

Evaluation of Approaches to Identifying Hail Damage to Crop Vegetation Using Satellite Imagery

JORDAN R. BELL

Department of Atmospheric Science, University of Alabama in Huntsville, Huntsville, Alabama

ANDREW L. MOLTHAN

NASA Marshall Space Flight Center/Earth Science Office, Huntsville, Alabama

(Manuscript received 31 December 2015; review completed 9 June 2016)

ABSTRACT

During the growing season in the central United States, severe thunderstorms frequently occur and produce large hail that damages the underlying vegetation, often in agricultural areas. Satellite remote sensing provides a tool for identifying these damaged areas. Previous studies have used changes in the normalized difference vegetation index (NDVI) to identify and examine these areas of damage, but have done so in a manual, time-consuming manner. This study examines an automated approach to detecting areas of hail damage in satellite imagery. Two techniques are evaluated: (i) use of an NDVI change threshold and (ii) detection of anomalies that occur in both daily NDVI and land surface temperature imagery. The two techniques are scored against one another using three different case studies. Two of the case studies occurred late in the growing season in August, and the third occurred in the growing season in early June. The NDVI threshold performed well in the two August case studies with a final probability of detection (POD) ranging from 0.497 to 0.647, whereas the anomaly detection for these two case studies had a lower POD of 0.317 to 0.587. The early June case study highlighted the limitations of using an NDVI threshold and the strengths of using anomaly detection. The POD for the NDVI threshold technique was 0.07–0.08 with a false alarm ratio (FAR) of 0.661–0.758, whereas the anomaly detection had a POD of 0.399–0.418 and a FAR of 0.540–0.681 for this third case study.

1. Introduction

The primary growing season for many agricultural crops (e.g., corn and soybeans) in the central United States runs from late May through late August. This period coincides with the occurrence of severe thunderstorms that bring damaging winds, large hail, and occasional tornadoes, resulting in damage to these crops and other surface vegetation. On average, hail causes anywhere from ~\$161.4¹ million [Storm Prediction Center (SPC) 2016] to ~\$580 million (Changnon et al. 2009) in damage to crops annually. Areas of hail-damaged crops may be localized, but in some severe weather events, the areal coverage of the damage can be quite extensive. Manual surveys of hail-damage areas with large spatial extent have been

completed, but these surveys have been difficult and time-consuming to perform.

Satellite remote sensing has been used previously in some surveys, and automated tools can be incorporated to increase the skill of future surveys. Remote sensing of hail and severe weather damage to crops has been well-documented. Klimowski et al. (1998) used the *Geostationary Operational Environmental Satellite-8 (GOES-8)* to observe hail damage in South Dakota. *GOES-8* only provided visible wavelength (0.52–0.72 μm) imagery with 1-km spatial resolution at nadir. However, the damage area was observed to be 120-km long with varying widths ranging from 7 to 11 km, totaling 200 000 acres (~809 km²) of destroyed vegetation and another 300 000 acres (~1214 km²) of significant damage.

In July 2003, several rounds of severe thunderstorms moved through southeastern North Dakota, northeastern South Dakota, and western Minnesota,

¹ Annual average for 2007–2015 from SPC storm report data.

leaving several hail-damage swaths across the landscape. Parker et al. (2005) observed and analyzed these swaths using the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments aboard Terra and Aqua, two polar-orbiting NASA research satellites. MODIS instruments have 36 channels that include observations in the visible, near-infrared, and thermal-infrared portions of the electromagnetic spectrum. This allowed Parker et al. to difference the 7-day pre- and post-event normalized difference vegetation index (NDVI) composites at 500-m spatial resolution for identifying areas of damage. NDVI is computed by the following equation:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

and for MODIS is based on the near-infrared (NIR, 0.865 μm) and red (0.645 μm) bands. NDVI is used to monitor the greenness (health) of the vegetation on a scale of -1 to 1 , with values closer to 1 indicating greener vegetation (Kriegler et al. 1969). In agricultural areas, NDVI values for crops are for the most part synchronous with the growing season, except those few that are winter-based (Fig. 1). The majority of crops across the Corn Belt will begin to green up in late spring and continue greening as they mature. These crops (corn and soybeans) will reach their peak NDVI values just prior to being harvested (Fig. 1).

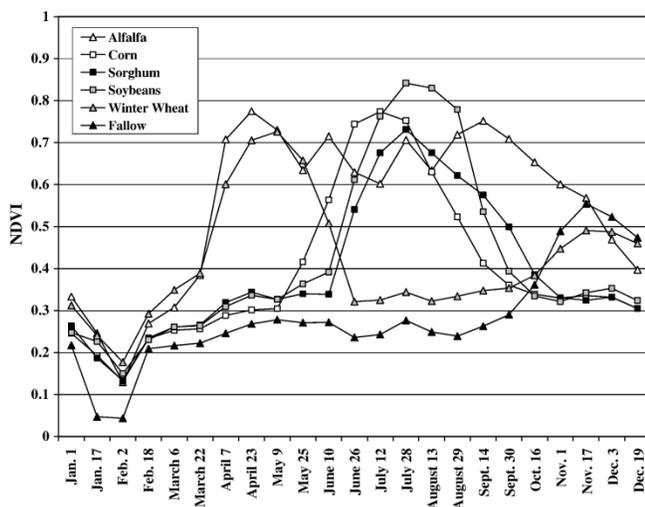


Figure 1. Normalized difference vegetation index (NDVI) values of common agricultural crops in the state of Kansas (reproduced from Fig. 3 in Wardlow and Egbert 2008). *Click image for an external version; this applies to all figures hereafter.*

Parker et al. (2005) observed warmer brightness temperatures in *GOES* infrared imagery (11.7 μm) in

and around the damaged areas when compared to the surrounding undamaged areas. The damaged areas reported an increase in temperature (0.9°C) and a decrease in dewpoint (0.3°C), which in turn increased convective available potential energy and decreased convective inhibition (CIN). Segele et al. (2005) observed similar trends in a 1997 hail streak in South Dakota. Daily temperature (dewpoint) trends observed for seven days after the damage occurred showed a 2°C increase (2.6°C decrease). Parker et al. (2005) performed numerical simulations with a portion of the first hail swath represented within the modeled land surface. Results showed that the damage swath was able to help with the development of deep convection, especially within simulations that contained a modest amount of CIN. Following the creation of the damage scar, these changes in sensible heat and buoyancy appeared to influence storms that formed later in July. Segele et al. (2005) also performed numerical simulations to see how this change in vegetation would change the boundary layer. They noted that updated daily land surface information was ideal for capturing the effect that a hail streak would have on the boundary layer.

On the morning of 9 August 2009, central Iowa was affected by severe thunderstorms with damaging hail, and vegetation damage was visible in MODIS imagery. Gallo et al. (2012) compared pre- and post-storm NDVI in the damage area to available radar data and ground reports of hail. NDVI imagery from before the event showed a uniform, green vegetation pattern across portions of central Iowa, with sporadic lower values representing urban areas. Imagery from the post-event timeframe showed multiple large areas of lower NDVI as a result of the storms. Hail-affected vegetation resulted in NDVI decreases in the week after the severe storms, ranging from -0.021 to -0.357 . Gallo et al. noted that 77% of the pixels that received radar-estimated hail of any size resulted in a decrease of NDVI between the pre- and post-storm composites. When the threshold for radar-estimated hail size was increased to >2.54 cm (1 in) and >5.08 cm (2 in), the percentage of pixels with a decrease in NDVI increased to 87% and 92%, respectively.

Molthan et al. (2013) examined hail swaths produced from severe thunderstorms that occurred on 18 August 2011 across extreme southeastern South Dakota, eastern Nebraska, southwestern Iowa, and northwestern Missouri. Molthan et al. performed a similar analysis as Gallo et al. (2012) but also incorporated observations with higher spatial resolutions from addi-

tional earth-observing instruments, including Landsat-7 Enhanced Thematic Mapper Plus (30-m spatial resolution), Terra Advanced Spaceborne Thermal Emission and Radiometer (15 m), Système Pour l'Observation de la Terre-4 (SPOT-4, 20 m), and SPOT-5 (10 m). Molthan et al. (2013) obtained similar results to those of Gallo et al. (2012) with respect to the correlations between NDVI change and maximum estimated size of hail (MESH) size. Molthan et al. (2013) used the higher spatial resolution sensors to confirm field-scale damage and NDVI reductions in areas that corresponded to the MESH hail-fall streaks and coarser-resolution MODIS observations.

This study builds on previous work by comparing damage identified by NDVI change (thresholding) and an image-processing technique (anomaly detection) to identify areas of hail damage in Aqua MODIS imagery. This study used only Aqua MODIS because the daily observation time is close to the peak of diurnal heating, and the use of a land surface temperature (LST) product is included in the anomaly detection approach. Diurnal heating allows greater contrast between damaged and undamaged areas. Techniques described herein can be converted to an automated algorithm to attempt an objective mapping of damage to vegetation in a near-real-time (NRT) product. Such an algorithm will assist in mapping hail damage following severe thunderstorms. The NDVI change threshold is based on the difference between a pre-event NDVI composite and post-event single-day NDVI image, as used previously by Gallo et al. (2012) and Molthan et al. (2013). The anomaly detection technique uses single-day NDVI and LST anomalies to identify potential areas of damage. The two techniques are each tested on three case studies and compared against a subjective, manual survey of hail-damage areas. Comparisons highlight their relative strengths and weaknesses and determine a technique appropriate for an NRT algorithm. The success of both techniques depends heavily on the impacted vegetation and land cover. Areas of the Midwest that are more uniform in their vegetation patterns and over a flat landscape provide a simpler background for anomaly and threshold-based detections, with more success than in the Great Plains and southern plains where the vegetation is less uniform and the terrain is more complex. The development of an NRT algorithm described herein focuses on areas of the Midwest with significant coverage of spring and summertime agriculture coinciding with severe thunderstorms.

2. Data and methods

a. Data

1) MODIS IMAGERY

MODIS surface reflectance was acquired through the NASA Land Processes Distributed Active Archive Center (available online at lpdaac.usgs.gov/). The MYD09GQ product provides daily surface reflectance in the red (0.645 μm) and near-infrared (0.865 μm) wavelengths at 250-m spatial resolution (Justice et al. 2002). This Level 2 product is atmospherically corrected for gaseous and aerosol scattering and absorption, cirrus contamination, and atmosphere coupling effects. Additional corrections were made for the bidirectional reflectance distribution function and adjacency effects caused by variations in the land cover (Vermote and Vermeulen 1999). The MYD09GQ product was used to create single- and multi-day composite NDVI imagery. Although other vegetation indices (e.g., enhanced vegetation index, Jiang et al. 2008) incorporate information from other spectral bands, the use of red and near-infrared reflectance from MODIS focuses on imagery with the highest spatial resolution (250 m). The Level 2 MYD09GA product provides all MODIS visible and near-infrared bands at 500-m spatial resolution (Justice et al. 2002) and was used to create various true- and false-color composites for visual inspection and identification of hail-damaged areas.

The MYD11A1 product provided daily LST and quality assurance (QA) data at 1-km spatial resolution (Justice et al. 2002). The LSTs are derived from MODIS bands 31 (11.03 μm) and 32 (12.02 μm) with QA flags included to identify clouds and cloud shadows, poor observations, and other low-quality observations (Roy et al. 2002). The QA dataset from the MYD11A1 product was used to identify pixels that may be cloud-contaminated, which were then removed from the analysis in both the MYD11A1 and MYD09GQ products. Because clouds typically produce negative values of NDVI—and cloud shadows may reduce the NDVI despite no actual damage to the vegetation being present—the quality of damage detection is contingent upon the quality of the MODIS QA band.

The MODIS products described herein are delivered on a fixed sinusoidal grid that allows end users to focus only on their areas of interest. The coverage area for this study focused on the central plains, covered by four of the grid tiles. For every cell in each tile of these

products, one or two daytime viewings by the Aqua MODIS instrument are possible each day, depending on whether two swaths overlap a given pixel. Gridded MODIS products select the observation with the best (nearest nadir) viewing angle and cloud-free conditions. In some cases, a tile can have pixels with observations from two different overpasses (times of day). This will have a limited effect on the MYD09GQ and MYD09GA datasets because they are observing the surface reflectance and will be impacted primarily by changes in cloud cover and sun angle between the two observations. The MYD11A1 product will be impacted by the variation in observation time because of diurnal heating and cooling, rainfall, and other weather effects. This can lead to occasional artifacts (seams) observed in the data, but these artifacts did not interfere with the results of this study. The MODIS Reprojection Tool (lpdaac.usgs.gov/tools/modis_reprojection_tool) was used to stitch the various tiles together, with nearest-neighbor resampling of coarser 1-km products used to map them to the 250-m grid for analysis.

2) MAXIMUM ESTIMATED SIZE OF HAIL

Hail occurrence and size estimates for this study were taken from the National Severe Storm Laboratory (NSSL) MESH product. MESH is a 1-km \times 1-km composited radar product available across the domain of the Weather Surveillance Radar-1988 Doppler network (Stumpf et al. 2004). This composite is derived from the severe hail index (SHI, Witt et al. 1998), which was developed as a primary predictor for severe hail sizes (Witt et al. 1998). SHI was used as the baseline for a model of the maximum hail size by comparing SHI to actual observed hail-size reports. Witt et al. noted that thunderstorms generating large hailstones likely would have other, smaller hailstones falling at the same time. The authors argued that the most likely hailstone size reported would be a smaller size that fell over a larger spatial area rather than the largest stone that fell over a smaller spatial extent. MESH was defined as the hail diameter larger than 75% of the hail observed from a given storm. Ortega et al. (2009) noted a wide range of measured hail sizes within a certain MESH value. Given these issues, MESH is used here as a proxy for likely hailfall size and location, rather than confirmation of the actual size of hail reaching the surface.

3) NATIONAL LAND COVER DATABASE

The National Land Cover Database (NLCD) is a comprehensive database that provides land classifications across the United States. These classifications were determined by categorizing data acquired by the Landsat missions, with the first NLCD released in 1992 (Homer et al. 2015). Since the initial release, there have been three updates: 2001, 2006, and 2011. The NLCD decision tree classifies every 30-m pixel across the United States into one of eight possible primary categories or among the 20 subcategories available, ranging from known water areas, to wetlands, various modes of agriculture, and various degrees of urbanization. Urban areas have similar spectral signatures (lower NDVI, higher LSTs) to that of damaged areas, and any damage to vegetation scattered throughout an urban area would occur at spatial scales not detected by MODIS. Areas of water result in small or negative values of NDVI and lower LST values owing to a higher specific heat capacity. In order to prevent pixels of urban areas and areas of water from being included, a mask was created from the NLCD to omit urban and water pixels from damage detection attempts.

b. Methods

1) NDVI CHANGE THRESHOLDING

Previous studies focused on identifying areas of damage by identifying NDVI change from a pre-event composite and post-event image (Parker et al. 2005; Jedlovec et al. 2006; Gallo et al. 2012; Molthan et al. 2013). In this study, an NDVI change product (threshold hereafter) is developed from the methods described by Molthan et al. (2013) and Gallo et al. (2012). The pre-event composite is developed from a 14-day maximum of NDVI across the study region, as a 2-week period allows for multiple cloud-free views (Fig 2a). For example, a pre-event composite valid for a given event day represents the cloud-free maximum NDVI for that day and the prior 13 days, resulting in a uniform, pre-event composite to help improve the detection of decreased NDVI within potential damage areas.

Post-event single-day NDVI images were calculated from the MYD09GQ product beginning the day after an event (Fig. 2b). The post-event single-day NDVI (Fig. 2c) was then differenced against the pre-event composite (Fig. 2d). Although the cloud-free portion of a damage swath is visible, lingering cloud

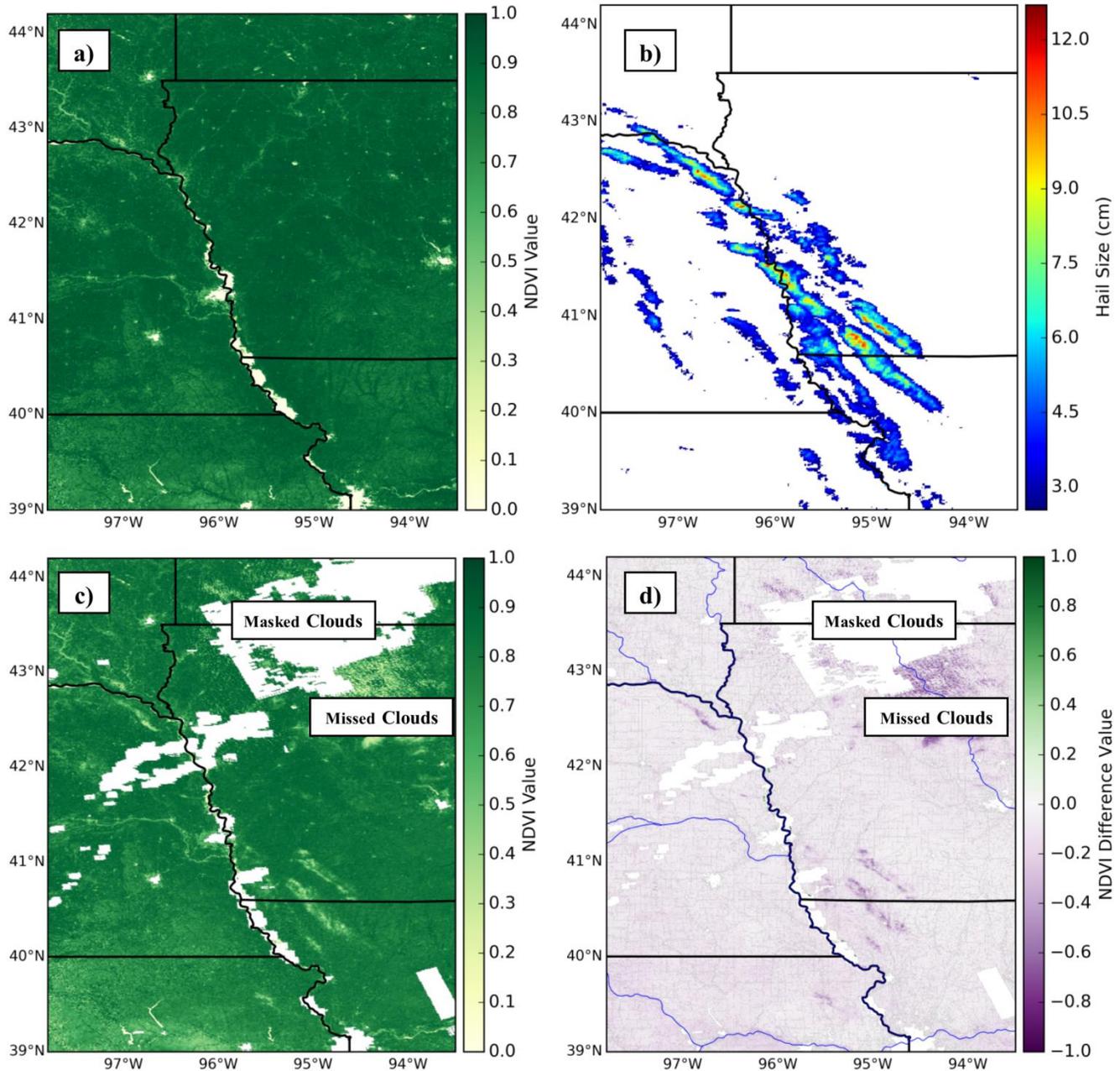


Figure 2. a) 14-day maximum NDVI composite from 5 August 2011 to 18 August 2011. b) Hail sizes from MESH over the period of 1200 UTC 18 August 2011 to 1200 UTC 19 August 2011. c) Single-day NDVI image from 31 August 2011. d) NDVI threshold product created by the difference of the single-day NDVI image and composite image. In panels c) and d), the white pixels are clouds defined by the mask provided as part of the quality assurance band.

cover may partially obscure a damage swath. To capture the full extent of the damage, difference images were generated daily for the 15 days following an event. These difference images were always created using the same pre-event composite (valid for the day of the event). Gallo et al. (2012) used visual comparisons of NDVI difference values and flagged pixels as

damage with an observed NDVI change of -0.15 or less. This study adopted the -0.15 change in NDVI used by Gallo et al. to classify a threshold product pixel as damaged because both studies shared the same regional focus. However, such a threshold may be regionally or seasonally dependent based on the affected vegetation.

2) ANOMALY DETECTION

Image-processing algorithms allow identification of damaged pixels from single-day imagery versus the generation of lengthy composites and a daily NDVI change product. In this study, an image-processing algorithm was used to identify damaged and undamaged areas by first creating daily anomaly images as the difference between a small subset of pixels and the immediate background, converting the results to grayscale, and then passing them through a segmentation filter known as Otsu's Method (Otsu 1979; Sahoo et al. 1988). Otsu's Method creates a histogram from a grayscale image, and then automatically determines the point at which the histogram is divided into two segments labeled as the "background" and "objects." For this study, the background represents undamaged areas of vegetation, and objects that stand out from the background represent potential damage areas in the vegetation. To improve the appearance of damage areas within the anomaly detection technique, individual single-day NDVI and LST (Figs. 3a–b) anomaly images characterize a pixel value against the local background (Figs. 3c–d). These daily anomaly images are created using two moving kernels: one larger, outer kernel representing the background kernel, and one smaller, inner kernel that is differenced against the background to identify the anomalies. The outer kernel is 50 km × 50 km, and anomalies are calculated as the difference between the inner-kernel pixels and the median of cloud-free observations in the outer kernel. The 12.5 km × 12.5 km inner kernel is large enough to resolve the typical 7–12-km width of previously documented hail swaths (Frisby 1963; Klimowski et al. 1998; Parker et al. 2005) when compared to the surrounding background, but could miss smaller features. Figure 4 shows how both kernels move throughout the image to calculate the anomalies, shifting 12.5 km through each iteration (first west-to-east before shifting north once the iterator reaches the end of the domain). Outer kernels of various sizes (12.5, 25, and 50 km) were examined objectively to determine the appropriate size of the final kernel for this study (Bell 2015). If the outer kernel is too small of an area, the kernel will not allow proper comparison between the inner kernel pixels and adjacent areas. Conversely, if the outer kernel is too large, then the inner kernel would also become too large to resolve potential hail-damage streaks. These kernel sizes were for the MODIS instrument only and may need to be adjusted for other sensors of higher spatial resolution (i.e.,

Landsat-7 or 8, Sentinel 2, and DigitalGlobe Worldview).

Kernels were applied to a single-day image to produce daily NDVI and LST anomalies. NDVI anomalies of –0.3 to 0.0 and LST anomalies of 0–3 K were stretched to byte values of 0–255 to create corresponding grayscale images with a spatial resolution of 250 m (Figs. 5a–b). Values were selected based on previous studies identifying damage areas with a lower NDVI and warming of the land surface. This caused the suspected damage areas to stand out in contrast to darker, undamaged pixels (background). Otsu's Method was then performed separately on the NDVI and LST anomaly images to create a set of two binary images, one for suspected damage identified by NDVI anomalies (Fig. 5c) and another for suspected damage identified by LST anomalies (Fig. 5d). Pixels are flagged as damage when identified in both the NDVI and LST anomaly outputs (Fig. 6).

3) TECHNIQUE VALIDATION

Unlike tornado damage surveys, formal ground surveys by the National Weather Service (NWS) are not conducted following widespread known or suspected hail damage. In order to evaluate the performance of the threshold or anomaly detection techniques, post-event satellite imagery was used to identify areas of likely hail damage for each case study. These analyses were created subjectively by identifying damaged areas using MODIS true-color and false-color composites, SPC storm reports of hail, and NSSL MESH data. True-color and false-color composites (Fig. 7) were analyzed for the clearest day observed from 11 to 15 days following a given event. Potential damage areas were identified and outlined using a polygon shapefile in ArcGIS. Parker et al. (2005) also used a geographic information system platform to create a satellite-analyzed dataset to outline the swaths of hail damage. This 11–15-day period encapsulates the post-storm period when crop owners will determine if crops will be able to recover or should be salvaged (R. Connelly 2015, personal communication). As there are other ways of changing NDVI over short time-periods (e.g., harvest, burning, construction, etc.), hail fall was confirmed in the potential damage areas by consulting the NSSL MESH for sizes meeting or exceeding the 2.54 cm (1 in) NWS hail diameter criterion for a severe thunderstorm warning (Ferree 2016). A 5-km buffer was added around all MESH values >2.54 cm. (1 in) to account for any

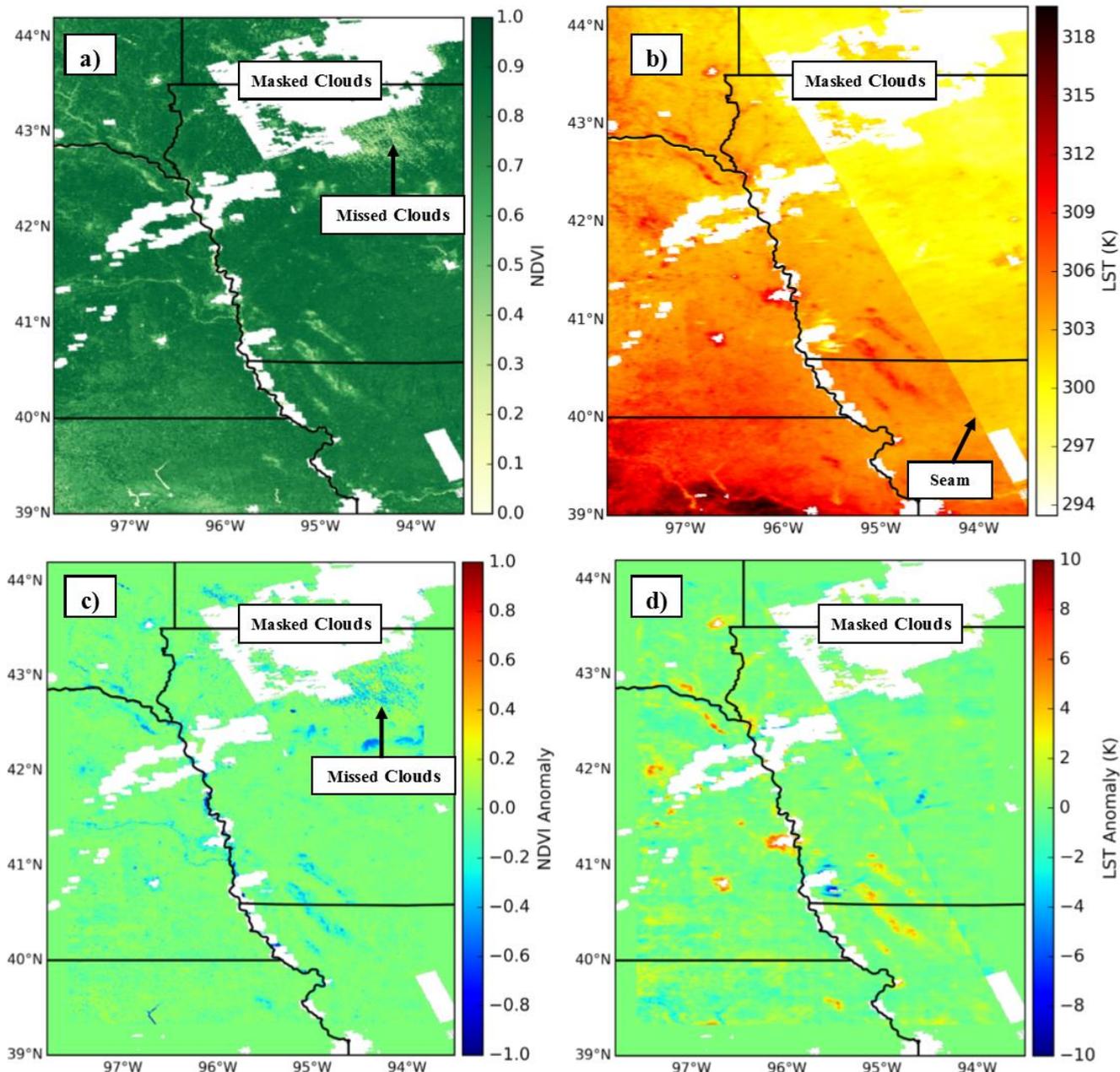


Figure 3. a) Single-day NDVI image for 31 August 2011. b) Single-day LST image for 31 August 2011. The seam was created because of two different observation periods. c) NDVI anomaly image for 31 August 2011. d) LST anomaly image for 31 August 2011. The difference in LST from two different views from Aqua MODIS is apparent as a seam oriented from northwest to southeast in the anomaly calculation. Positive LST anomalies in northeastern Kansas were from an event that occurred on 19 August 2011.

possible radar sampling and displacement errors (Fig. 8) noted in the previous studies by Gallo et al. (2012) and Molthan et al. (2013). The displacement errors likely increased the damage detected downwind of the MESH swaths, but likely increased false detections upwind because the 5-km buffer was uniform around all areas where MESH was >2.54 cm (1 in). Each polygon shapefile was then converted to a 250-m reso-

lution raster to match the spatial resolution of the other datasets.

The MESH and NLCD mask datasets were used as constraints in the calculation of skill scores to limit analyses to vegetated areas that likely experienced severe hail. Only pixels that were observed to have overlapped with MESH values ≥ 2.54 cm (1 in) and were not classified as urban or water were considered

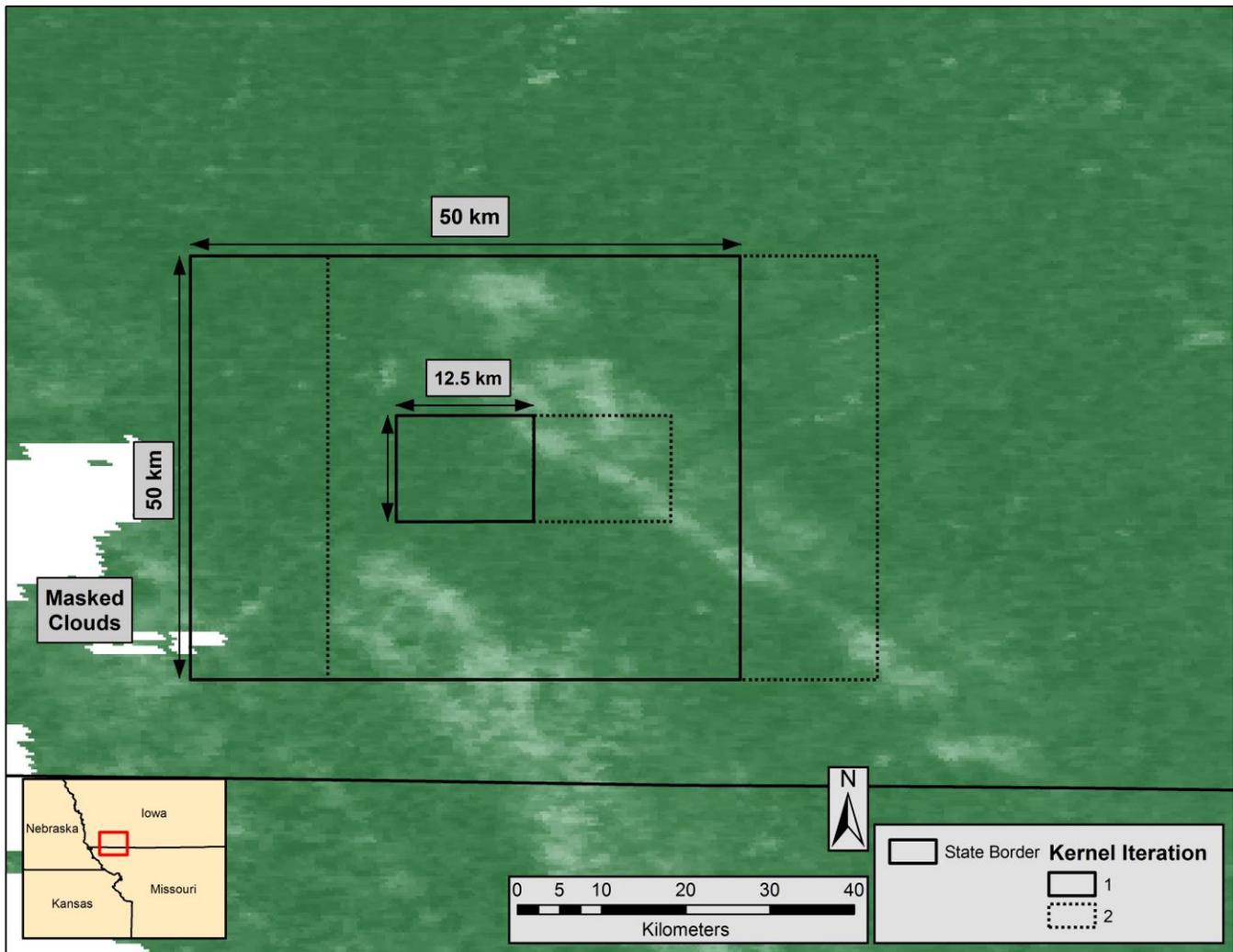


Figure 4. Diagram showing how the kernel filtering moves through a series of iterations on an NDVI image from 31 August 2011. The kernels are projected to the projection of the satellite imagery.

in the skill-scoring process. Several skill scores were calculated to quantify the performance of the two techniques and evaluate their effectiveness in an NRT setting. Imagery for up to 15 days after an event was broken into three 5-day periods following each event to find the highest quality observations (cloud-free, preferred satellite viewing angle) over the area of interest. The clearest day from each of the three 5-day periods was selected in assessing skill scores (Table 1) for both techniques. Definitions of a hit, miss, false alarm, and correct null were determined as follows and summarized in Fig. 9. A pixel was determined to be a hit if it was identified by the technique and inside the boundaries of the satellite-analyzed surveys. A pixel was classified as a miss if it was inside the satellite-analyzed surveys, but was not identified by the technique. A pixel was determined to be a false alarm if it

was identified by a technique but was not within the boundaries of the satellite-analyzed surveys and was identified to have experienced MESH values ≥ 2.54 cm (1 in). The raw NSSL MESH > 2.54 cm (1 in) was used to simulate what datasets would be available in an NRT setting and what potential problems may arise with using the raw MESH values.

Table 1. Validation periods, names, and days following an event that comprise each period.

Time Period Name	Days After Event
Short-term	1–5 days
Mid-term	6–10 days
Long-term	11–15 days

Skill scores included the probability of detection (POD), false alarm ratio (FAR), and critical success

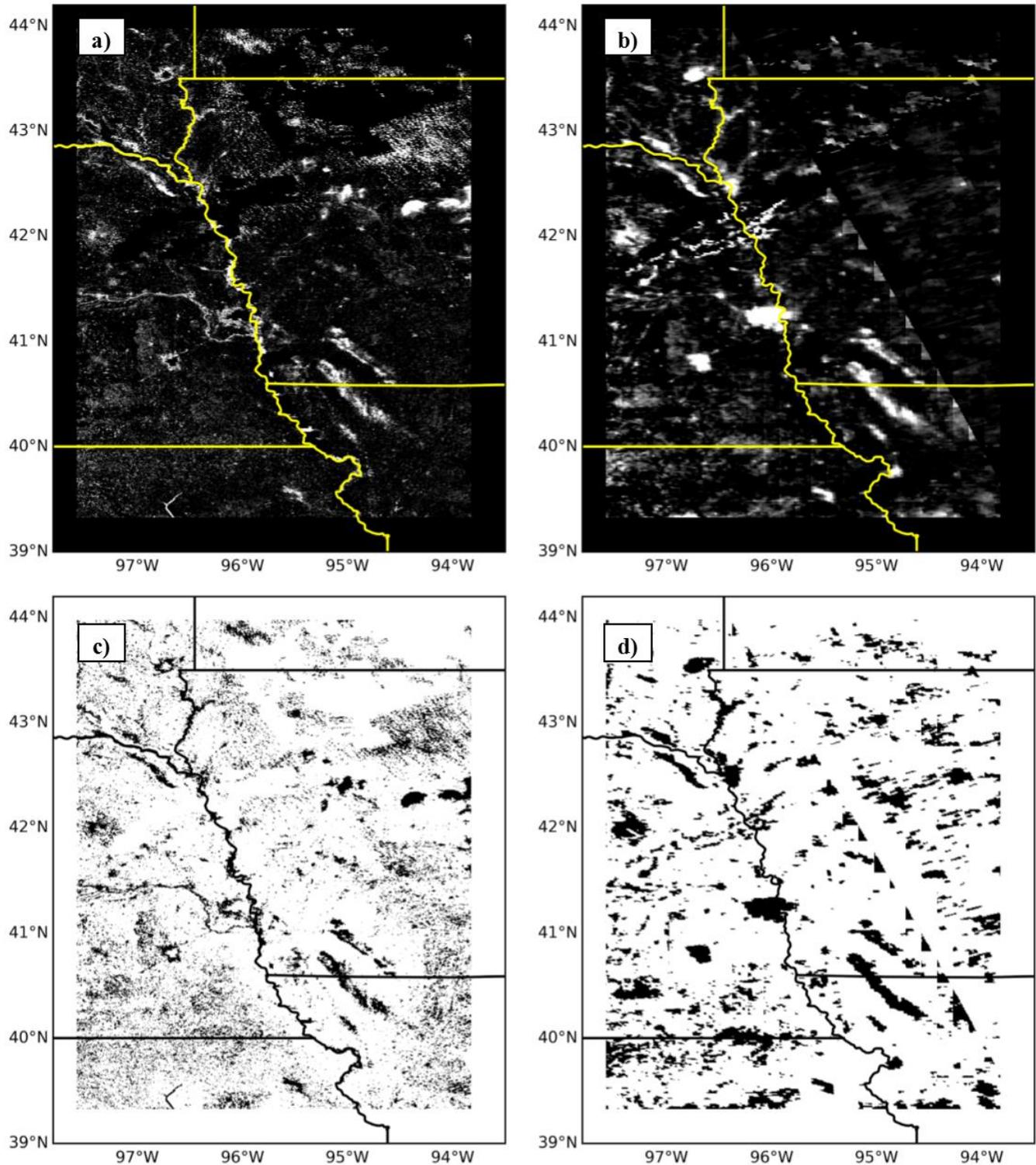


Figure 5. a) Grayscale image for NDVI anomaly image. b) Grayscale image for LST anomaly image. The seam shown in Fig. 2 is erroneously detected when calculating local anomalies and converting to grayscale. c) The classifier (background white, damage black) as a result of anomaly detection for the grayscale NDVI anomaly image. d) Same as c) but for grayscale LST anomaly, with the seam present. The damaged areas are consistent with the damaged areas in the NDVI imagery (Fig. 3a).

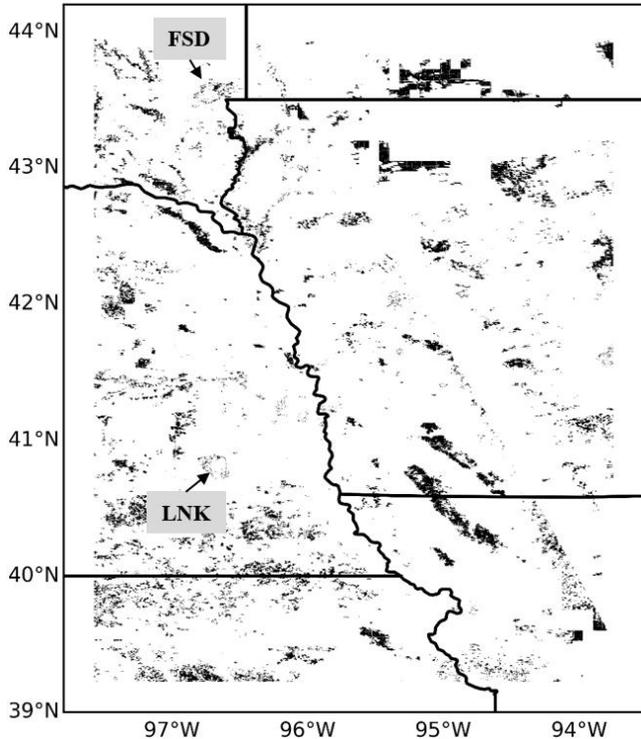


Figure 6. Final anomaly detection product as the agreement among anomalies produced by the NDVI and LST anomaly processing. The anomalies detected around Lincoln, NE (LNK), and Sioux Falls, SD (FSD), are false anomalies (low NDVI, high LST) surrounding urban areas that are masked out by the land-cover product.

index (CSI) following the techniques of Schaefer (1990):

$$POD = \frac{Hits}{Hits+Misses} \quad (2)$$

$$FAR = \frac{False\ Alarms}{Hits+False\ Alarms} \quad (3)$$

$$CSI = \frac{Hits}{Hits+Misses+False\ Alarms} \quad (4)$$

POD (2) describes the ratio of correct hits to the total number of hits and misses, or how well the individual techniques are predicting damage within the satellite-analyzed surveys. FAR (3) is the ratio of false alarms to the total number of hits and false alarms and ensures that each technique is not misidentifying damage pixels outside of the satellite-analyzed survey. CSI (4) is a verification score that compares the number of hits to the total number of hits, misses, and false alarms detected by the techniques. The ideal scenario would be a POD of 1.0 to indicate that the two techniques correctly identified pixels as damage in a perfect match to

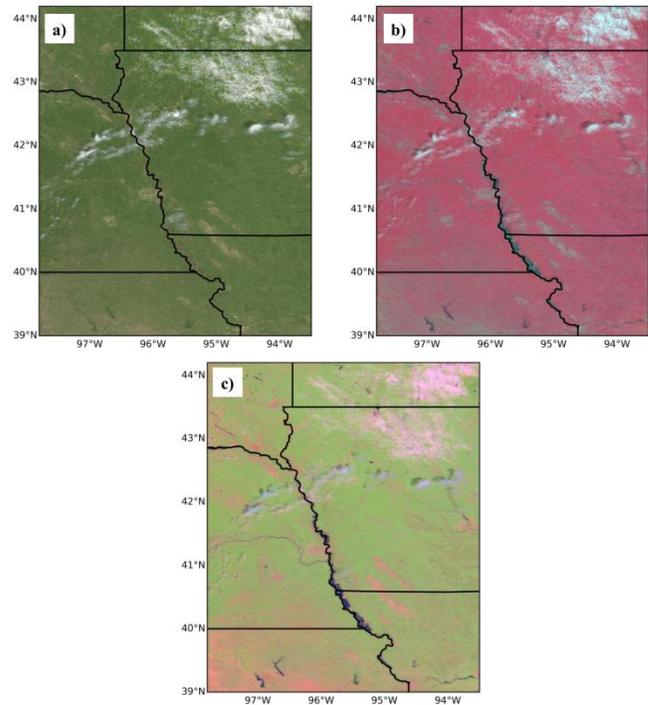


Figure 7. a) True-color red-green-blue (RGB) composite. b) Color infrared RGB composite. c) False-color RGB composite. All three of these RGB composites were used to create satellite-analyzed surveys for validation in each case study.

the satellite-analyzed surveys with no false alarms or misses. However, this ideal scenario is not likely given the nature of the satellite-analyzed surveys. These satellite-analyzed surveys may over-sample areas of potential damage, or they may miss areas of potential damage from case to case. Overall, an ideal algorithm will achieve a high POD while also achieving a low FAR, resulting in a relatively high CSI.

3. Analysis and discussion

The two techniques were each tested on three case studies to evaluate their effectiveness for an NRT application. All case studies occurred in agricultural areas of the Midwest. Two of the case studies were repeated from the analysis of events documented by Gallo et al. (2012) and Molthan et al. (2013).

a. 18 August 2011

During the late afternoon of 18 August 2011, several supercells developed in southeastern South Dakota and propagated southeastward into Nebraska, Iowa, and eventually into Missouri during the overnight hours. The SPC 1630 UTC convective outlook (www.spc.noaa.gov/products/outlook/) had this region

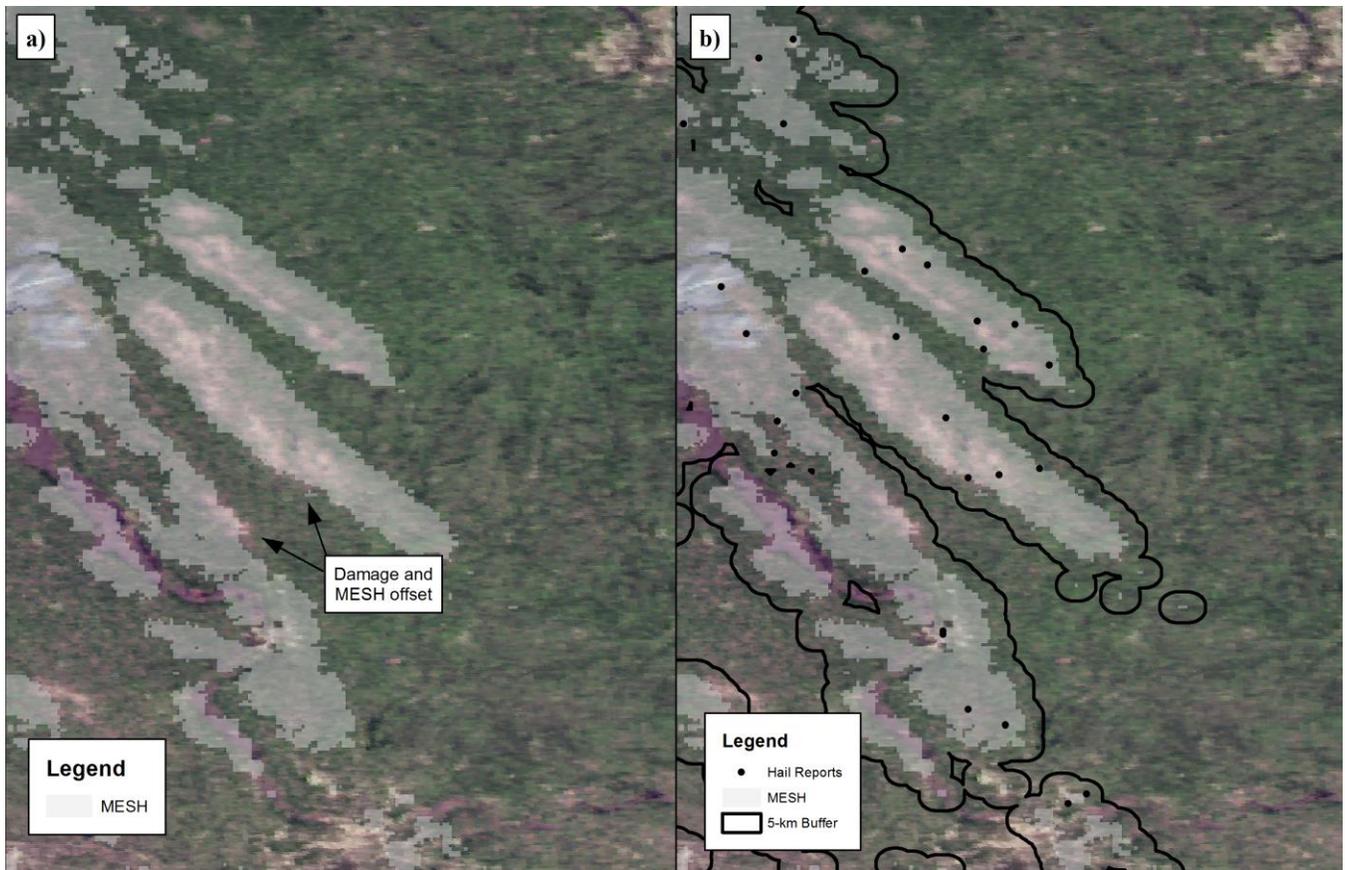


Figure 8. a) True-color RGB with MESH ≥ 2.54 cm (1 in) masked with partial transparency. Areas of damage that are offset from the MESH are annotated. b) Same image as a) but with 5-km MESH buffer outlined in black.

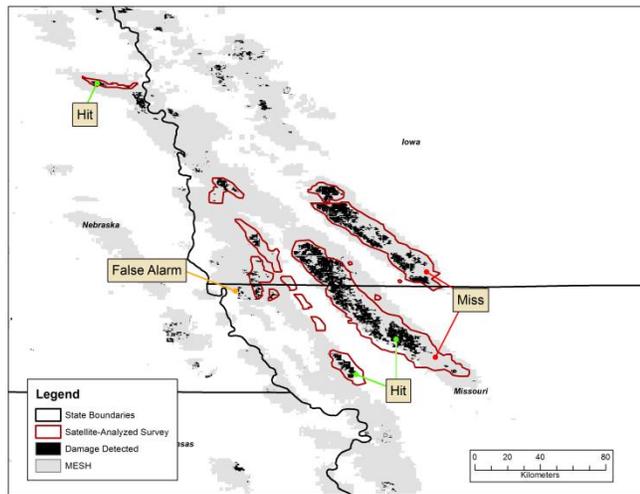


Figure 9. Diagram to visually represent how components of the skill scores were determined. Representative examples of various categories are shown. Damaged pixels (black) represent damage detected by a hypothetical technique. The data for this example were taken from the 18 August 2011 case referenced in the text.

under a slight risk with the highest probability of specific severe weather being damaging winds [30% like-

lihood of ≥ 50 kt (58 mph)] with a slightly lower probability of severe hail [15% ≥ 2.54 cm (1 in)]. For this event, the SPC storm reports received nearly 200 reports of damaging winds and >150 reports of severe hail. There were 19 significant hail reports [≥ 5.08 cm (2 in)] with the largest report of 10.8 cm (4.25 in) occurring southwest of Yankton, South Dakota.

Vegetation damage was apparent in satellite imagery the next day (19 August 2011) with damaged areas prevalent in southwestern Iowa, northwestern Missouri, southeastern South Dakota, and northeastern Nebraska. Molthan et al. (2013) used higher resolution satellite datasets to document the change, in addition to looking at MODIS NDVI change. The threshold technique showed an area of small, negative NDVI changes in the same areas where the damage was visible. When evaluated with the satellite-analyzed surveys, the threshold technique had a POD of 0.284, and the anomaly detection technique had a POD of 0.298. Both methods produced slightly higher FARs (Table 2) for the short-term validation period, but these FARs decreased for the mid-term validation

Table 2. Skill scores for the threshold (THRSH) and anomaly (ANOM) techniques used in the 18 August 2011 case study.

Validation Period	Date	POD		FAR		CSI	
		THRSH	ANOM	THRSH	ANOM	THRSH	ANOM
Short-term	19 August 2011	0.284	0.298	0.373	0.270	0.243	0.287
Mid-term	25 August 2011	0.404	0.338	0.188	0.256	0.369	0.303
Long-term	31 August 2011	0.497	0.317	0.384	0.197	0.380	0.294

period. The POD of both techniques increased, with the threshold technique having the highest POD for the mid-term validation period at 0.404. By the long-term validation period, the highest POD for the thresholding technique was achieved as the vegetation had 13 days to respond to impacts from the damaging hail (survive, wilt, or decay). The true-color red-green-blue (RGB, Figs. 10a–b) highlights well-defined areas of damage present throughout the domain by the long-term validation period. The threshold technique had the highest POD of the two techniques at the long-term validation period with a POD of 0.497; and the POD of the anomaly detection technique was 0.317 (Figs. 10c–d). The FAR of the threshold technique increased from 0.188 during the mid-term period to 0.384 for the long-term period. The FAR increasing in the thresholding technique from the mid-term and late-term validation periods potentially can be attributed to NDVI values of corn and soybeans beginning to drop rapidly by late August (Fig. 1). Because the thresholding technique utilizes pre-event imagery, the change between then and now may meet the threshold, which is used to identify damage. The FAR of the anomaly detection technique dropped from 0.256 in the mid-term validation period to 0.197 during the long-term validation period (Figs. 10e–f). The CSI of the two techniques increased between the short-term (threshold = 0.243 and anomaly = 0.287) and long-term (threshold = 0.380 and anomaly = 0.294) validation periods.

b. 9 August 2009

During the mid-morning hours of 9 August 2009, severe storms with damaging winds and large hail moved eastward across Iowa and resulted in a line of nearly continuous SPC storm reports across the state. Many of these hail reports were <5.08 cm (2 in), but one storm report in central Iowa had hail of 7.62 cm (3 in) in diameter. These storms were responsible for \$169.9 million in damage to agricultural crops (Gallo et al. 2012).

The short-term validation period threshold technique showed swaths of negative NDVI changes in a west-to-east orientation. These same areas lined up with the satellite-analyzed surveys and led to an initial

POD of 0.428, FAR of 0.397, and CSI of 0.334 (Table 3). Most of the false alarms were identified along the MESH swaths east of the satellite-analyzed surveys. The anomaly technique applied in the short-term validation period identified most of the damaged areas but failed to identify some of the western extent. These missing damaged areas led to a POD of 0.433 and CSI of 0.353 for the anomaly technique. The anomaly technique had a lower FAR than the threshold technique at 0.344, indicating that this technique was not picking up on many damaged pixels outside the analyzed survey areas.

This event did not have a mid-term validation period as days 6–10 were too cloudy to validate the threshold or anomaly detection techniques. Clearing skies allowed for a long-term validation period with the damaged areas being more apparent than in the short-term event. The long-term, true-color RGB displays the damage in a west-to-east orientation with more definition than during the short-term validation period owing to continued wilting and other processes (Figs. 11a–b). This was confirmed with the POD of the threshold technique increasing to 0.647 and CSI increasing to 0.571. In contrast to the increase in both POD and CSI, the FAR of the threshold technique dropped to 0.170 (Figs. 11c–d). The long-term POD of the anomaly technique also increased to 0.587 as the technique was able to resolve damage on the western edge. The long-term FAR for anomaly detection also dropped to 0.268 (Figs. 11e–f).

c. 3 June 2014

Several rounds of severe storms moved southeastward from south-central South Dakota through the eastern third of Nebraska on 3 June 2014. The storms began around mid-morning with the last round occurring during the evening hours. The local weather service offices received multiple reports of large hail in excess of 10.16 cm (4 in) in diameter. Several damage swaths were seen over the area in the weeks that followed. Although the algorithms applied attempted to detect damage for all of the damage swaths during all validation periods, daily cloud cover restricted observations over portions of two of the swaths. This

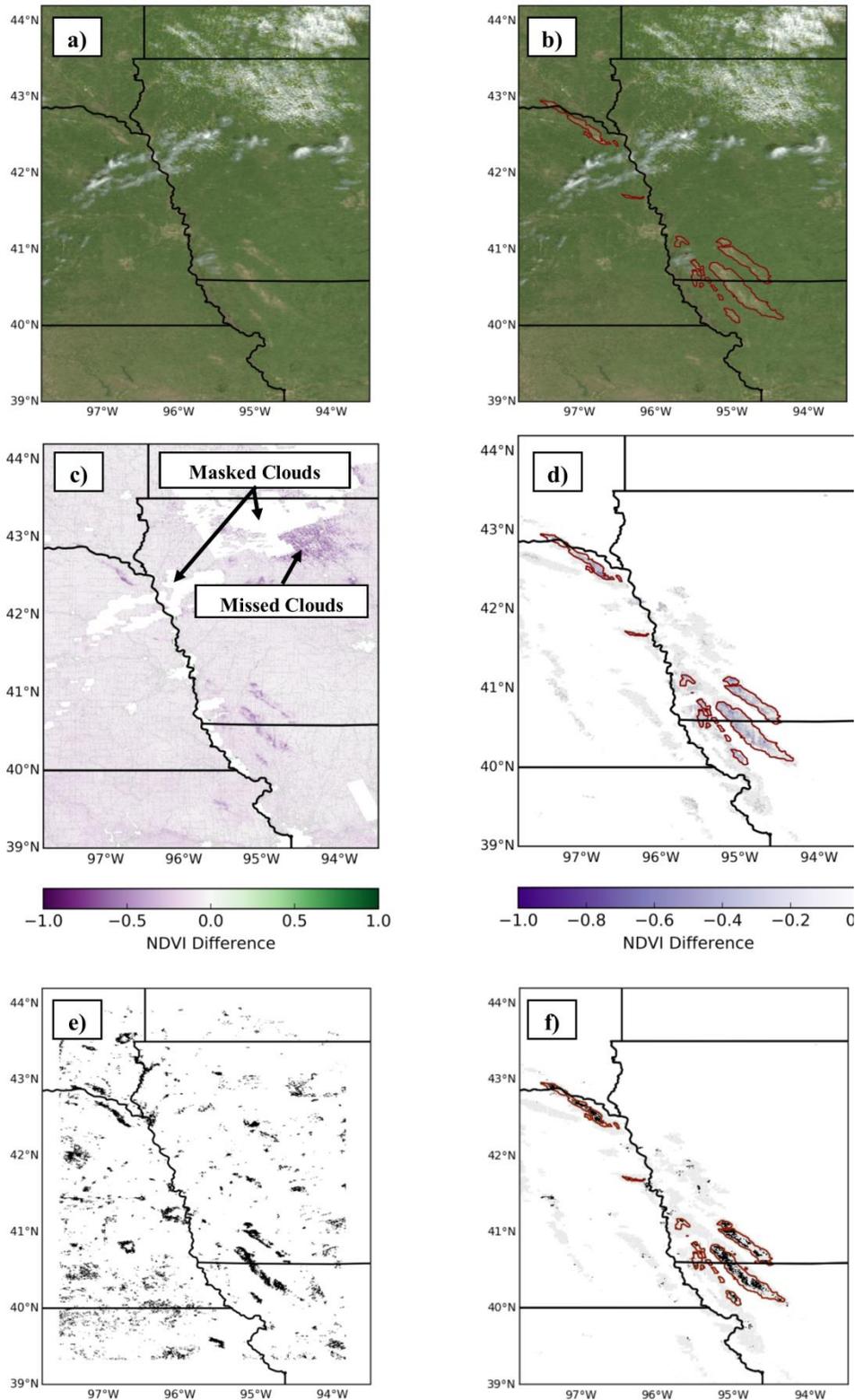


Figure 10. a) True-color RGB from the 31 August 2011 long-term validation period. b) Same as a) but with satellite-analyzed surveys outlined in red. c) Long-term validation period NDVI change threshold product. d) NDVI change threshold product with MESH values ≥ 2.54 (1 in) in grayscale. The satellite-analyzed surveys are shown as an outline. e) Long-term validation period anomaly detection. f) Same as d) but for the anomaly detection product.

Table 3. Same as Table 2 except for the 9 August 2009 case study.

Validation Period	Date	POD		FAR		CSI	
		THRSH	ANOM	THRSH	ANOM	THRSH	ANOM
Short-term	12 August 2009	0.428	0.433	0.397	0.344	0.334	0.353
Mid-term	N/A						
Long-term	24 August 2009	0.647	0.587	0.170	0.268	0.571	0.483

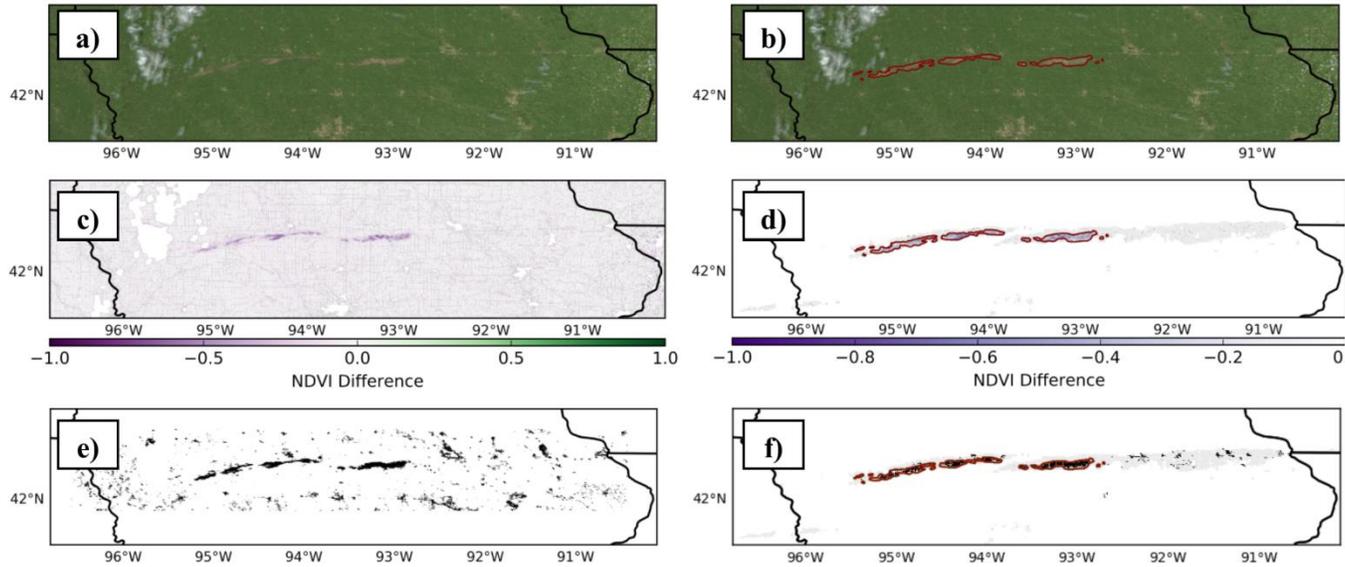


Figure 11. Same as Fig.10 but valid for the long-term validation periods on 24 August 2009.

case study was able to focus on one streak that was observed under mostly clear skies, but only for the mid- and long-term validation periods.

Persistent cloud cover prevented areas of the mid-term validation period from having areas for comparison. In the middle of the domain there were several areas of negative NDVI change. Despite these damaged areas being visible in the threshold technique product, the NDVI change threshold did not pick up on a large number of damaged pixels. When comparing the damaged areas to the satellite-analyzed survey, the threshold technique had a POD of 0.079 and CSI of 0.063 (Table 4). The low CSI resulted from the high FAR (0.758), indicating that the majority of the pixels that were identified by the NDVI change threshold were outside the satellite-analyzed surveys. The anomaly technique performed much better in the mid-term validation period with a higher POD (0.418) and CSI (0.221) and a lower FAR (0.681). Anomaly detection was successful in picking up small NDVI and LST anomalies that were present in the images. The anomaly detection technique identified the majority of damage in the southeastern polygon associated with the analyzed survey, whereas the threshold technique did not.

Like the previous two case studies, damage is apparent in the long-term, true-color RGB for the 3 June 2014 case study (Figs. 12a–b). The damage in south-central Nebraska occurred where radar imagery (not shown) identified a supercell splitting into two damaging storms, resulting in two separate streaks with a small gap between them. In the previous two case studies, vegetation continued to wilt or decay with time, and damage became more apparent, creating a stronger signal for detection. In this case, wilting was less apparent because the vegetation in early June continued to grow, especially in the areas surrounding the damaged vegetation. NDVI values in June increased rapidly between the beginning and end of the month (Fig. 1), so any vegetation that was damaged likely was overtaken by non-damaged vegetation that continued growing. The lack of apparent damage lowered the POD and CSI of the threshold technique to 0.068 and 0.060, respectively (Figs. 10c–d). The long-term FAR for the threshold technique decreased to 0.661. Damaged vegetation in the threshold technique may have been more easily identified by increasing the threshold of NDVI change to a value greater than -0.15 (e.g., -0.10 or -0.05). The FAR and CSI associated with the anomaly detection technique improved between the

Table 3. Same as Table 2 except for the 3 June 2014 case study.

Validation Period	Date	POD		FAR		CSI	
		THRSH	ANOM	THRSH	ANOM	THRSH	ANOM
Short-term	N/A						
Mid-term	13 June 2014	0.079	0.418	0.758	0.681	0.063	0.221
Long-term	17 June 2014	0.068	0.399	0.661	0.539	0.060	0.272

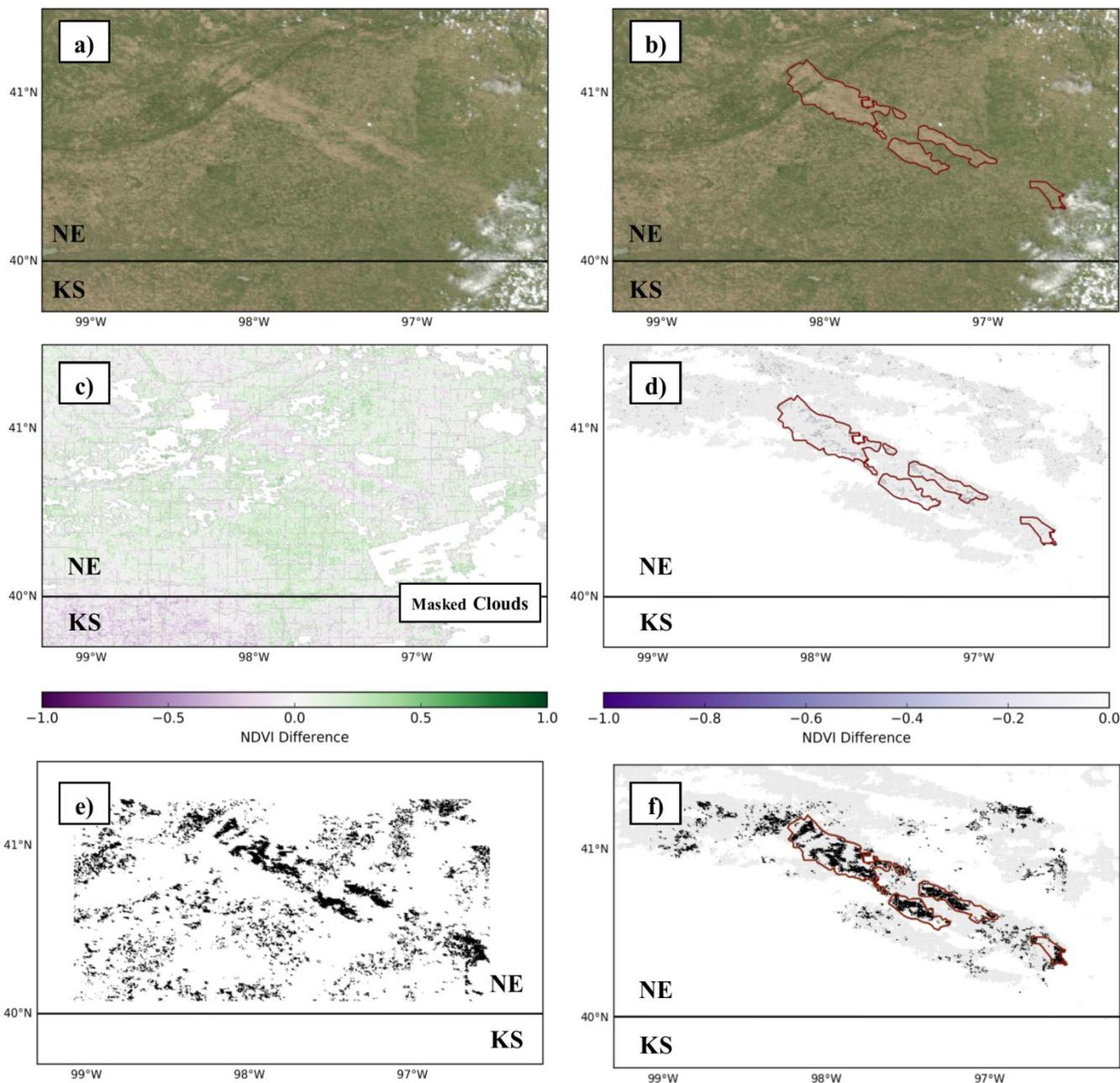


Figure 12. Same as Fig. 10 but valid for the long-term validation period of 17 June 2014.

mid-term and long-term validation period. The FAR decreased to 0.539 and the CSI increased to 0.272. The long-term validation period POD for the anomaly de-

tection technique did decrease to 0.399, but was still much higher than the threshold technique (Figs. 10e-f).

d. Discussion

Three case studies compared two different techniques that could be used in an NRT product for identifying hail-damage swaths. These case studies clarified strengths and limitations of the two techniques—important for determining which to implement in an NRT setting.

The NDVI change threshold technique resulted in the highest POD for damage detection during the 18 August 2011 and 9 August 2009 case studies and the lowest POD for the 3 June 2014 case study. Both the POD and CSI increased between the first available validation period and the long-term validation period for the two case studies that occurred in late summer. The increase in POD and CSI reaffirmed that a stronger NDVI change may appear further in time from an event, caused by the wilt and decay or potential harvesting of the damaged vegetation. The threshold technique did not perform well in the 3 June 2014 case study—an early growing season event. A negative NDVI change was observed, but the threshold of -0.15 made it difficult to identify pixels as damage that would be counted towards validation because the pre-event NDVI was relatively low with young crops still maturing, and damage did not produce an NDVI change at the -0.15 threshold. The -0.15 threshold presents two challenges for early season cases. First, it can be difficult to obtain a large decrease in NDVI when crops are beginning to grow and NDVI is relatively low. Second, crops may be less susceptible to damage early in the season or are able to rebound more easily, resulting in a decrease in the magnitude of NDVI change. Earlier in the season, agricultural areas will not see the same pre- and post-event NDVI change signal as they would in events later in the growing season.

In all three case studies, the anomaly detection technique produced similar or higher POD scores compared to the NDVI change technique. The anomaly detection technique identified areas of damage that were visually apparent in all of the short-term validation periods, including the 3 June 2014 case study in which the mid-term validation period was first. These damaged areas continued to remain present throughout all the remaining validation periods. The POD and CSI scores of the three events increased between the short- and mid-term validation periods and the long-term validation periods, although FAR scores decreased. As with the threshold technique, damage became more apparent with longer time for crops to wilt and decay,

resulting in a larger NDVI change that followed a severe hail event. In the case of the anomaly detection technique, anomalies in the NDVI and LST imagery become more apparent as vegetation wilts and decays, exposing the surface. This allows the final anomaly detection product (combined individual NDVI and LST anomaly products) to identify more objects as damage in the correct areas when evaluated against the satellite-analyzed surveys.

These case studies compared two different techniques that could be used in an NRT product for identifying swaths of hail damage to crops. Based on their performance, the anomaly detection technique would be recommended for an NRT product because of its consistent performance over the entire growing season and the use of single-day imagery, whereas the threshold technique relied on multi-day composites. Both the threshold and anomaly detection techniques performed well with late-season events (i.e., July and August), but anomaly detection performed better than the threshold technique in the early season case (i.e., 3 June 2014). This anomaly detection technique would benefit from better cloud detection, as the MODIS QA dataset missed several areas of clouds and cloud edges, which returned false alarms.

4. Conclusions and future work

Satellite remote sensing has been used to identify areas of damage to vegetation as a result of severe thunderstorms that bring damaging winds and large hail. These areas of damage are especially prevalent across the central United States during the peak growing season. In previous studies (Parker et al. 2005; Gallo et al. 2012), these areas of damage were identified through manual analysis, which is inefficient and time-consuming. This study evaluated two different techniques to identify areas of damage and assessed the strengths and weaknesses of each one in three case studies before determining which one would be best suited for an NRT algorithm.

The first technique looked at short-term changes in NDVI that met a certain threshold, similar to previous studies. Differences were obtained from the pre-event 14-day maximum and a post-event, single-day NDVI image. Pixels were determined to be damaged if they had an NDVI change of ≤ -0.15 . The thresholding technique had a higher POD in the long-term validation period than the anomaly technique in two of the three case studies, but struggled with an early season event. The POD increased in two of the three case

studies, indicating that as time moved further from the actual event, the effects of damage on vegetation were more easily identifiable.

The second technique detected local NDVI and LST anomalies in the imagery by passing the single-day imagery through kernel filters that could detect hail swaths. Damaged areas have a lower (higher) NDVI (LST) when compared to the immediate background. Each of the NDVI and LST anomaly images were converted to grayscale and processed further with Otsu's Method, identifying damaged areas as objects against an undamaged background. This technique performed well in all three case studies, including an early season event during seasonal green-up.

The techniques described herein can be used to identify areas of vegetation damaged by hailstorms instead of performing a manual analysis. Use of NDVI and LST products for damage detection remains limited, and validation of products remains challenging without an official ground validation dataset. Future work continues to address the challenge of validating these techniques against available ground survey datasets. Multiple methodologies exist that can successfully detect damaged areas, and future work and analysis should continue to explore all possible combinations of methodologies, satellite, and non-satellite datasets to create and maintain the best possible detection. Given improvements in the automated detection of crop damage, these efforts may benefit estimates of crop loss for comparisons against reported loss estimates. Expanding this NRT algorithm to the identification of vegetation damage caused by other aspects of severe weather (e.g., damaging winds and tornadoes) could be done by using additional radar products such as azimuthal shear and rotation tracks from NSSL.

Acknowledgments. The authors thank professors Drs. Rob Griffin and Udaysankar Nair from the University of Alabama in Huntsville for their input and guidance on this study. The authors also thank Kiel Ortega at the NSSL Cooperative Institute for Mesoscale Meteorological Studies for input regarding the acquisition of MESH data and Phil Schumacher of NWS Sioux Falls, South Dakota, for sharing information and experience when attempting to document these areas of hail damage. Finally, the authors thank the three anonymous reviewers who provided feedback to improve the quality of this manuscript.

REFERENCES

- Bell, J. R., 2015: The development of a near-real time hail damage swath identification algorithm for vegetation. M. S. thesis, Dept. of Atmospheric Science, University of Alabama in Huntsville, 134 pp.
- Changnon, S. A., D. Changnon, and S. D. Hilberg, 2009: Hailstorms across the nation: An atlas about hail and its damages. Illinois State Water Survey Contract Report 2009-12, 101 pp. [Available online at www.isws.illinois.edu/pubdoc/CR/ISWSCR2009-12.pdf.]
- Ferree, J., cited 2016: National change of the hail criteria for severe storms from $\frac{3}{4}$ inch to 1 inch beginning January 5, 2010. [Available online at www.nws.noaa.gov/oneinchhail/docs/One_Inch_Hail.pdf.]
- Frisby, E. M., 1963: Hailstorms of the upper Great Plains of the United States. *J. Appl. Meteor.*, **2**, 759–766, [Crossref](#).
- Gallo, K., T. Smith, K. Jungbluth, and P. Schumacher, 2012: Hail swaths observed from satellite data and their relation to radar and surface-based observations: A case study from Iowa in 2009. *Wea. Forecasting*, **27**, 796–802, [Crossref](#).
- Homer, C. G., and Coauthors, 2015: Completion of the 2011 National Land Cover Database for the conterminous United States – Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.*, **81**, 345–354. [Available online at www.asprs.org/a/publications/pers/2015journals/PERS_May_2015/HTML/index.html#346/z.]
- Jedlovec, G. J., U. Nair, and S. L. Haines, 2006: Detection of storm damage tracks with EOS data. *Wea. Forecasting*, **21**, 249–267, [Crossref](#).
- Jiang, Z., A. R. Huete, K. Didan, and T. Miura, 2008: Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.*, **112**, 3833–3845, [Crossref](#).
- Justice, C. O., J. R. G. Townshend, E. F. Vermote, E. Masuoka, R. E. Wolfe, N. Saleous, D. P. Roy, and J. T. Morisette, 2002: An overview of MODIS Land data processing and product status. *Remote Sens. Environ.*, **83**, 3–15, [Crossref](#).
- Klimowski, B. A., M. R. Hjelmfelt, M. J. Bunkers, D. Sedlacek, and L. R. Johnson, 1998: Hailstorm damage observed from the GOES-8 satellite: The 5–6 July 1996 Butte–Meade storm. *Mon. Wea. Rev.*, **126**, 831–834, [Crossref](#).
- Kriegler, F. J., W. A. Malila, R. F. Nalepka, and W. Richardson, 1969: Preprocessing transformations and their effects on multispectral recognition. *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, pp. 97–131. [Available online at adsabs.harvard.edu/abs/1969rse.conf...97K.]

- Molthan, A. L., J. E. Burks, K. M. McGrath, and F. J. LaFontaine, 2013: Multi-sensor examination of hail damage swaths for near real-time applications and assessment. *J. Operational Meteor.*, **1** (13), 144–156, [Crossref](#).
- Ortega, K. L., T. M. Smith, K. L. Manross, K. A. Scharfenberg, A. Witt, A. G. Kolodziej, and J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530, [Crossref](#).
- Otsu, N., 1979: A threshold selection method from gray-level histograms. *IEEE Trans. Syst., Man, Cybern.*, **SMC-9**, 62–66. [Available online at engineering.purdue.edu/kak/computervision/ECE661.0/OTSU_paper.pdf.]
- Parker, M. D., I. C. Ratcliffe, and G. M. Henebry, 2005: The July 2003 Dakota hailswaths: Creation, characteristics, and possible impacts. *Mon. Wea. Rev.*, **133**, 1241–1260, [Crossref](#).
- Roy, D. P., J. S. Borak, S. Devadiga, R. E. Wolfe, M. Zheng, and J. Descloitres, 2002: The MODIS Land product quality assessment approach. *Remote Sens. Environ.*, **83**, 62–76, [Crossref](#).
- Sahoo, P. K., S. Soltani, and A. K. C. Wong, 1988: A survey of thresholding techniques. *Comput. Vision Graph.*, **41**, 233–260, [Crossref](#).
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, [Crossref](#).
- Segele, Z. T., D. J. Stensrud, I. C. Ratcliffe, and G. M. Henebry, 2005: Influence of a hailstreak on boundary layer evolution. *Mon. Wea. Rev.*, **133**, 942–960, [Crossref](#).
- SPC, cited 2016: Storm Prediction Center WCM page. [Available online at www.spc.noaa.gov/wcm/.]
- Stumpf, G. J., T. M. Smith, and J. Hocker, 2004: New hail diagnostic parameters derived [sic] by integrating multiple radars and multiple sensors. Preprints, *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., P7.8. [Available online at ams.confex.com/ams/pdfpapers/81451.pdf.]
- Vermote, E. F., and A. Vermeulen, 1999: Atmospheric correction algorithm: Spectral reflectances (MOD09). Algorithm Technical Background Document, version 4, 107 pp. [Available online at modis.gsfc.nasa.gov/data/atbd/atbd_mod08.pdf.]
- Wardlow, B. D., and S. L. Egbert, 2008: Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sens. Environ.*, **112**, 1096–1116, [Crossref](#).
- Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, [Crossref](#).