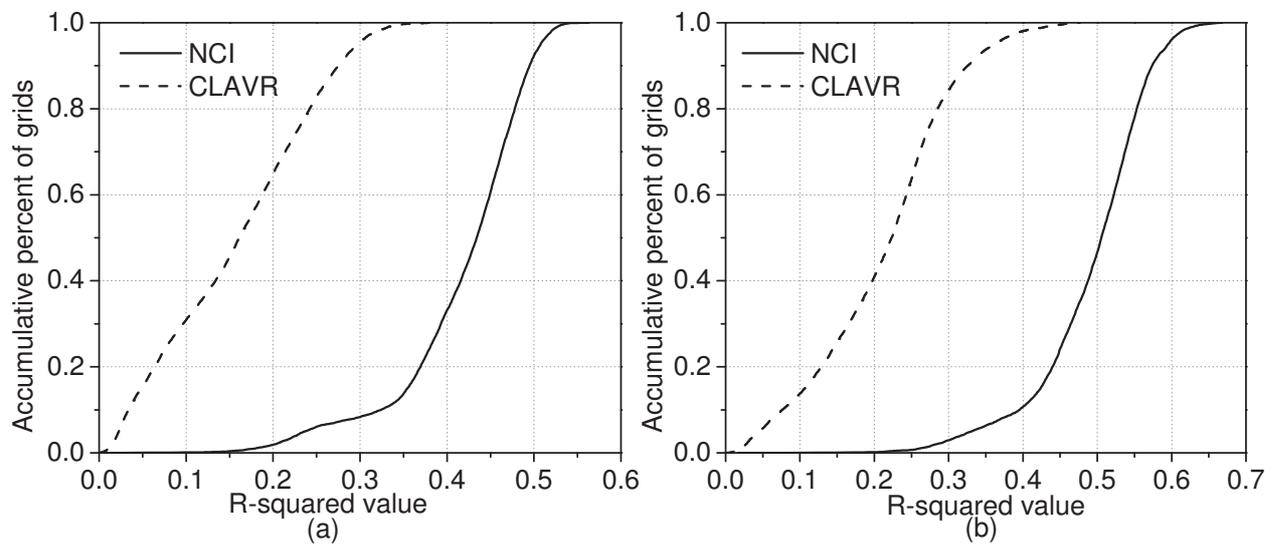
Figure 7: (a) Cumulative distribution of R^2 associated with NCI/cloudiness and CLAVR/cloudiness; (b) cumulative distribution of R^2 associated with NCI/SCI and CLAVR/SCI.
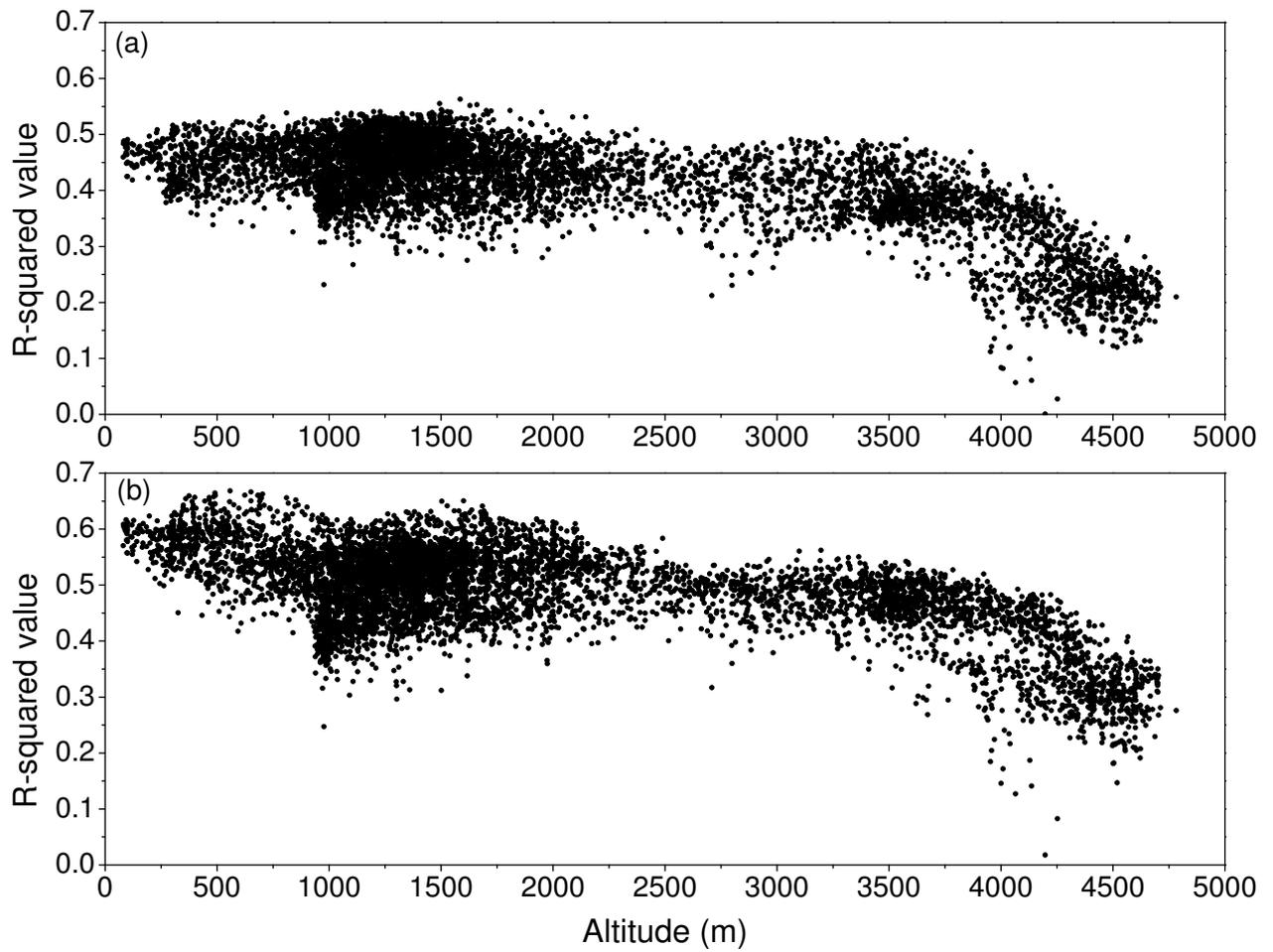
Figure 8: (a) R^2 values for the NCI and cloud amount as a function of altitude; (b) R^2 values for the NCI and SCI as a function of altitude

26
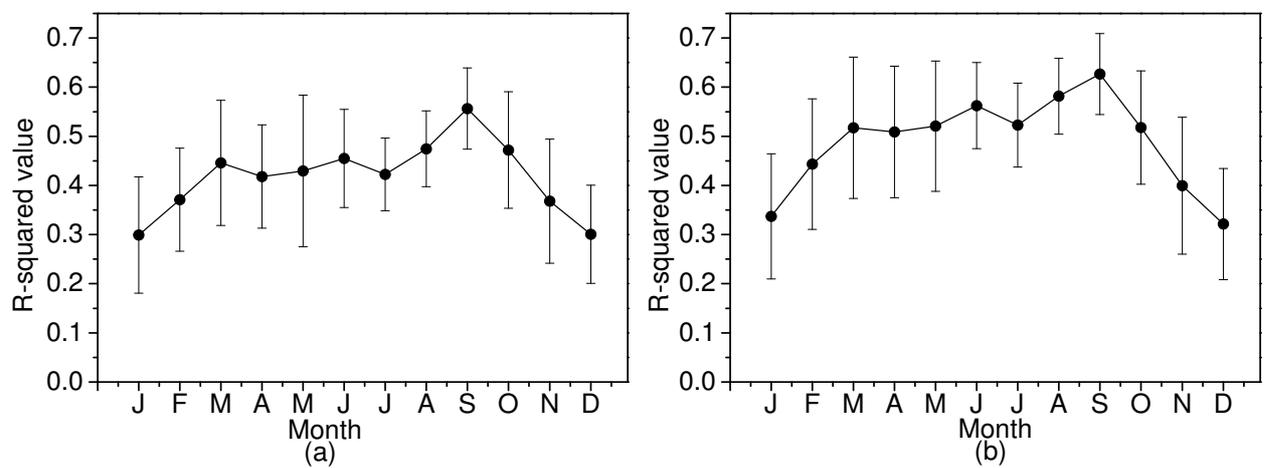
Figure 9: (a) Seasonal variation in R^2 values for the NCI and cloud amount; (b) seasonal variation in R^2 values for the NCI and SCI
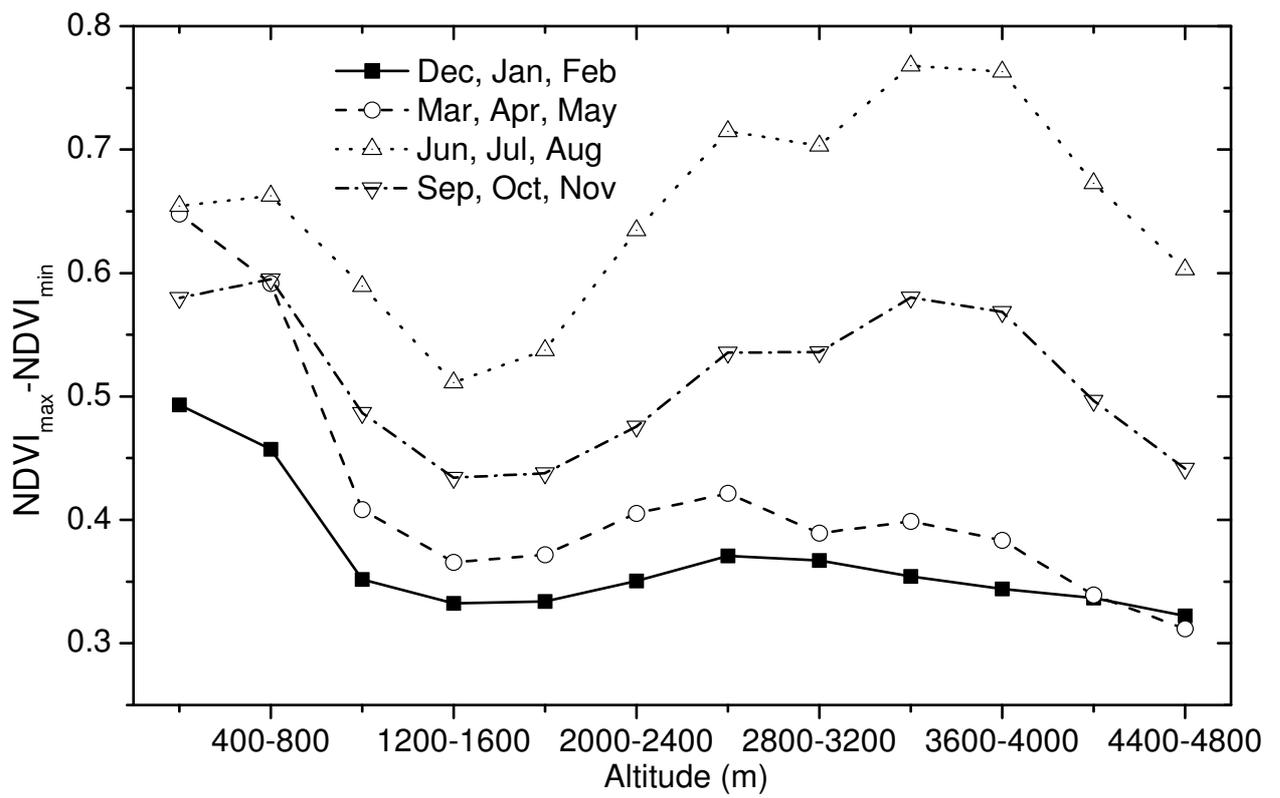
Figure 10: The difference between monthly maximum NDVI value and the minimum NDVI value.
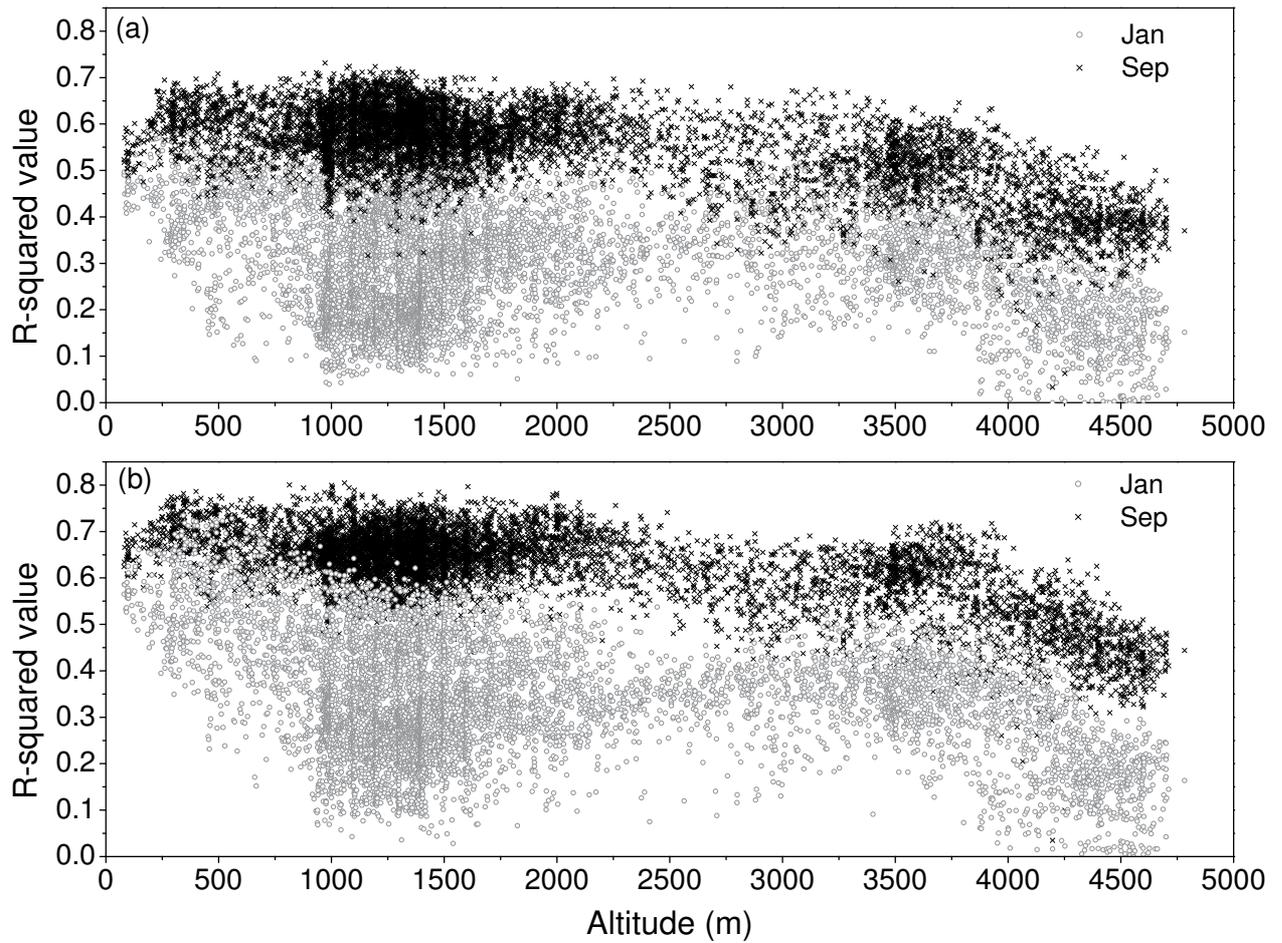
Figure 11: (a) R^2 values for the NCI and cloud amount in January and September as a function of altitude; (b) R^2 values for the NCI and SCI in January and September as a function of altitude.

List of Tables

Table 1: Daily NCI, CLAVR values, and ground observations

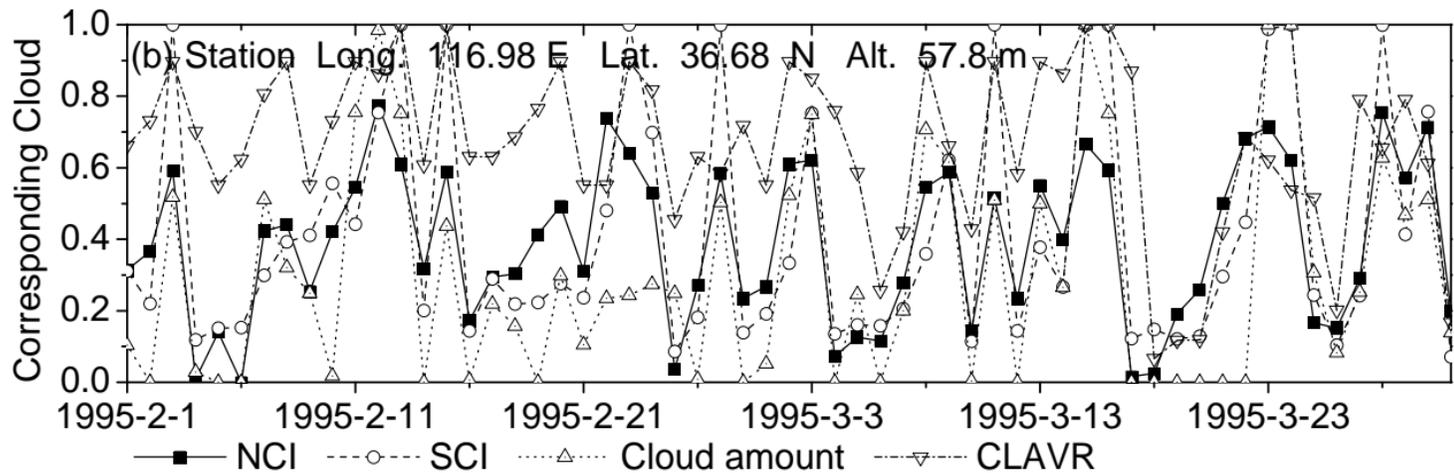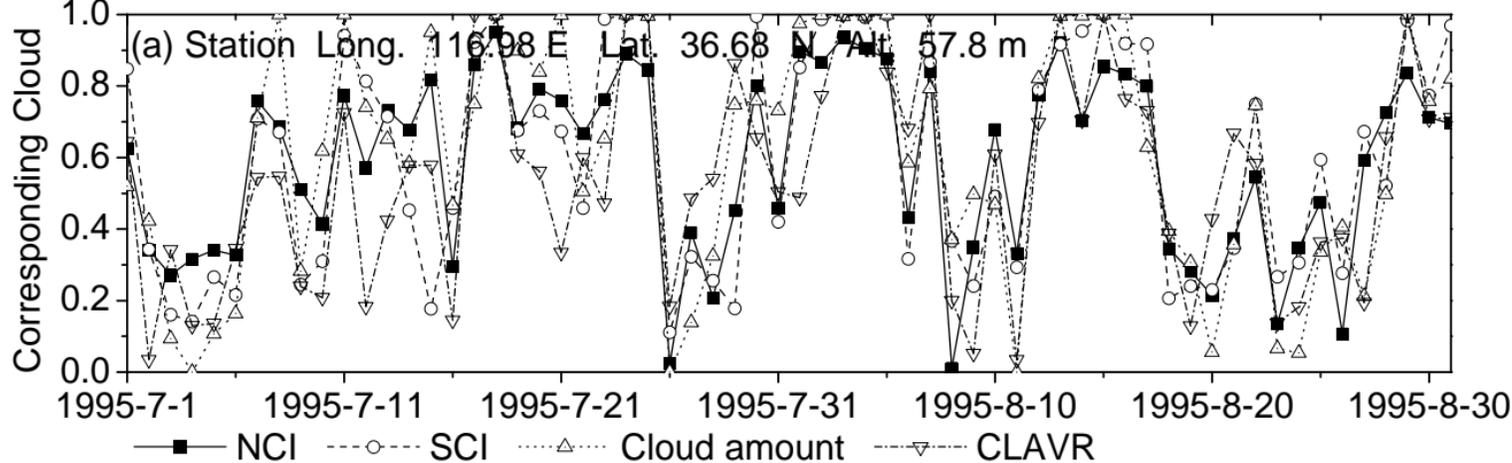| USGS Land Use	NCI/Cloud		NCI/SCI		CLAVR/Cloud		CLAVR/SCI	
	R^2	SD	R^2	SD	R^2	SD	R^2	SD
Urban and Built-Up Land	0.42	0.05	0.53	0.07	0.18	0.10	0.25	0.11
Dryland Cropland and Pasture	0.42	0.06	0.51	0.07	0.17	0.08	0.23	0.10
Irrigated Cropland and Pasture	0.45	0.04	0.58	0.04	0.26	0.07	0.33	0.09
Cropland/Grassland Mosaic	0.45	0.04	0.53	0.03	0.20	0.07	0.28	0.09
Cropland/Woodland Mosaic	0.47	0.07	0.58	0.04	0.32	0.07	0.28	0.09
Grassland	0.38	0.08	0.48	0.07	0.11	0.07	0.16	0.08
Shrubland	0.40	0.07	0.49	0.07	0.20	0.11	0.26	0.15
Mixed Forest	0.39	0.02	0.45	0.09	0.14	0.02	0.21	0.06
Grand summary	0.41	0.07	0.50	0.07	0.16	0.09	0.22	0.11

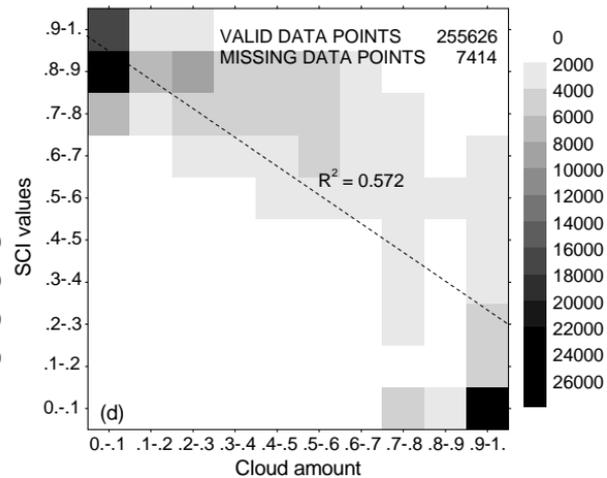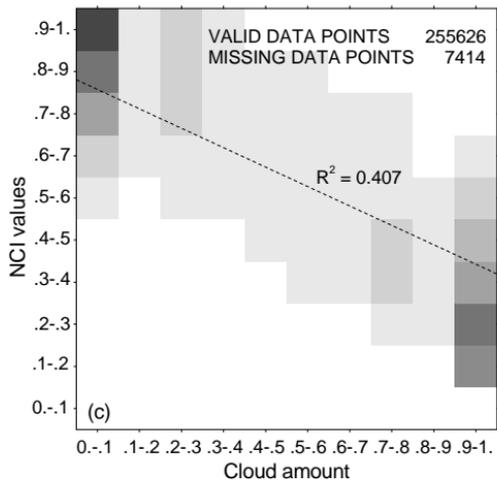Table 2: Daily NCI values and cloud indices derived from observations

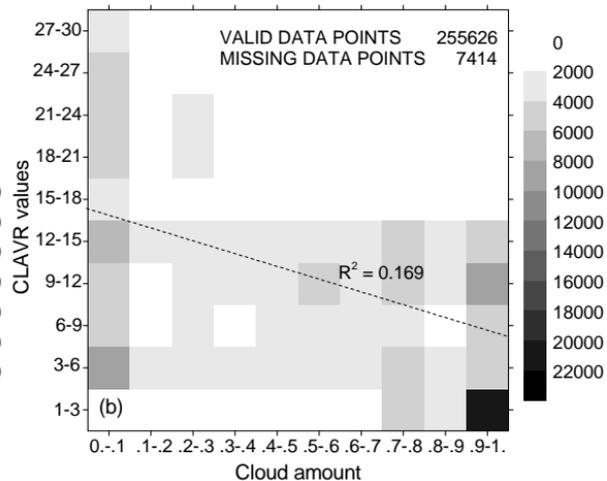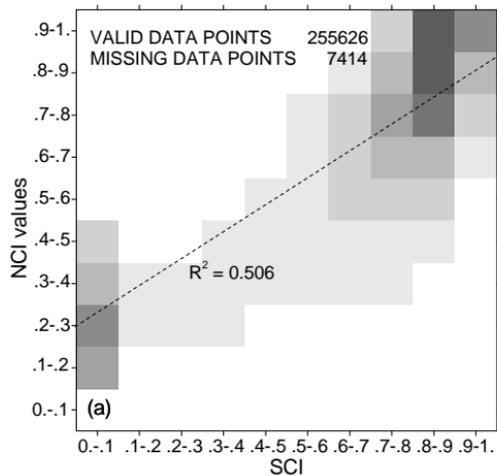| USGS Land Use	Cloud R^2	Cloud SD	SCI R^2	SCI SD
Urban and Built-Up Land	0.37	0.06	0.49	0.06
Dryland Cropland and Pasture	0.44	0.05	0.51	0.07
Irrigated Cropland and Pasture	0.46	0.04	0.57	0.05
Cropland/Grassland Mosaic	0.47	0.04	0.55	0.05
Cropland/Woodland Mosaic	0.48	0.04	0.54	0.04
Grassland	0.39	0.08	0.47	0.08
Shrubland	0.43	0.05	0.50	0.06
Mixed Shrubland/Grassland	0.27	0.06	0.38	0.08
Savanna	0.47	0.05	0.51	0.04
Deciduous Broadleaf Forest	0.49	0.03	0.52	0.04
Deciduous Needleleaf Forest	0.43	0.05	0.51	0.02
Mixed Forest	0.46	0.05	0.52	0.05
Water Bodies	0.35	0.13	0.43	0.14
Barren or Sparsely Vegetated	0.37	0.05	0.45	0.06
Wooded Tundra	0.26	0.10	0.33	0.10
Grand summary	0.42	0.08	0.49	0.08

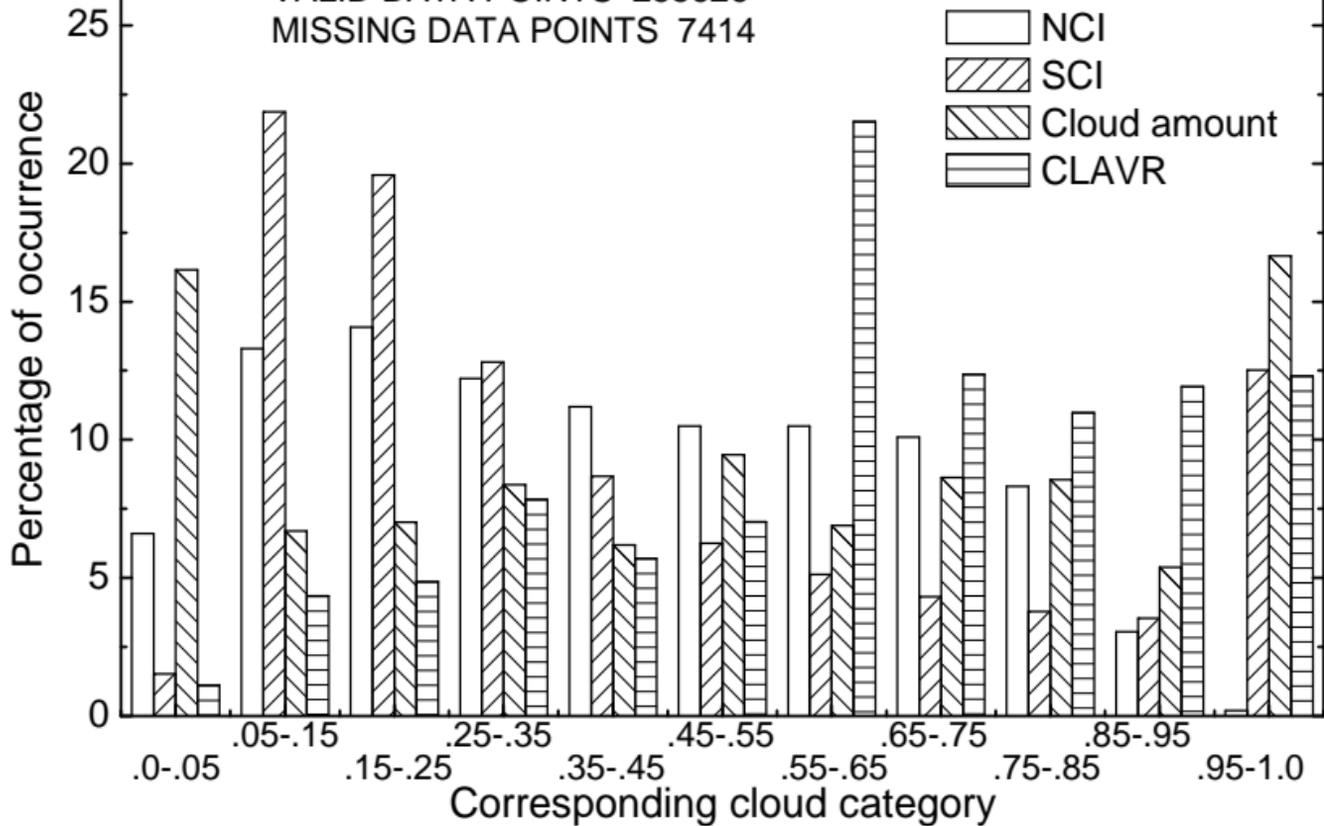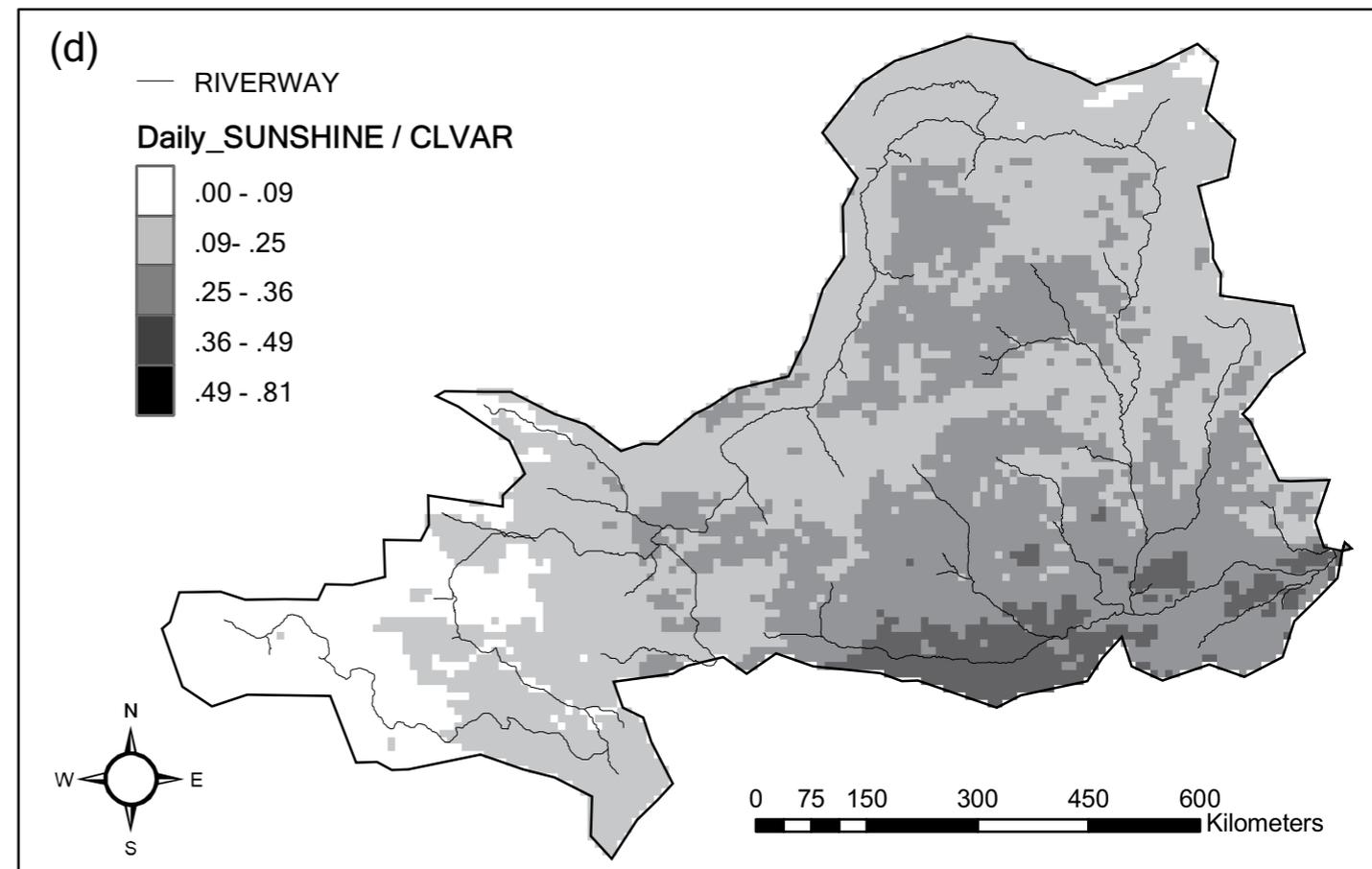$$NCI = (NDVI_r - NDVI_{min}) / (NDVI_c - NDVI_{min}) \quad [0, 1]$$

(a) Station Long. 116.98 E Lat. 36.68 N Alt. 57.8 m

(b) Station Long. 116.98 E Lat. 36.68 N Alt. 57.8 m
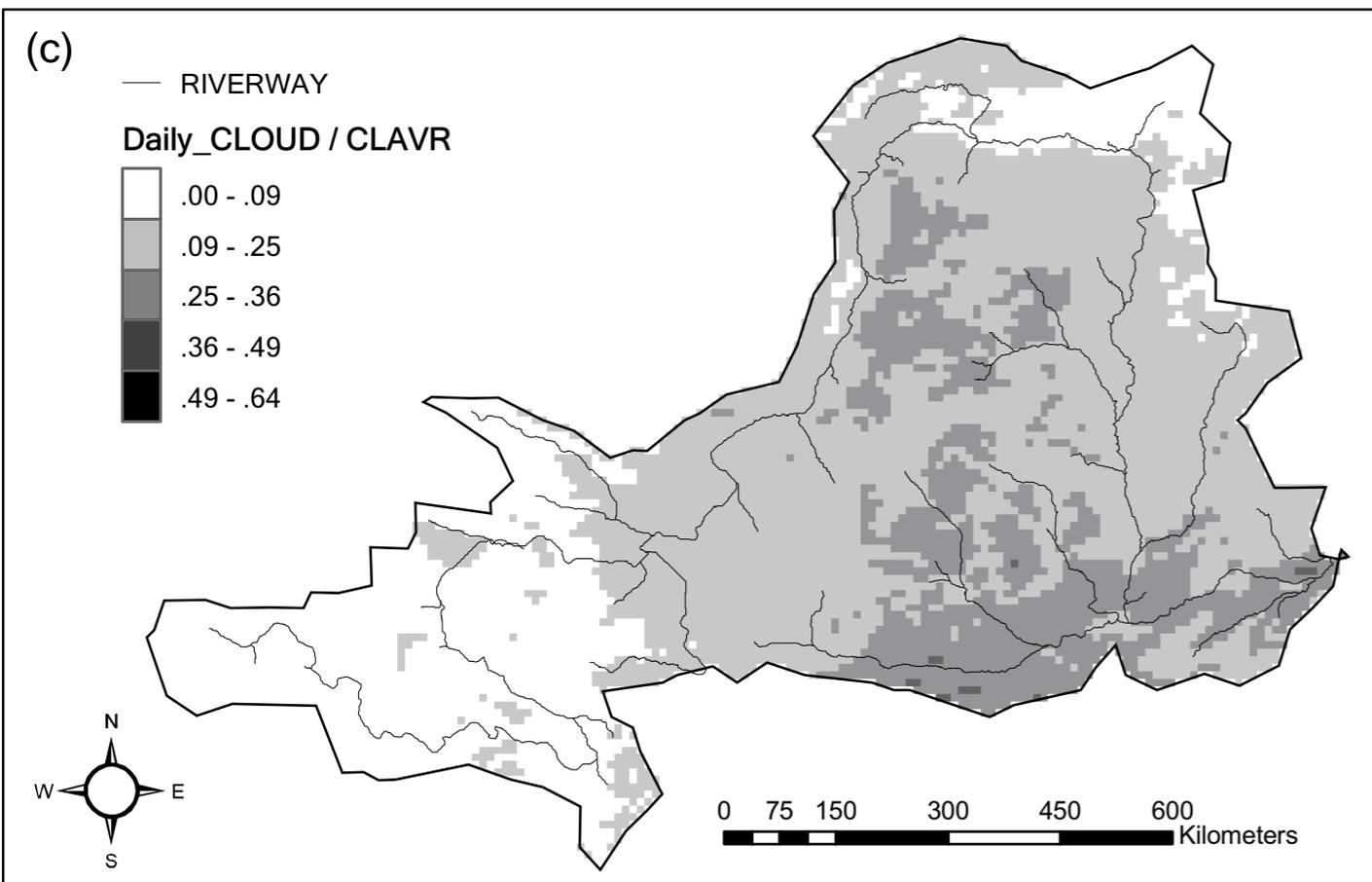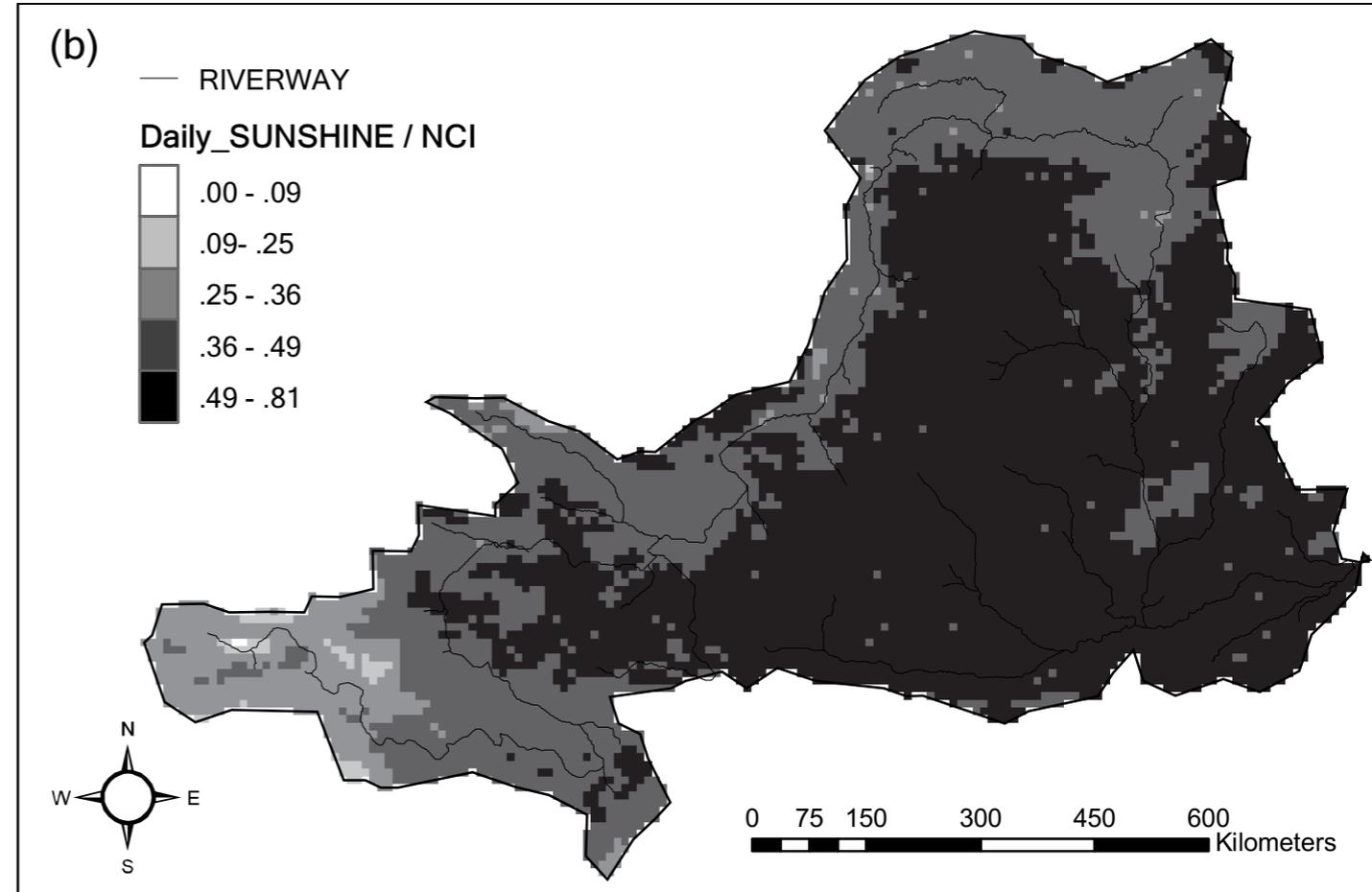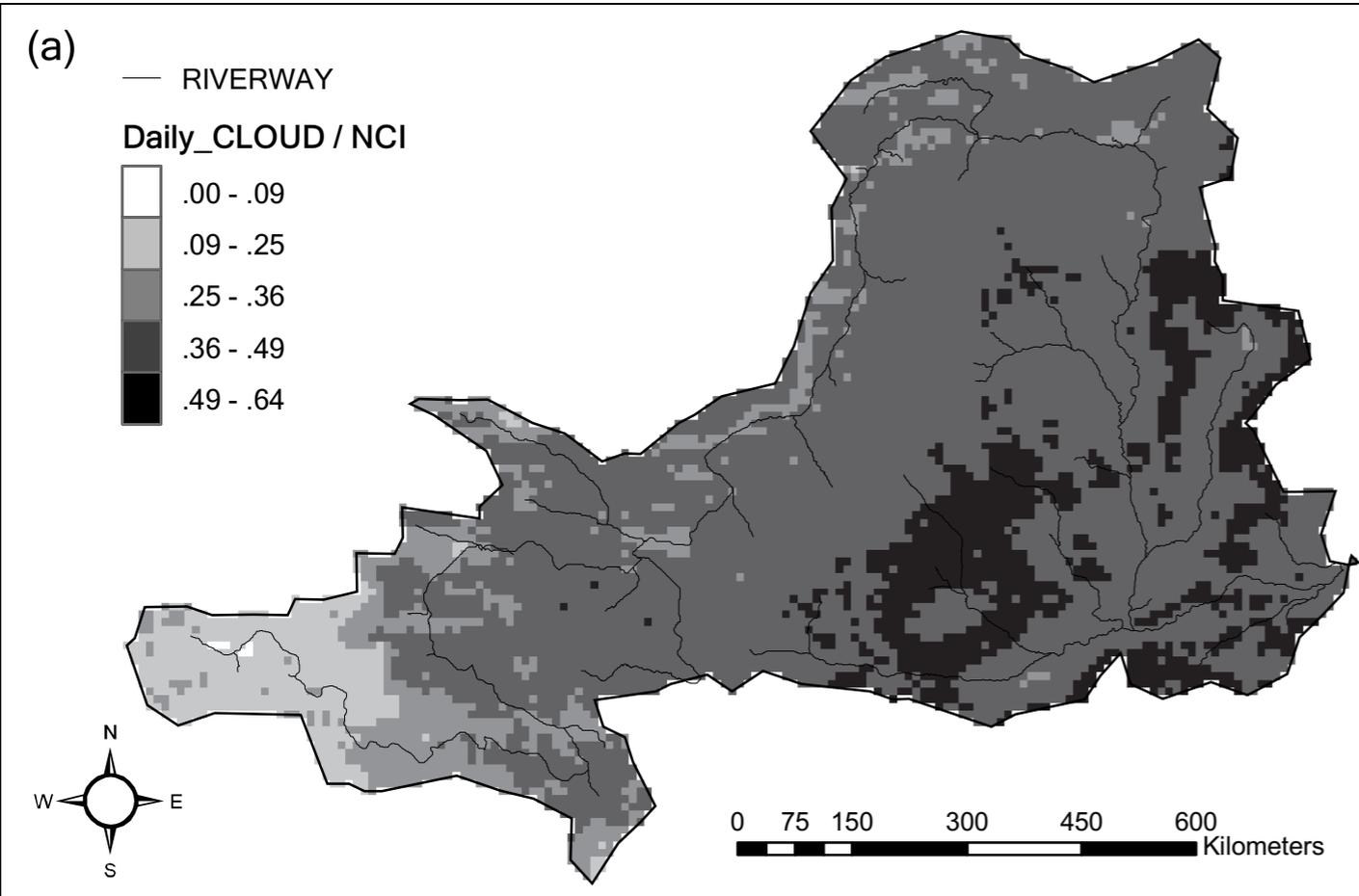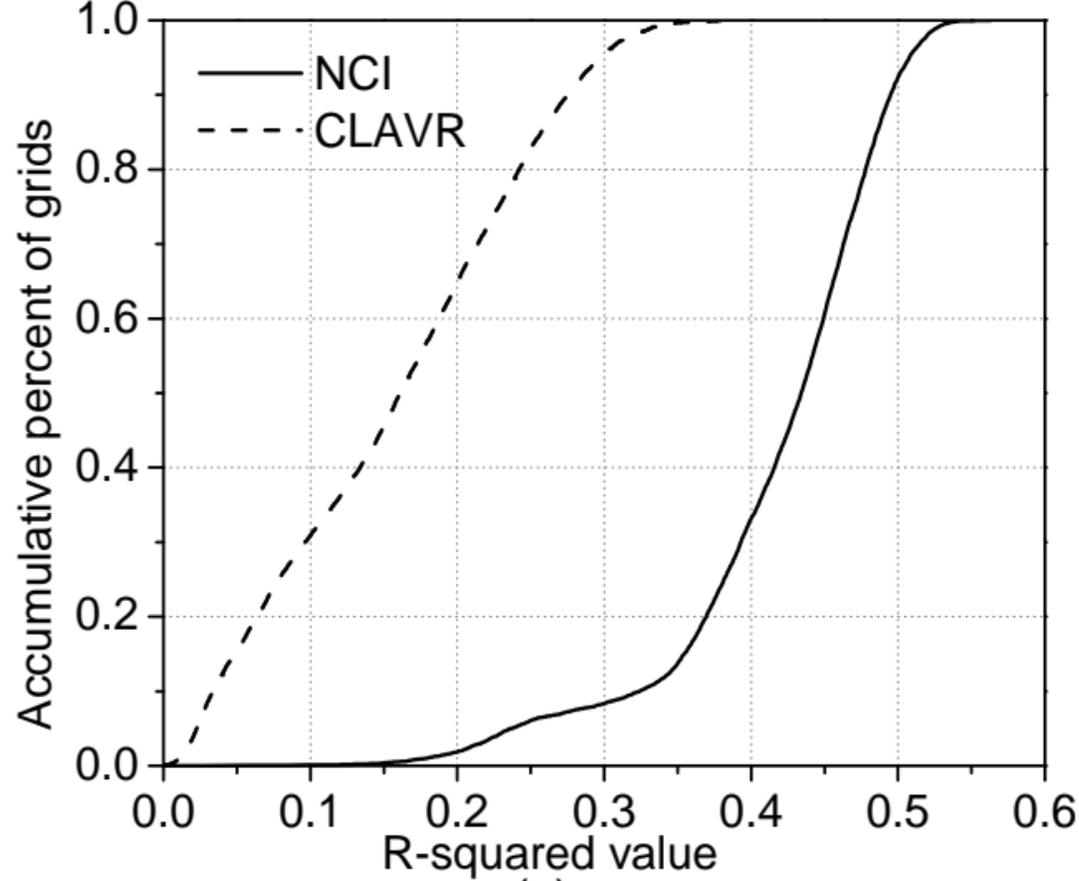
NCI   SCI   Cloud amount   CLAVR
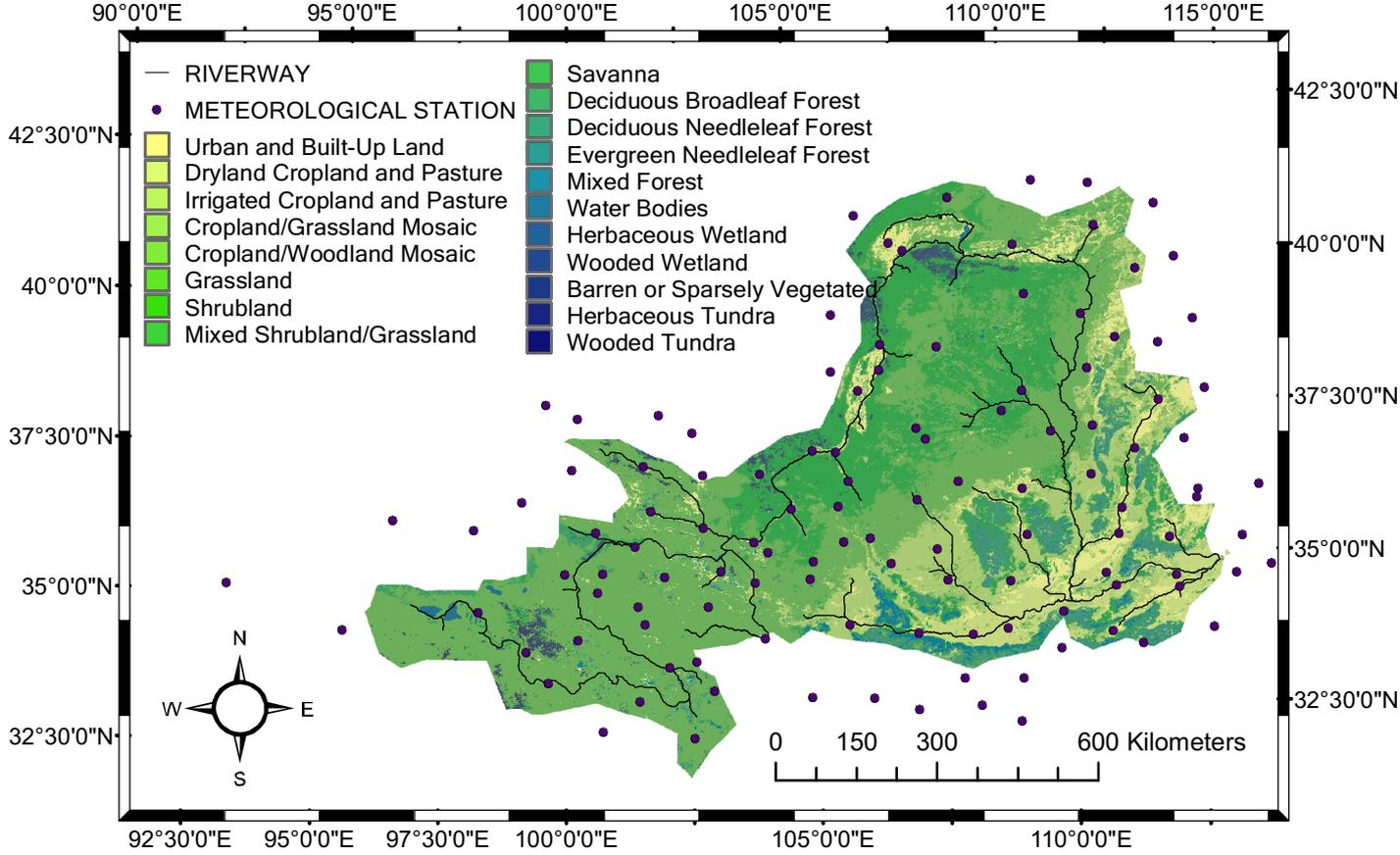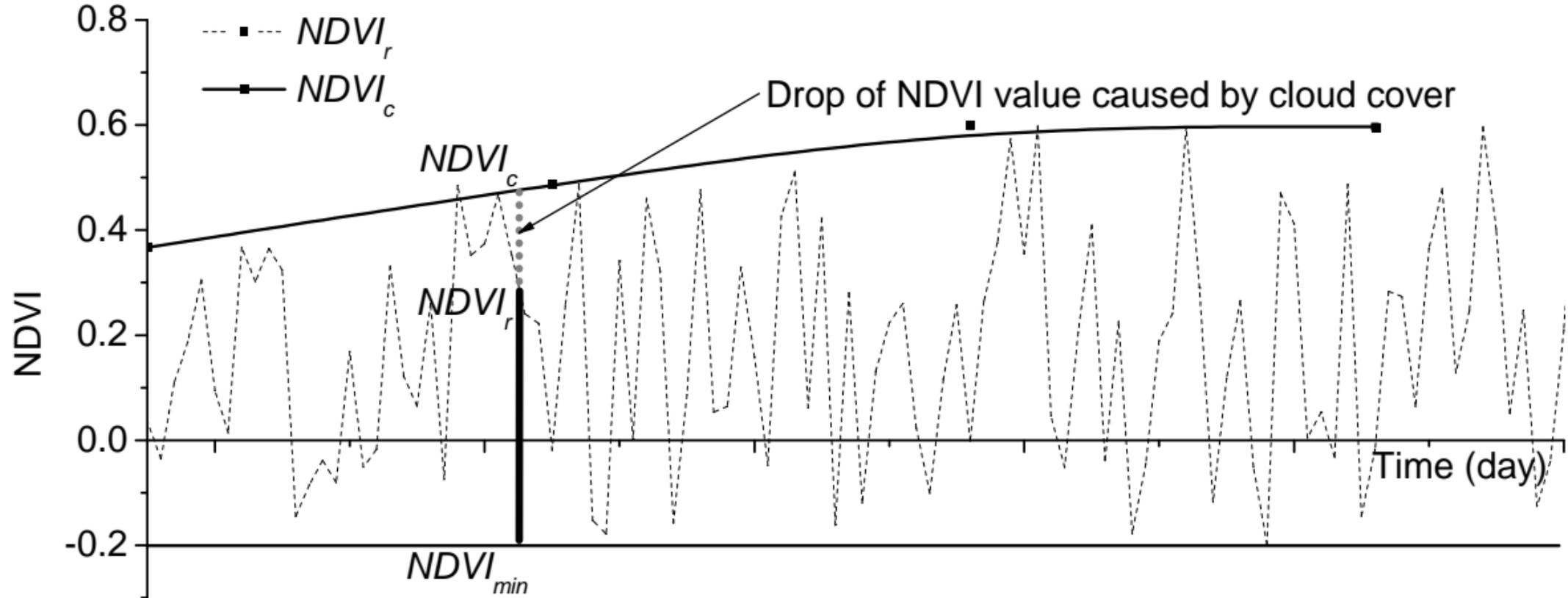
YELLOW RIVER BASIN JAN 1, 1995 - DEC 31, 2000
VALID DATA POINTS 255626
MISSING DATA POINTS 7414

NCI
SCI
Cloud amount
CLAVR

Percentage of occurrence

Corresponding cloud category

(a) Daily_CLOUD / NCI
- .00 - .09
- .09 - .25
- .25 - .36
- .36 - .49
- .49 - .64

(b) Daily_SUNSHINE / NCI
- .00 - .09
- .09 - .25
- .25 - .36
- .36 - .49
- .49 - .81

(c) Daily_CLOUD / CLAVR
- .00 - .09
- .09 - .25
- .25 - .36
- .36 - .49
- .49 - .64

(d) Daily_SUNSHINE / CLVAR
- .00 - .09
- .09 - .25
- .25 - .36
- .36 - .49
- .49 - .81

(a)   (b)

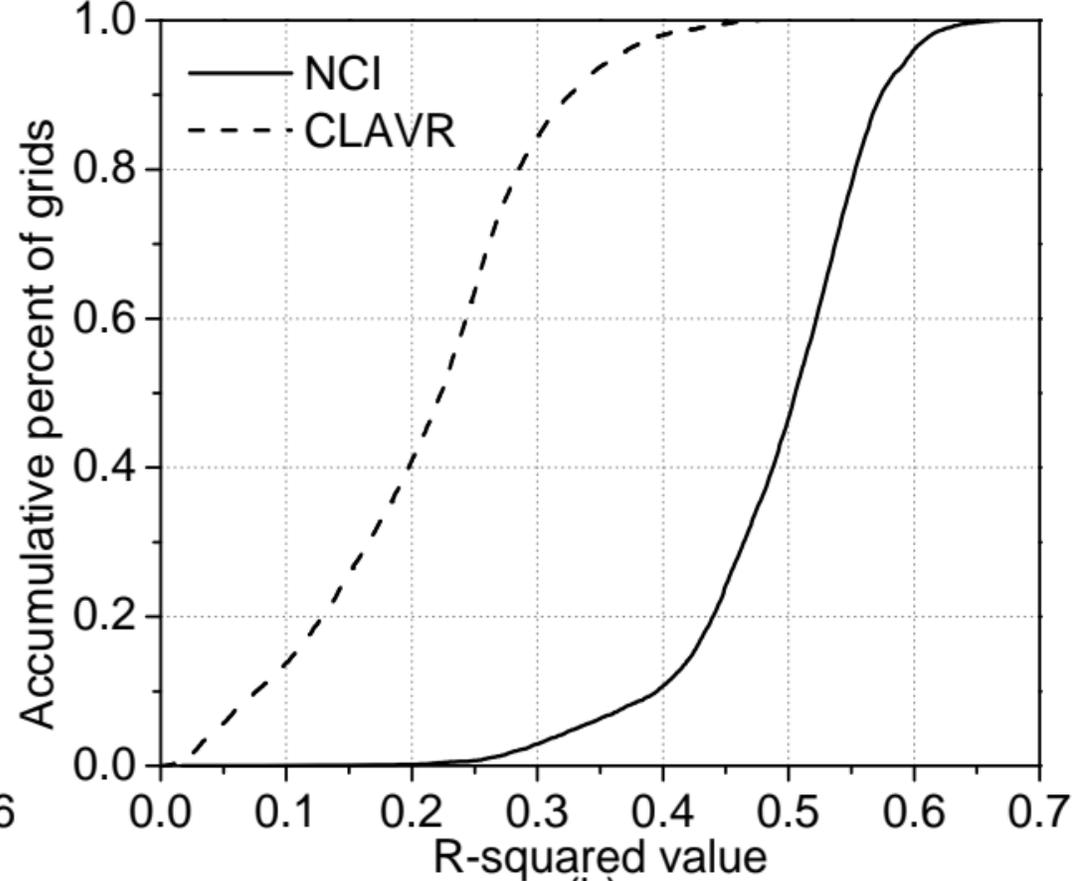(a)   (b)