# SUGARCANE APHID MOVEMENT MODEL FOR OKLAHOMA, WEATHER AND GEOGRAPHIC EFFECTS ON SUGARCANE APHID MIGRATION IN OKLAHOMA

By

SEOKIL LEE

Bachelor of Science in Agricultural Economics
Chungbuk National University
CheongJu, Korea
2013

Master of Science in Agricultural Economics
Chungbuk National University
CheongJu, Korea
2016

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2022

SUGARCANE APHID MOVEMENT MODEL FOR OKLAHOMA, WEATHER AND GEOGRAPHIC EFFECTS ON SUGARCANE APHID MIGRATION IN OKLAHOMA

Dissertation Approved:

Dr. Dayton M. Lambert

Dissertation Adviser

Dr. Jefferey Vitale

Dr. B. Wade Brorsen

Dr. Kristopher L. Giles

Name: SEOKIL LEE

Date of Degree: JULY, 2022

Title of Study: SUGARCANE APHID MOVEMENT MODEL FOR OKLAHOMA, WEATHER AND GEOGRAPHIC EFFECTS ON SUGARCANE APHID MIGRATION IN OKLAHOMA

Major Field: AGRICULTURAL ECONOMICS

Abstract: The first paper discusses developing a predictive model of SCA intrusion trajectories for Oklahoma sorghum producers. The SCA survival rates and migration distribution are generated using the temperature, precipitation, wind direction, and wind speed. A clustering algorithm was used to generate groups of low and high infestation probabilities, and the fields that could be infected. Each day's initial and movement probability depends entirely on the wind direction recorded on that day. However, as time passes, the importance of wind direction diminishes. The probability distributions have advantages in terms of data availability and cost, unlike multispectral and visual reporting. The SCA survival rates and migration distribution could provide a framework for future research regarding SCA survivability and migration modeling.

The second paper measures the effects of weather, geographic, and biological characteristics of SCA on the infestation of sorghum fields in Oklahoma. Infestation likelihood curves were used to measure the single effect of each covariate. Infestation likelihood curves suggest there is a difference in the infestation probability between the high- and low-density wing/un-winged SCA groups. The larger the ratio of winged to un-winged SCA, the more likely their movement to other sorghum fields. The PHM results reaffirmed that weather is an important factor in determining infestation hazard. Field location and the distance between sampled fields was not a significant factor. The total population of SCA per plant in the most recently infested field was negatively correlated with infestation hazard. The infestation curve and PHM results provide field-level information about more detailed infestation hazard. These aids may improve field-level decision-making for pest control planning, such as coordination of pesticide use and harvest timing.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

# CHAPTER I

# SUGARCANE APHID MOVEMENT MODEL FOR OKLAHOMA

## Introduction

Sugarcane aphid (SCA), *Melanaphis sacchari* (Zehntner) (Hemiptera: Aphididae), is an economic pest found throughout Asia, Africa, Australia and South America with sorghum, sweet sorghum, some millet varieties, and Johnsongrass as its main hosts (Singh et al., 2004). SCA was introduced in the 1970s into the United States (US), but at the time it was considered only a pest of sugar cane. However, SCA infestations continue to this day, as evidenced by the economic damage to sorghum by SCA in 2013 (Nibouche et al., 2018).

The largest producers of sorghum in the US are Kansas (5,923,280 metric tons) and Texas (1,577,340 metric tons) (Oklahoma is 304,800 metric tons) (USDA NASS, 2020). SCA was first observed in the Texas Gulf Coast and Louisiana in 2013 (Knutson et al., 2016a). SCA was found in Louisiana, Oklahoma, and Mississippi in late 2013, and in more than 400 counties in 17 states in 2015, accounting for 97% of U.S. sorghum cropland (7,405,000 acres) and affecting 98% of production (14,230,426 metric tons) (Bowling et al., 2015).

A recent SCA study focused on the economic damage of SCA in Kansas and Texas. Zapata et al. (2018) measured the change in profit due to increased pest monitoring and control costs caused by SCA infestation of sorghum in the Rio Grande Valley (RGV), Texas, 2014 and 2015. Between 2014 and 2015, average losses due to SCA infestation was $64.54 per acre, and losses in revenue were $49.56 per acre. Specifically, in the RGV region in 2014, there were losses of labor income ($27.08 million), losses in value added to the economy ($31.7 million), losses of production ($38.78 million), and a loss of 103 jobs.

Information on the location and range of SCA migration pathways is crucial for sorghum producers in their effort to manage SCA. Multispectral imagery (MI) and infestation maps over time provide information on pest infestation severity and the route of sugarcane aphid reported ‘on-the-ground’ by producers (Texas A&M AgriLife Extension, 2020). However, with MI it is difficult to detect the hourly or even daily movement of SCA. On-the-ground tracking by producers facilitates confirmation of SCA infestation through visual inspection. Integrating MI with ground reconnaissance is effective for documenting the location and intensity of SCA infestation, but forecasting where SCA move next remains a difficult task. Spatial gaps in SCA infestation also contribute to the challenge of predicting where SCA infestation will go and whether producers are able to document visually infestation events.

Predicted SCA migration paths may provide an early warning system for sorghum producers. Information from accurate forecasts could help minimize the economic loss incurred due to subsequent SCA infestations. Related detection systems focus on the mountain pine beetle (*Dendroctonus ponderosae*). Wulder et al. (2006) estimated the attack probability of the mountain pine beetle in Western Montana. Their study found that infestation was dependent on terrain and weather data, but the model did not consider the pests' travel routes or other pertinent information. Periodic information, such as SCA infestation status and weather forecasts, would enhance SCA infestation models because these variables are important for predicting field infestation status and trajectory.

The purpose of this study is to develop a SCA movement model for Oklahoma sorghum producers using the State's Mesonet weather information system (Mesonet, 2020). Previous studies tracking the movement of pests using weather data and organismic traits are numerous (Acreman & Dixon, 1989; Angilletta, Jr., & Dunham, 2003; Angilletta, 2004; Asin & Pons, 2001; Auad et al., 2009; Bale et al., 2002; Boate & Otayor, 2020; Chen et al., 2019; Huberty & Denno, 2004; Kobori & Amano, 2003; Michael, 2019; Souza & Davis, 2020). Other studies tracked insect movement and distance as a function of wind speed and direction (Mann et al., 1995; Rodríguez-del-Bosque et al., 2020; Shao et al., 2020). It is reasonable to assume that the spread of SCA, like the other small insects analyzed in these previous studies, is strongly dependent on weather.

In 2018, there were 17 reports of SCA over 71 days in Oklahoma, suggesting that SCA populations are likely to remain in infested fields after migration from other fields (June 15 through August 25, 14 counties). Prediction of SCA movement using only weather data is possible, but the accuracy of predicted paths diminishes as time after sighting passes because previous periods cannot be used. Estimating the predicted paths of SCA could help producers develop management plans, but a more important issue for producers is determining if, or when, a field will be infested. SCA do not recognize county borders. The closer fields are to each other, the greater the influence of the weather on aphid movement between locations. This means that the field level, rather than the county level, is the most desirable observational unit if that data is available. In addition, long-term patterns of temperature and precipitation, which are major components of weather, determine the weather of a region (Kukal & Irmak, 2016). That is, in neighboring regions, weather variables have similar values and exhibit similar temporal patterns. In addition, because SCA is an organism, the survival and migration of populations are affected by real-time and historical weather variables. This means that 1) the temporal patterns of weather variables should be considered in an analysis of SCA transmission, and 2) time-accumulated paths of migration routes are required for accurate forecasting.

This chapter formulates a probability distribution model for weather variables that affect SCA survival and migration direction. No previous literature links weather variables, including precipitation, wind direction, and temperature, to SCA migration. Thus, while the migration probabilities generated with this model are theoretical and based on relationships reported in current literature, the assumption that any migration of SCA is solely driven by these variables is reasonable. Migration probability distributions are used to forecast the movement of SCA from one to many fields on a daily time step. The types of migration probability distributions available to conduct this analysis are limited by the lack of previous research pertaining to SCA survivability and movement as a function of weather variables. Yet, the distributions formulated here retain the usual properties of probabilities. The probabilities are positive, sum to one, and are bound between zero and one. The migration probabilities generated by linking the biological properties of SCA to weather variables can assist sorghum producers in the formulation or modification of production and pest management plans. In addition, easier access to weather information from public resources such as Oklahoma's MESONET is a cost-effective alternative to multispectral imaging.

## Data

The data are daily Mesonet observations of the average air temperature (TAVG), rainfall (RAIN), average wind speed (WSPD), and primary wind direction (PDIR) observed in Oklahoma from 15 to 25 June 2013 to 2020. Weather variables after June 15 in 2018, when SCA was first reported in Oklahoma, showed similar trends for nearby counties (Figures 19-21 in Appendix for the distribution of each weather variable by date).

TAVG is the average of all 5-minute-averaged temperature observations each day (degrees Fahrenheit). RAIN is liquid precipitation measured each day (inches). WSPD is the average of all 5-minute wind speed observations each day (miles per hour), PDIR is most common wind direction for a day and it is based on 16-point compass heading with a 16-point cardinal direction (Mesonet, 2020). These variables were obtained from Mesonet (2020)'s daily records and USDA NASS (2020). Data was collected from weather stations located in each of Oklahoma's 77 counties. Some counties have more than one weather station. In total, there are 119 active weather stations across Oklahoma (Table 1).

<< Table 1>>

Oklahoma's sorghum production is concentrated in the northwest region of the state (Figure 1, see Figure 17 in the Appendix for location and size of each field). Observational data for SCA is available for Kansas, Oklahoma, and Texas (EDDMapS, 2020). Collection of these data began in 2013. The data include the date and county where SCA infestations were documented on fields by observers (EDDMapS, 2020).

<<Figure 1>>

SCA can live for up to one month, but typically have a 10-day lifespan (Knutson et al., 2016b). Therefore, the weather for 10 days after SCA were observed on a field most likely influences their survival and movement. Table 2 summarizes the statistics of the counties where sugarcane aphid were reported in 2018, along with average of weather variables for 10 days from the reporting date. In Grady, Washita, and Noble County, there are multiple weather stations, and so observations on the weather variables were averaged over their respective number of weather stations. Weather stations without observations are excluded.

<< Table 2>>

## Methods and Procedures

Daily observations of weather variables are used to generate SCA movement probability distributions for each field included in the sample. Figure 2 summarizes the steps used to derive the probabilities. First, each probability distribution is created using the weather variables. At this time, the parameters for each probability distribution are determined using information obtained from studies on the ecological characteristics of SCA (reduction of the SCA population according to temperature and precipitation (Bale et al., 2002; Chen et al., 2019; Huberty & Denno, 2004; Souza & Davis, 2020), mobility according to wind direction, and movement distance due to wind speed (Mann et al., 1995; Rodríguez-del-Bosque et al., 2020; Shao et al., 2020)). Daily movement probabilities of SCA are calculated by combining the probability distributions associated with each weather variable. After that, a cumulative probability of daily movement over time is calculated. Spatial clustering procedures are used to group probabilities most similar to each other across fields. Lastly, the field group to which the SCA will most likely move to is classified using the migration probabilities (Figure 2).

<< Figure 2>>

### *Parametric Probability Distributions for Weather Variables*

A parametric probability distribution for each weather variable is developed to predict the SCA movement as a function of weather variables. Rainfall and temperature variables are related to the survival rate of SCA. Wind direction and wind speed are related to the probability of SCA movement. That is, rainfall and temperature affect SCA populations in a given field, which in turn affects the probability that the SCA will move to another sorghum field. On the other hand, wind direction and speed determine the movement direction and distance of the SCA.

### *Probability Distribution for SCA Survival Based on Precipitation*

Precipitation affects the survival rate of SCA (RAIN, measured each day in inches). There are several studies on precipitation and its relationship with the survival and behavior of animals (Chen et al., 2019; Huberty & Denno, 2004; Kobori & Amano, 2003; Ukoroije & Abalis, 2020). SCA is a small insect, and precipitation greatly influences SCA survival rates and movement. As the precipitation increases, the number of SCA will decrease. Using this information, the distribution of precipitation to proxy SCA survivability is

(1) $$P_{rain} = \exp(-\beta \cdot RAIN)$$

where $P_{rain}$ is the probability of survival as a function of rainfall and $\beta$ is positive parameter. When rainfall is zero, the SCA survival rate is 100 percent. As rainfall increases, the survival rate decreases exponentially. As the size of the parameter $\beta$

increases, the survival rate of SCA increases (left, Figure 3). There is no experimental data on a reasonable value for $\beta$, which ostensibly links SCA survival to rainfall. Future research would need to relate Eq. 1 to empirical data on SCA survival and precipitation. As applied here, $\beta$ was chosen such that the resulting distribution corresponded with the observed average rainfall for sorghum producing areas. Increasing $\beta$ reflects drier conditions, while decreasing it corresponds with wetter growing conditions.

<< Figure 3>>

*SCA Survival Probabilities and Temperature*

The distribution of SCA survival rates as a function of temperature is generated using TAVG (average of all 5-minute averaged temperature each day, degrees Fahrenheit). There are several studies on the reproduction, survival, and behavior of organisms related to temperature, with each organism exhibiting unique responses to temperature and depending on the its biological characteristics and habits (Acreman & Dixon, 1989; Angilletta & Dunham, 2003; Angilletta, 2004; Asin & Pons, 2001; Auad et al., 2009; Bale et al., 2002; Boate & Otayor, 2020; Michael, 2019; Souza & Davis, 2020). A common finding of these studies is that temperature directly affects the organism's life. However, there are no immediate studies on the correlation between temperature and the probability of SCA survival, although studies on the effects of low temperatures on SCA mortality exist.

Souza et el. (2019) derived the longevity and fecundity of SCA over the 5 to 35°C temperature range. Longevity was highest at 15°C, and fecundity was highest at 25°C. In other words, in the range of 15 to 25°C, SCA exhibit the highest survival rate and fertility. On the other hand, outside of the range of 15 to 25°C, longevity and fecundity decrease. Based on this information, the survival probability distribution has the form of a quadratic function with respect to temperature. For this analysis, the distribution with the greatest survival rate in the range of 15 to 25°C is:

(2) $$P_{TAVG} = \kappa \cdot TAVG + \lambda \cdot TAVG^2$$

where $P_{temp}$ is probability of survival as a function of temperature, $TAVG$ is the average of all daily temperature observations recorded at 5-minute intervals, and $\kappa$, $\lambda$ are parameters ($\kappa$ > 0, $\lambda$ < 0). The parameter value with the largest survival rate in the range of 15 to 25°C was set to $\kappa$ = 0.1, $\lambda$ = -0.0026 (right, Figure 3). The values for these parameters were selected based on findings from Souza & Davis (2020) and Ukoroije & Abalis (2020). The parameters were also chosen such that the median value of the 15 to 25°C range corresponded with the highest probability mass, which is 1 at 20°C.

*SCA Movement Probabilities and Wind Speed*

Several studies analyzed the effects of wind speed on insect movement (Mann et al., 1995; Rodríguez-del-Bosque et al., 2020; Shao et al., 2020). These studies find that wind speed directly affects smaller organisms such as SCA and distance traveled. The greater the wind speed, the higher the probability that SCA migrate over longer distances. However, considering that SCA travel 5 hours per day on average the moving distance probability is maximum at time of five hours. A probability distribution reflecting this relationship is:

(3) $$P_{WSPD} = 1 - \frac{|w \cdot (t-5)|}{5w}$$

where $P_{WSPD}$ is probability of movement according to $WSPD$, $WSPD$ is average of all 5-minute wind speed observations each day (miles per hours), $w$ is wind speed (miles per hours), and $t$ is the time SCA travel per day. The movement probabilities as a function of

WSPD appear in Figure 3 (center panel). Equation (3) predicts that SCA most likely travel 5 hours per day according to the wind speed, and the probability decreases as flight deviates from the expected 5 hours flights. The distance between the initial field infected and other inter-field distances were calculated using the Haversine formula[1].

*SCA Movement Probabilities and Wind Direction*

There are no previous studies documenting the correlation between wind direction and SCA movement, but studies conclude that wind affects SCA travel distance (Mann et al., 1995; Rodríguez-del-Bosque et al., 2020; Shao et al., 2020). The assumption that wind direction is an important factor in determining the direction of SCA movement can be justified. When the wind direction and a certain field at an initial location are situated at the same angle, the migration probability approaches 1. The triangular distribution depicts this relationship. In this study, the angular range of ±11.25 of the corresponding primary wind direction was used. That is, the range of ±11.25 left and right is set as the angle corresponding to the direction of a 16-point cardinal direction. For example, in the 16-point cardinal direction, a value of ‘8’ means that the wind direction is north and the corresponding angle is 90 degrees with respect to the west direction (0 degrees). In the 16-point cardinal direction, the angle between one value and its neighbor is 22.5 degrees (for example, the angle difference between values 8 and 7 is 22.5). If the value of the 16-

[1] For distance calculations consisting of latitude and longitude, Euclidean distance calculations are not suitable for calculating the distance of a globe with a spherical shape, so a great-circle distance should be used for the distance between two points on the earth’s plane. The Haversine formula is better controlled numerically for small distances. Using the latitude and longitude in the two coordinates, the equation is shown below. $d_h = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\theta_2-\theta_1}{2}\right) + \cos(\theta_1)\cos(\theta_2)\sin^2\left(\frac{\delta_2-\delta_1}{2}\right)}\right)$, where $r$ is the radius of the sphere, $\theta_2, \theta_1$ are latitude points in radians, $\delta_2, \delta_1$ are longitude points in radians.

point cardinal direction is ‘8’, then the range used for the probability distribution is 90 ± 11.25 (78.75-101.25 degrees). The probability distribution depicting this relationship is:

(4) $$P_{PDIR} = 1 - \left((|d_i| * \sigma)/11.25\right)$$

where $P_{PDIR}$ is the probability of SCA movement according to the primary wind direction, $PDIR$ is most common wind direction recorded for a day, σ is a parameter (σ > 0), and $d$ is the difference between the PDIR angle and the angle between the initial location and the sorghum field. As σ increases (perhaps due to higher winds speeds), so too does potential travel distance. For example, if the PDIR value is ‘8’, the angle is 90 degrees. As another example, if the angle between the centroid of Kiowa County and one sorghum field is 45 degrees (northeast), $d$ = 45.

*Cumulative Probability Function for Predicting Movement of SCA*

A probability function for predicting the path of SCA as a function of the weather variables is constructed by combining the weather variable probabilities assuming independent probabilities. Under the assumption of independence, the product of the probabilities is the joint probability for predicting SCA movement and survivorship. In this case, the survival rate, movement direction are all-important factors for the movement of the SCA. The probability function of each field is:

(5) $$P_t^{ij}(\mathbf{x};\boldsymbol{\theta}) = P_{t,rain}^{ij} \cdot P_{t,temp}^{ij} \cdot P_{t,PDIR}^{ij} \cdot P_{t,WSPD}^{ij}$$

where $P_t^{ij}$ is the probability of moving from initial location $i$ to $j$ sorghum field in Oklahoma at date $t$; $\mathbf{x}$ is a vector of rainfall, average air temperature for each county, primary wind direction observations, wind speed; $\boldsymbol{\theta}$ is a parameter vector; $P_{t,rain}^{ij}$is the probability of SCA survival based on rainfall data distribution in initial location to $j$ field at date $t$; $P_{t,temp}^{ij}$ is the probability of SCA movement based on air temperature from an initial location to $j$ field at date $t$; $P_{t,PDIR}^{ij}$ is the probability of SCA movement based on the primary wind direction from an initial location to $j$ field at date $t$; and $P_{t,WSPD}^{ij}$ is the SCA movement probability using the wind speed at an initial location to $j$ field at date $t$.

The above function determines the probability of SCA moving from an initial location to another field, given prevailing weather data. The cumulative probabilities for any date are found by calculating the probability distributions of Equation (5) as the average over the $n$ days of recorded observations:

(6) $$P_{t,n}^{ij}(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{n_t} \sum_t^{t+n} \left( P_{t,rain}^{ij} \cdot P_{t,temp}^{ij} \cdot P_{t,PDIR}^{ij} \cdot P_{t,WSPD}^{ij} \right)$$

where $P_{t,n}^{ij}$ is the cumulative probability of moving from initial location $i$ to $j$ sorghum field in Oklahoma at date $t$, $n_t$ is the number of days for record period $t$. The information on the field where SCA was reported is unknown, but the predicted movement probability was estimated by setting the centroid of Kiowa County, Oklahoma as the initial location. The reason is that in 2018, the first county reporting SCA was Kiowa County on June 15. The exact location was not given so the center of Kiowa County is set as the initial location. The centroid of Kiowa County and the angle and distance between each field were used to estimate the movement probability of the wind direction (PDIR) and wind speed (WSPD). For temperature (TAVG) and precipitation (RAIN), weather information enumerated by date for the county to which each field belongs was used.

*Movement Probability Clusters*

The probabilities derived in Equation (6) are the likelihood of SCA moving from an initial location (e.g., the center of Kiowa County) to each field. Although the field-unit movement probability provides producers with likelihoods that SCA will migrate to their field, it is difficult to ascertain movement to an area or group of sorghum fields throughout Oklahoma. On the other hand, the EDDMapS (2020) infection map only indicates if there was an infection at the county level. The EDDMapS information facilitates understanding of SCA movement trends at the county level, but it does not provide field-level information on infection or probabilities.

Clusters based on field-level movement probabilities provide a more detailed, most-likely infection route by generating movement probabilities for every field. Spatial clusters for SCA infestation – i.e., clusters with a relatively high probability of migrating sugarcane aphid – are generated for fields using the SCA migration probabilities.

One method widely used to analyze spatial clusters is K-means clustering. This procedure classifies clusters according to the average value of the center of the cluster with the closest movement probability value of each field  (Zhang et al., 2008). This property applies more to clusters where the radii and density of observations are relatively equal. This makes it difficult to apply the K-means clustering because Oklahoma's sorghum producing area is concentrated in the west and the predicted probabilities of SCA movement depend mainly on wind direction, which is a random variable. The classification of clusters can also be skewed by outliers and an unnecessarily large number of clusters. Unlike the K-means cluster procedure, density-based clustering procedures construct a relatively non-restrictive cluster by relaxing the assumption of the number of predetermined clusters (Campello et al., 2013). This study constructs clusters for movement probabilities to sorghum fields using density-based clustering methods.

## Results

*Probability Function Predicting SCA Movement*

In 2018, the first SCA reported county in Oklahoma was Kiowa County on June 15. Predicted probabilities of sugarcane aphid migration from Kiowa County to other sorghum fields are calculated for the period between June 15 to 25, 2013 to 2020. The probability of SCA movement, and the location of the field where SCA were observed on June 15 in 2018, is shown in Figure 4. The green dot in the figure is the center of Kiowa County, and the field with the highest probability of SCA movement is 0.367 (red point in Figure 4). Predicted SCA movement is toward Woodward County on June 15. This means that, in Kiowa County on June 15, the field with the highest probability of SCA moving to it is located in Woodward County.

<< Figure 4>>

Equation (6) was used to determine probability distribution of SCA movement starting from Kiowa County's centroid. The probability, on June 15, is the same as that reported in Figure 4. In the case of June 16, there was some likelihood that SCA would move in a different direction after June 15 (PDIR 7; 11.5 degrees; north-northwest) according to Kiowa County's PDIR. On June 16, fields in Ellis County had the highest probability associated with SCA infestation from Kiowa County (0.493 and 0.246, respectively). For the information using only the 25-day weather variables, fields located in Grady County had the highest probability of SCA infestation, with a probability of 0.05. On the other hand, the field with the highest probability of SCA moving elsewhere using the weather variable of 15 to 25 days (11days) is 0.254, which is located in Ellis County (red point in Figure 5).

<< Figure 5>>

From 15 to 25 June 2018, Kiowa County's PDIR value of 5 out of 11 days was 7 (11.5 degrees; north-northwest), meaning that the probability increased for the corresponding wind direction. This result suggests that the movement probability using a single date provides a short-term prediction for SCA movement. Since this probability includes the movement probabilities from previous periods, it also provides producers with an expected movement direction for SCA.

Equation 6 applies the same process from 2013 to 2020 to derive the movement probability by year. This field is used to construct the SCA movement probability cluster by year. Statistics of weather variables used by year are in the appendix (Table 4 in appendix). Counties with fields most likely to have SCA move from them on June 25, 2010 to 2020 were Woodward, Woodward, Custer, Beckham, Woodward, Blaine, Custer, Woodward, Ellis, Custer, and Ellis counties. These counties are located in the northern part of Kiowa County. This finding is consistent with the observed record of SCA movement to the north.

*SCA Movement Probability Clusters*

The density-based clustering analysis generated 12 probability clusters. The number of fields belonging to a cluster for each year is as follows. The closer the color is to red, the higher the cumulative probability of predicted movement to that location (Figure 6-13). The number in the figure is the average predicted cumulative probability of movement for each cluster, and the number in parentheses is the cluster number. Except for 2018, the number of clusters is 2 to 5, and the number of counties to which fields belong to cluster is 5 to 23. In most years, clusters located near Kiowa County had the highest SCA migration probability. This suggests that SCA movement is limited over some distance, even though wind speed affects movement. On the other hand, the frequency of the wind direction during the period used to calculate the movement probability influences the movement direction of SCA.

<< Figures 6-13>>

Counties where actual SCA infestations were reported are compared with fields in counties with relatively high forecast probabilities (95th percentiles of the probability distributions). This comparison is one way to gauge the accuracy of the movement probabilities because the level of analysis is different and the historical SCA reporting field cannot be accurately identified.

SCA reports obtained from EDDMapS (https://www.eddmaps.org/) were used for the comparison (Table 3). Since 2014 was the first year SCA was reported to counties, county-level SCA reporting was used for 2014-2019. The SCA reporting counties were compared to the counties in which the clustered fields belonged. For each year, SCA infestation activity was reported in four out of 10 counties (17 actual reporting counties) in 2014. In 2015, there were reports in nine out of 22 counties (19 actual reporting counties). In 2016, SCA was reported in three of the 12 counties (17 actual reporting counties). In 2017, five out of 12 predicted counties (16 actual reporting counties), one out of 3 predicted counties (32 actual reporting counties) in 2018, and 21 predicted counties (actual reporting counties) in 2019 SCA was reported in 8 out of 10 counties.

These results have limited usefulness for evaluating the accuracy of the SCA movement probability model. First, the unit of analysis used to calculate the probability is a sorghum field, not the county. On the other hand, the regional unit for which SCA is reported is the county. Also, as the value of the percentile (K) used for clustering decreases, accuracy increases, but this increase is an artifact due to the simple increase in the number of prediction counties. Second, since SCA reports for a sorghum field are made with the naked eye, some reports may never be entered even if there was an SCA infestation. In addition, in most cases, SCA-reporting counties are not adjacent, so it is difficult to identify clearly the route of SCA movement. There is also the possibility of SCA inflows from Texas and Kansas to Oklahoma. This, along with the relatively limited distance SCA can travel, makes it difficult to determine clearly the accuracy of the procedure.

### *Empirical SCA Movement Probabilities Clusters*

For the practical application of the moving probability model, data selected from sampled sorghum fields in Oklahoma, 2017 to 2019, is used. These data include the field location, the date of investigation, and the SCA population per plant. If the population of SCA per plant per day is positive, the sorghum field is considered infected. In a probability model using weather variables, the cumulative probability of other fields is derived for the field where the first infection was reported for each year.

In 2017, there were 28 fields in which sugarcane aphid infection was investigated (out of 276 fields). The date when sugarcane aphid was first reported in Oklahoma in 2017 was May 31, and an infection report occurred in a sorghum field in Caddo County. The last reported infection in Oklahoma in 2017 was October 16, in fields in Beaver, Texas, and Cimarron counties (Appendix Figure 22). Counties with reported SCA infections in 2017 were Beaver, Caddo, Canadian, Cimarron, Garfield, Grant, Harper, Jackson, Kay, Kiowa, Noble, Payne, Texas, Tillman, and Woodward counties. During the

11 days from May 31 to June 10, the counties where infections were reported were sorghum fields in Kiowa and Caddo counties (the yellow area in Figure 14). The cumulative probability for the sorghum field sampled from the field belonging to Kiowa County where the infection occurred on May 31 was calculated. Figures 14-16 show the infestation probabilities for each field using the cumulative probability model, covering May 31 to June 10. The cumulative probability is for June 5, and the probability of moving to the field belonging to Jackson County by June 11 is the highest (red square in Figure 14). On June 9th and 10th, fields in Beaver County also registered movement probabilities (blue dots). In 2018, 25 fields were investigated for infection, and there were 151 investigations. The first date of investigation was May 30, and the date and location of the first reported infection was a field belonging to Kiowa County on June 13. Counties with reported infections in 2018 were Caddo, Cimarron, Garfield, Grant, Jackson, Kay, Kiowa, Noble, Texas, and Tillman counties. Kiowa County is the only county in which field reported infections from June 13 to 23. Figure 15 shows the cumulative movement probability over 11 days. Fields belonging to Caddo County have the migration highest probability (red square) for all days in the analysis period. It also appears that other sorghum fields belonging to Kiowa County also have a movement probability. In 2019, there were 29 fields where infection was investigated, and a total of 170 infections were investigated. The date of the first investigation is June 24 and the date of the first report of infection is July 3, a field that will belong to the county of Jackson, Greer.

For fields sampled in 2019, counties with reported infections were Beaver, Caddo, Cimarron, Garfield, Grady, Grant, Jackson, Kay, Major, Noble, Texas, and Tillman counties. During the July 3 to 13 period, the counties in the field where infections were reported were Jackson, Greer, and Caddo counties. Figure 16 shows the cumulative movement probabilities for 11 days after July 3rd. From July 3 to 7, the field belonging to Jackson County had the highest movement probability, but from July 8 to 13, the field belonging to Greer County had the highest probability (red square).

## Conclusions

The purpose of this study was to develop a predictive model of SCA infestation trajectories for Oklahoma sorghum producers. In view of this, it may be possible to develop an early warning information system of SCA movement for sorghum producers. Most studies on SCA focus on economic damages, which are measured with multispectral imagery. The movement and location of SCA affects the production and harvesting of sorghum. This is because multispectral images provided by various institutions can identify large areas, but there remains the problem of whether there is an infestation occurring at a specific point in time and the accuracy (both spatially and temporally) of reports. Infestation maps can confirm the presence of SCA based on visual reporting, but a disadvantage of this method is that the primary unit is a county and detailed movement routes between fields is largely unknown. In addition, it is not possible to predict the location of potential damage because time sensitive information is generally unavailable, which complicates forecasting where SCA may relocate.

Sugarcane aphids are small organisms whose survival and migration are largely determined by weather. Although there are no studies directly analyzing the relationship between sugarcane aphid and the weather, studies that the survival rate and migration of insects are determined by precipitation, temperature, and wind assume that sugarcane aphid migration and survival are also determined by the weather. In addition, using weather information in predicting the migration path of sugarcane aphid has advantages in terms of data availability and cost, unlike multispectral and visual reporting. Under this assumption, a distribution of SCA survival rates and migration was generated using the temperature, precipitation, wind direction, and wind speed. Movement probabilities were generated for June 15 to June 25, for 2013 to 2020. The starting dates correspond with the first sightings of SCA in Oklahoma. The initial and movement probability of each day depends entirely on the wind direction recorded on that day. However, as time passes, the importance of wind direction diminishes.

The calculated movement probabilities were used to forecast SCA movement from an initial sorghum field to others between June 15 to June 25 for each year analyzed. A clustering algorithm was used to generate groups of low and high infestation probabilities, and the fields that could be infected. Based on actual sugarcane aphid infection counties, since 2016, the accuracy is 0.18 (3 out of 17), 0.31 (5 out of 16), 0.08 (1 out of 12), and 0.89 (8 out of 9).

The intention of the exercise was to develop a method whereby an early system predicting SCA infestation could be developed for sorghum producers. An ideal system would inform the development of preventative measures to mitigate the effects of SCA damage on crops. In addition, prediction of movement routes using weather variables may be an alternative to the high cost of using existing multispectral images. The present analysis could provide a framework for future research regarding SCA survivability and migration modeling. The probability distributions used here were ad hoc in that they were an inelegant solution to address a difficult problem of data unavailability. More structured experiments on the effects of the variables considered here and their relationship to SCA infestation on sorghum fields would provide better information on the types of distributions suitable for reflecting these relationships, in addition to the parameters governing the shape and scale of the distributions capturing the biological response of SCA to weather.

**Table 1-1. Summary Statistics for Weather in Oklahoma, 2018**

| Variable | TAVG (°F) | PDIR (16-point) | WSPD (miles/hours) | RAIN (inches) |
|---|---|---|---|---|
| Mean | 78.75 | 7.41 | 9.34 | 0.26 |
| Standard Deviation | 4.45 | 2.72 | 3.8 | 0.55 |
| Min | 66.25 | 1 | 2.47 | 0 |
| Max | 89.38 | 15 | 21.72 | 4.74 |

Note: TAVG is average of all 5-minute averaged temperature observations, PDIR is most common wind direction for the day, WSPD is average of all 5-minute wind speed, RAIN is liquid precipitation measured each day, sorghum area is sorghum area (acre) by county in Oklahoma

Source: Mesonet (https://www.mesonet.org), USDA NASS (https://www.nass.usda.gov/)

**Table 1-2. Counties Observed in Oklahoma SCA in 2018 and Average Weather for 10 Days after Observation**

| Report Date (M/D/Y) | Station ID | County | TAVG (°F) | RAIN (inches) | WSPD (miles/hours) | PDIR (16-point) |
|---|---|---|---|---|---|---|
| 6/15/2018 | HOBA | Kiowa | 81.1 | 0.5 | 14.9 | 7.7 |
| 6/20/2018 | APAC | Caddo | 80.5 | 0.7 | 12 | 7.8 |
| 6/21/2018 | ACME | Grady | 81 | 0.6 | 11.7 | 7.7 |
| 6/21/2018 | HOBA | Kiowa | 82.1 | 0.9 | 14.2 | 7.7 |
| 6/27/2018 | CARL | Payne | 81.7 | 0.9 | 6 | 6.5 |
| 7/3/2018 | BESS | Washita | 81.6 | 0.4 | 7.9 | 6.1 |
| 7/3/2018 | LANE | Atoka | 80.5 | 0.3 | 4.5 | 5 |
| 7/3/2018 | REDR | Noble | 82.2 | 0 | 6.3 | 5.2 |
| 7/10/2018 | BREC | Garfield | 84.8 | 0.1 | 7.8 | 5.6 |
| 7/10/2018 | MEDF | Grant | 83.5 | 0.9 | 6.9 | 5.1 |
| 7/25/2018 | BEAV | Beaver | 75.2 | 0.5 | 9.2 | 5.8 |
| 8/1/2018 | EVAX | Texas | 75.3 | 0.2 | 9.2 | 7.3 |
| 8/1/2018 | FAIR | Major | 81.9 | 0.7 | 8 | 7.9 |
| 8/1/2018 | LAHO | Major | 80.8 | 0.5 | 8.7 | 8.3 |
| 8/1/2018 | SLAP | Beaver | 78.6 | 0.2 | 9.7 | 9.8 |
| 8/22/2018 | BOIS | Cimarron | 75.4 | 0.1 | 12.6 | 8.1 |
| 8/25/2018 | ARNE | Ellis | 79.5 | 0.5 | 9.6 | 6.8 |

Note: TAVG is average of all 5-minute averaged temperature observations, PDIR is most common wind direction for the day, WSPD is average of all 5-minute wind speed, and RAIN is liquid precipitation measured each day.

Source: Mesonet (https://www.mesonet.org/)

**Table 1-3. SCA Migration Forecast Counties for 2013-2020**

| 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|
| Beaver | Alfalfa | Alfalfa | Beckham | Beckham | Beaver* | Alfalfa | Beaver |
| Beckham | Custer* | Beckham | Caddo* | Caddo* | Ellis | Beckham | Custer |
| Caddo | Garfield* | Blaine | Comanche | Comanche | Harper | Caddo* | Ellis |
| Comanche | Grant | Caddo* | Cotton | Cotton | | Comanche* | Harper |
| Cotton | Kay* | Comanche | Custer | Custer | | Cotton | Washita |
| Custer | Kingfisher | Cotton | Grady* | Grady* | | Custer | |
| Ellis | Major | Custer | Greer | Greer | | Garfield* | |
| Grady | Noble* | Garfield* | Jackson | Jackson* | | Grady* | |
| Greer | Osage | Grady* | Kiowa | Kiowa* | | Grant* | |
| Harper | Woods | Grant* | Stephens | Stephens | | Greer | |
| Jackson | | Greer | Tillman* | Tillman* | | Jackson* | |
| Kiowa | | Harmon | Washita | Washita | | Kay | |
| Stephens | | Jackson* | | | | Kingfisher | |
| Tillman | | Kay* | | | | Kiowa* | |
| Washita | | Kingfisher | | | | Major | |
| | | Kiowa | | | | Noble* | |
| | | Major | | | | Osage | |
| | | Noble* | | | | Stephens | |
| | | Osage* | | | | Tillman | |
| | | Stephens | | | | Washita | |
| | | Tillman* | | | | Woods | |
| | | Washita | | | | | |
| | | Woods | | | | | |

Note: * means that the SCA migration forecast counties belong to each year's SCA actual reporting counties.

Source: EDDMapS (https://www.eddmaps.org/), Texas A&M AgriLife Extension (2020) (https://agrilifeextension.tamu.edu/)

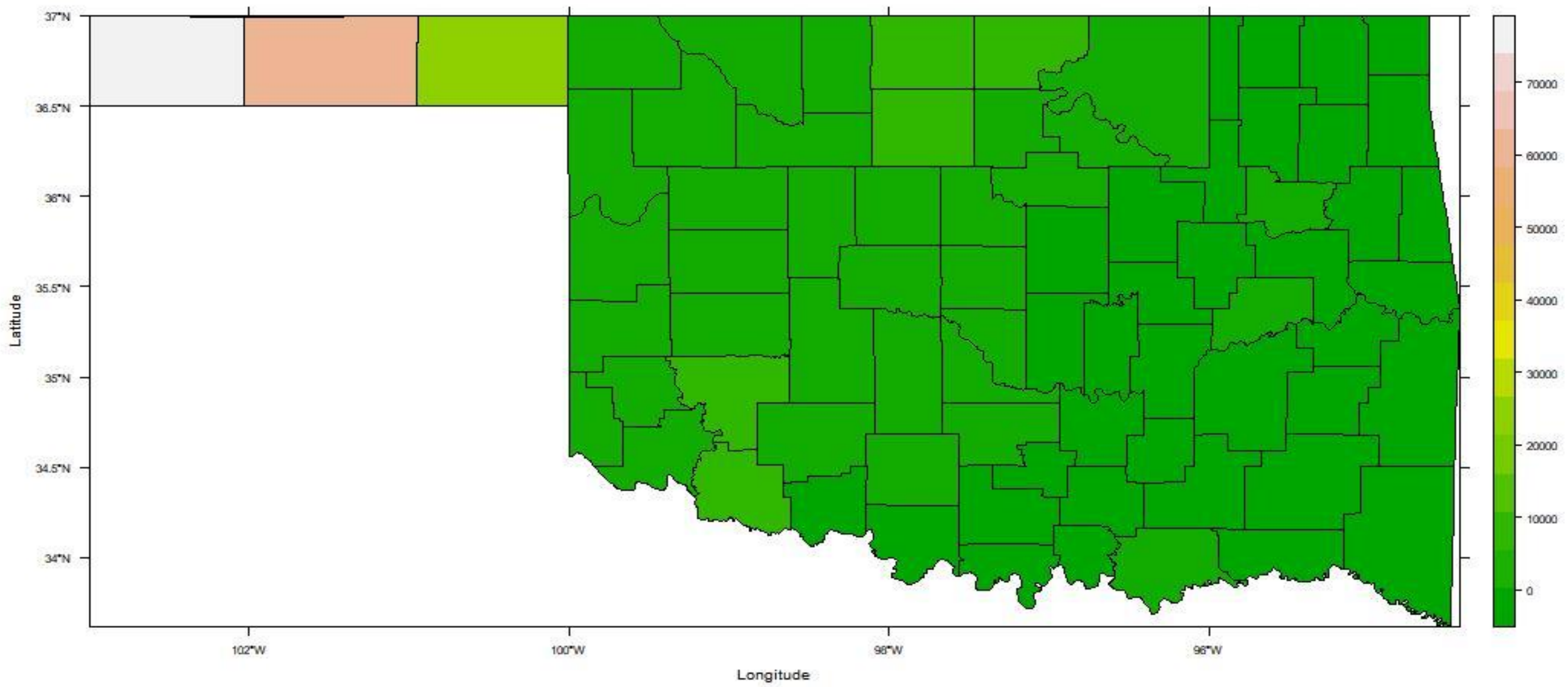


**Figure 1-1. Distribution of the Sorghum Area of Oklahoma in 2018 (Acres)**

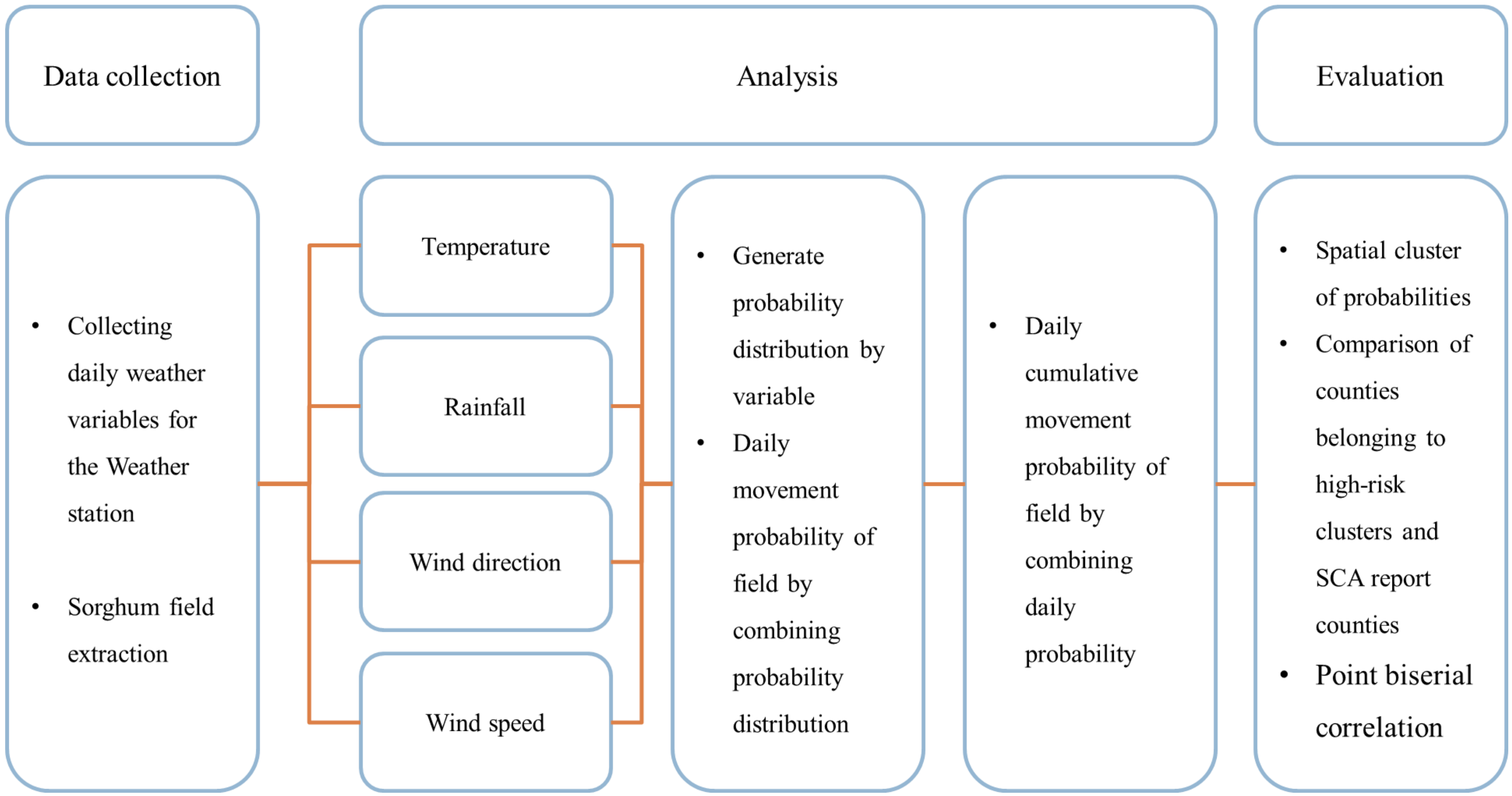


**Figure 1-2. Methodology Flow Chart**

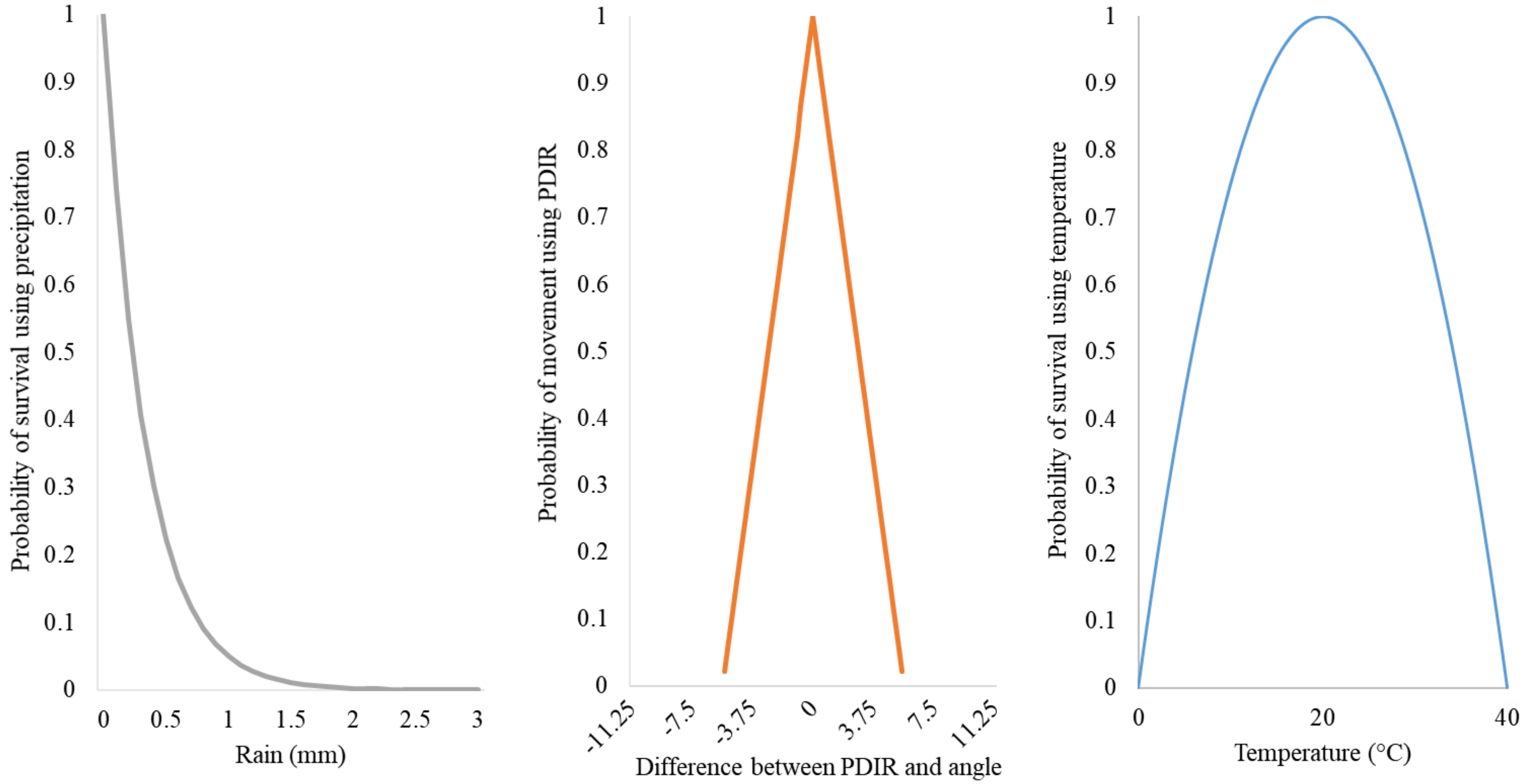


**Figure 1-3. Relationship between Weather Variables and Movement Probability (*β*=3, *w*=10, *κ*=0.1, *λ* =-0.0026)**

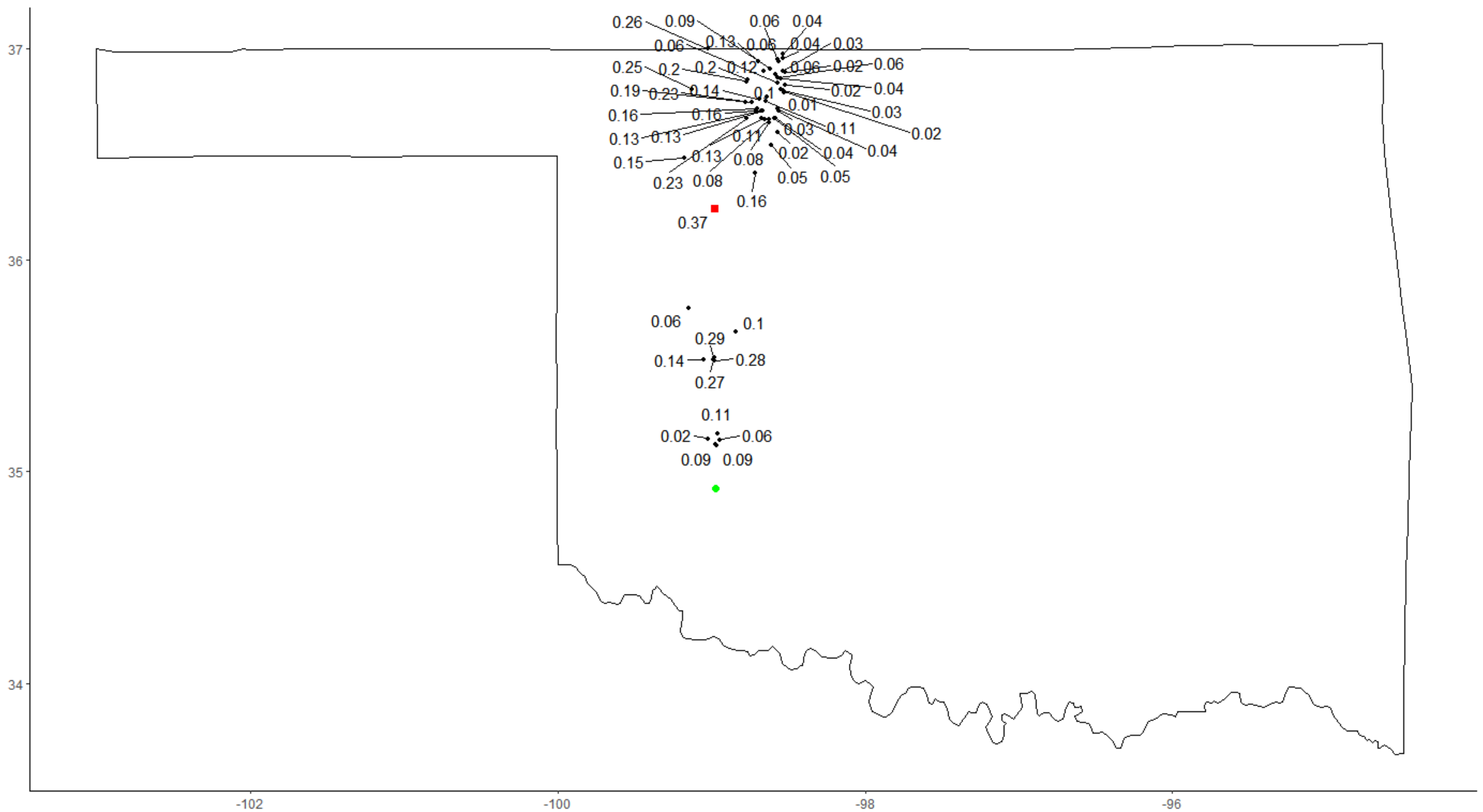


**Figure 1-4. SCA Predicted Movement Probabilities and Location based on Weather Variables, June 15 2018**

**Note: The green dot indicates the centroid of Kiowa County, and the red square indicates the field with the highest movement probability.**

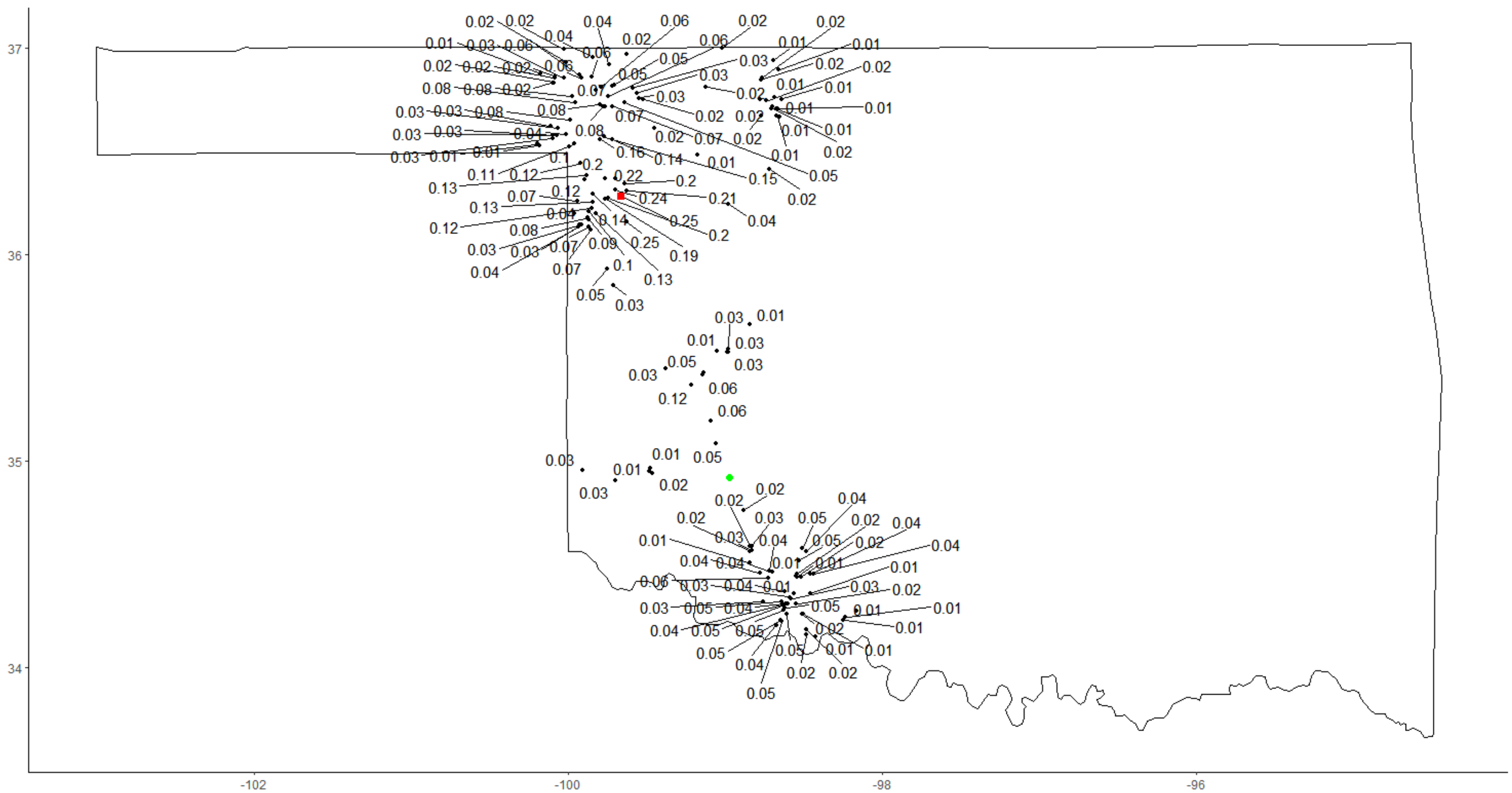


**Figure 1-5. SCA Predicted Movement Probabilities and Location using Weather Variables from 15 to 25 June 2018**

**Note: The green dot indicates the centroid of Kiowa County, and the red square indicates the field with the highest movement probability.**

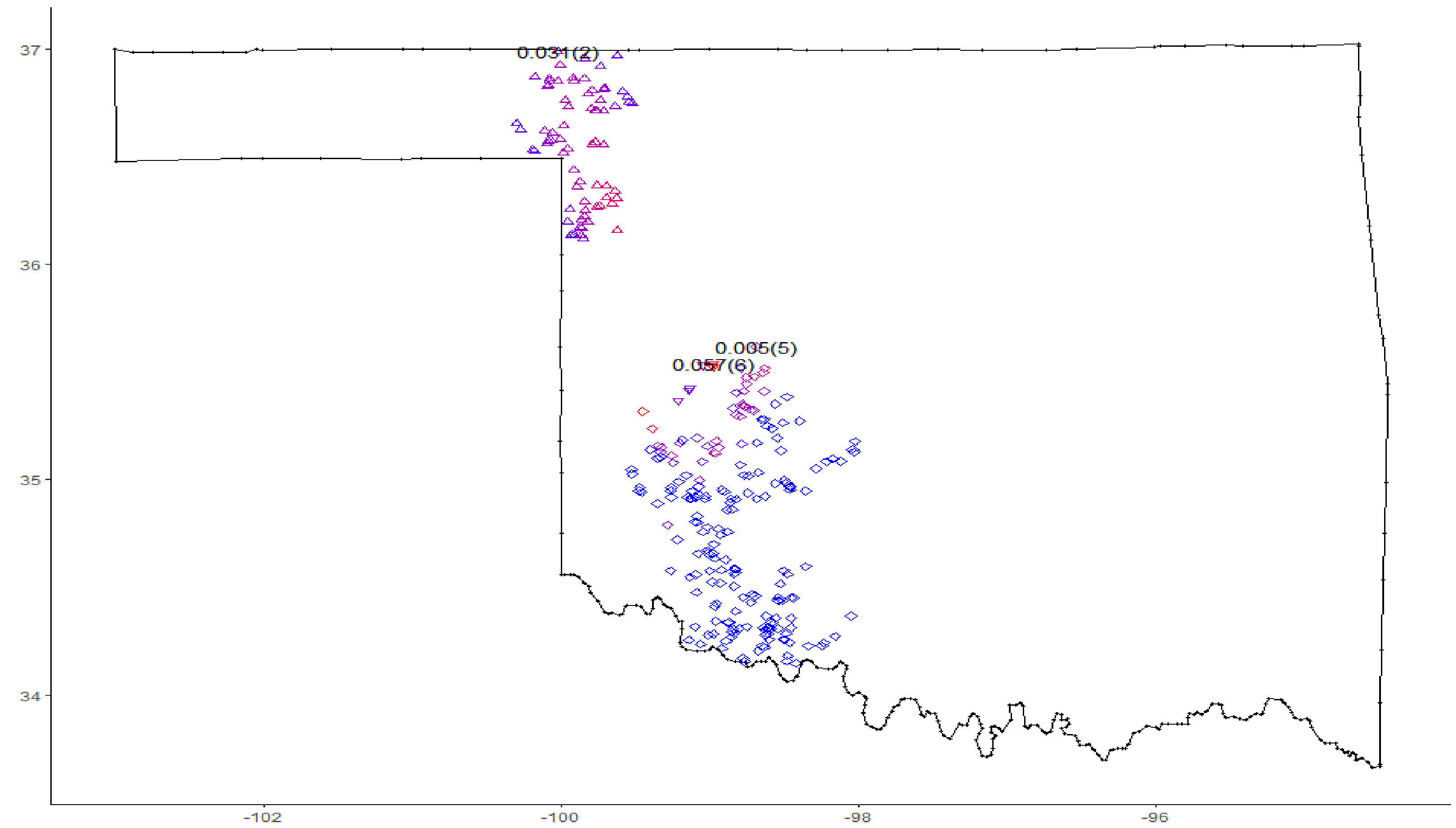


**Figure 1-6. A Cluster of SCA Movement Probability for Fields, 2013**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

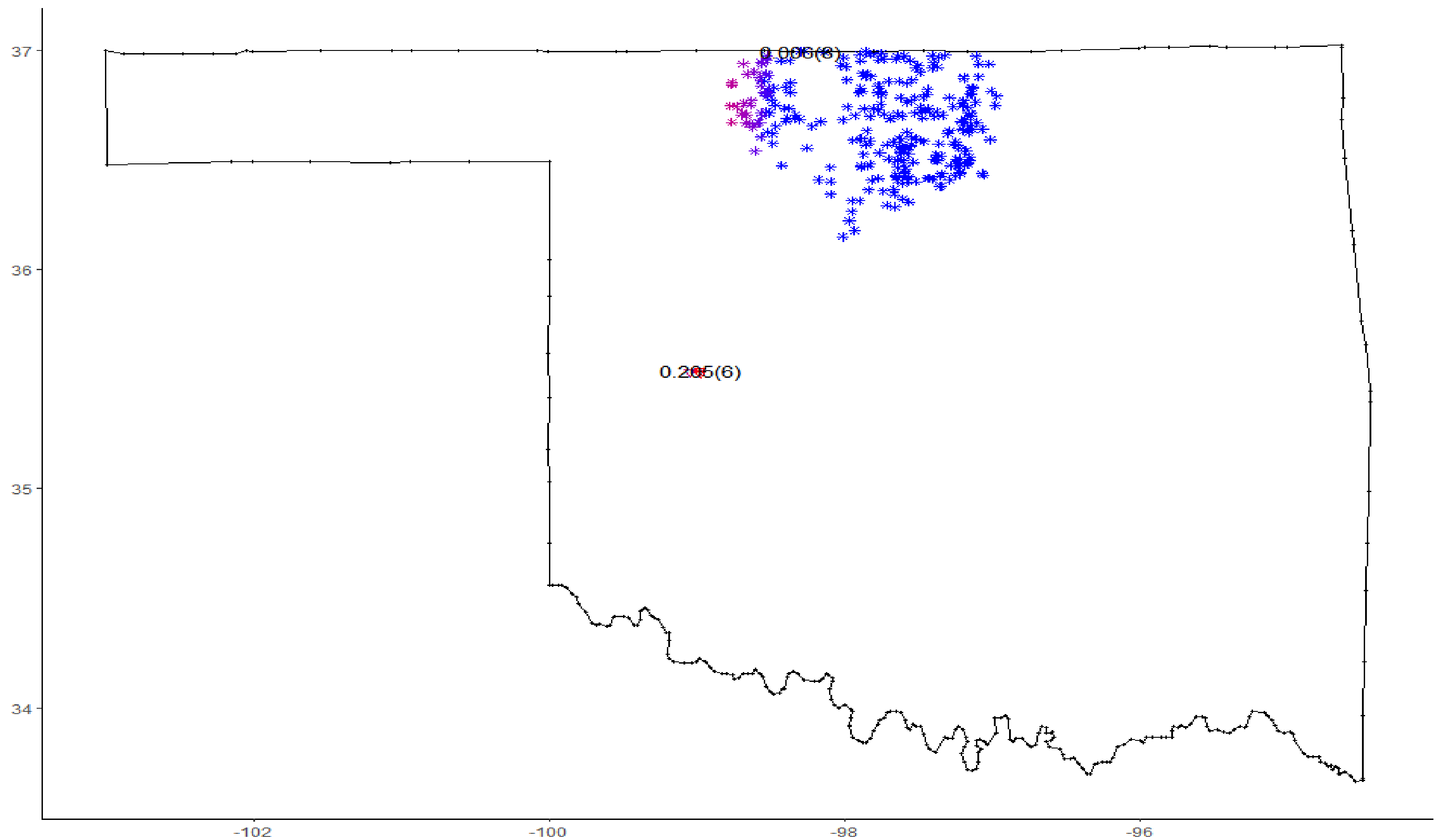


**Figure 1-7. A Cluster of SCA Movement Probability for Fields, 2014**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

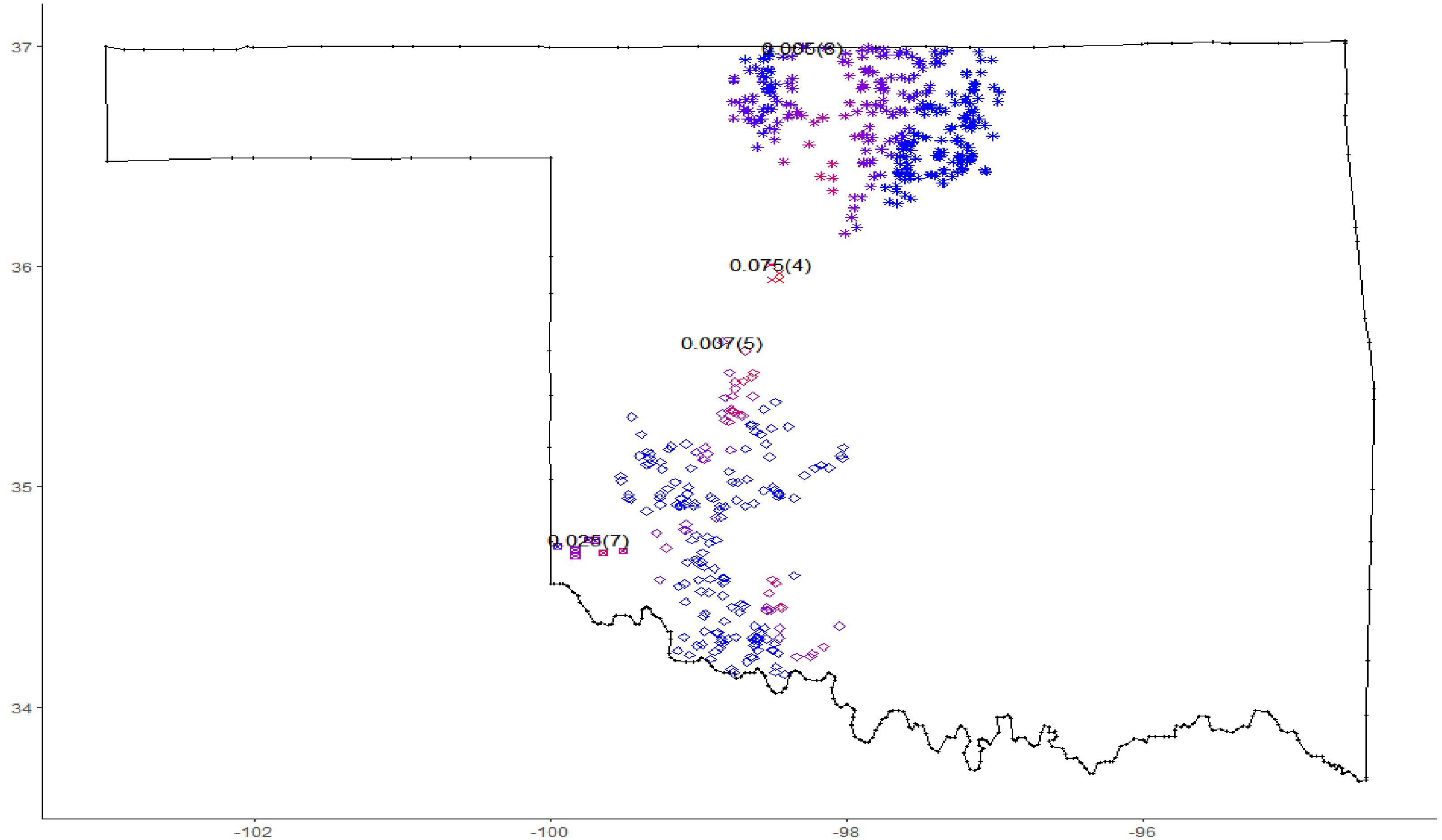


**Figure 1-8. A Cluster of SCA Movement Probability for Fields, 2015**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

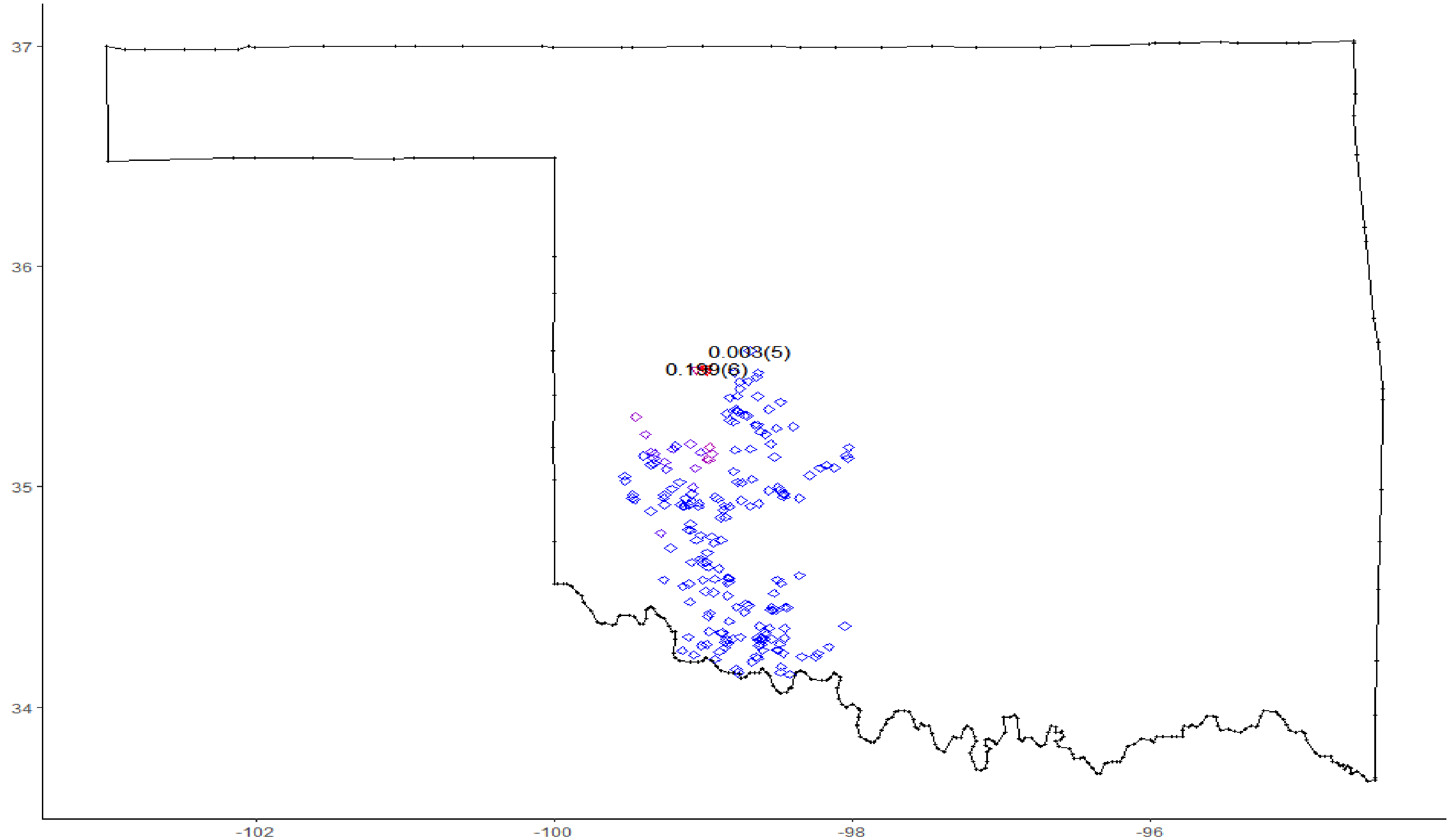


**Figure 1-9. A Cluster of SCA Movement Probability for Fields, 2016**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

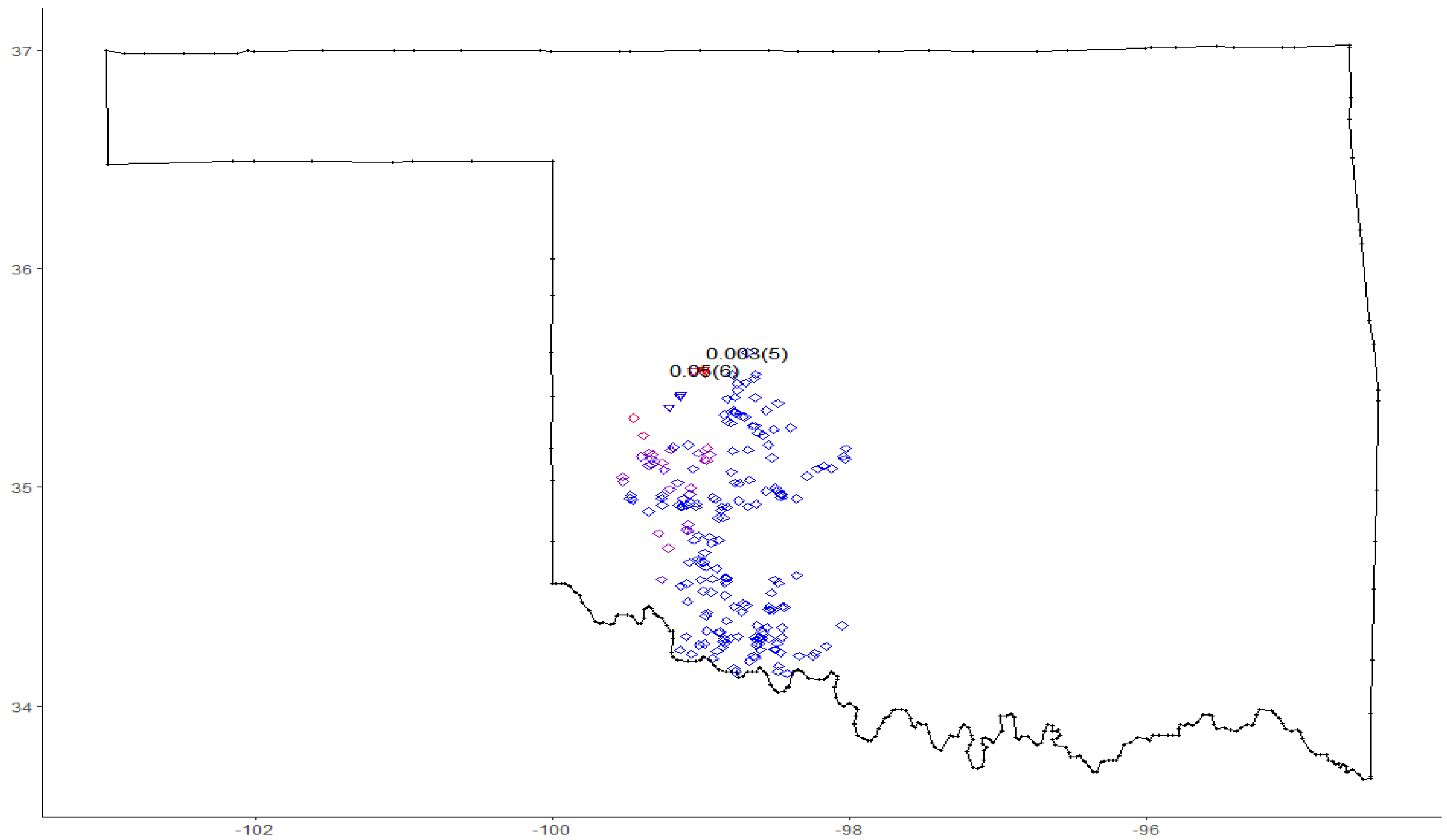


**Figure 1-10. A Cluster of SCA Movement Probability for Fields, 2017**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

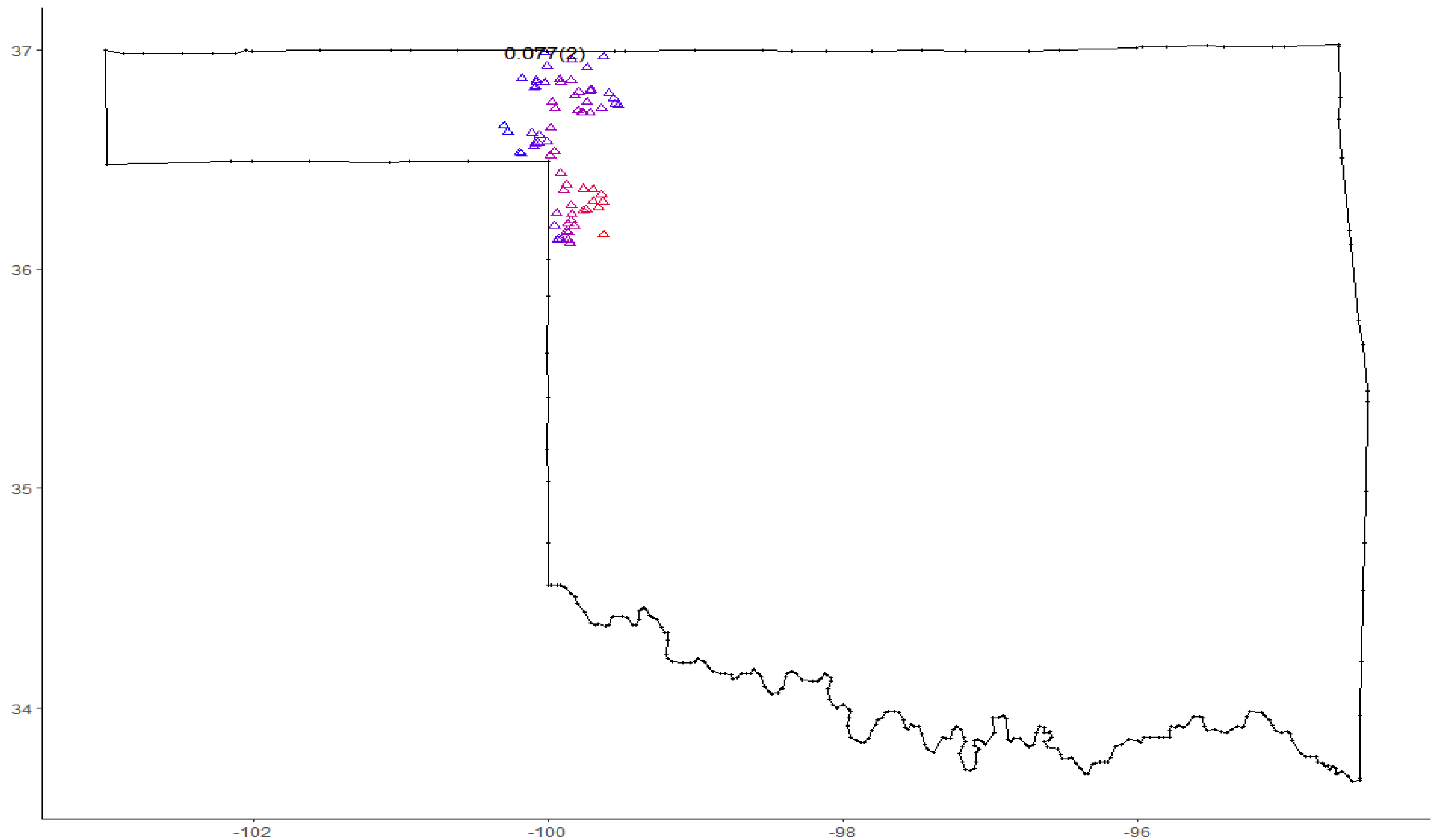


**Figure 1-11. A Cluster of SCA Movement Probability for Fields, 2018**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

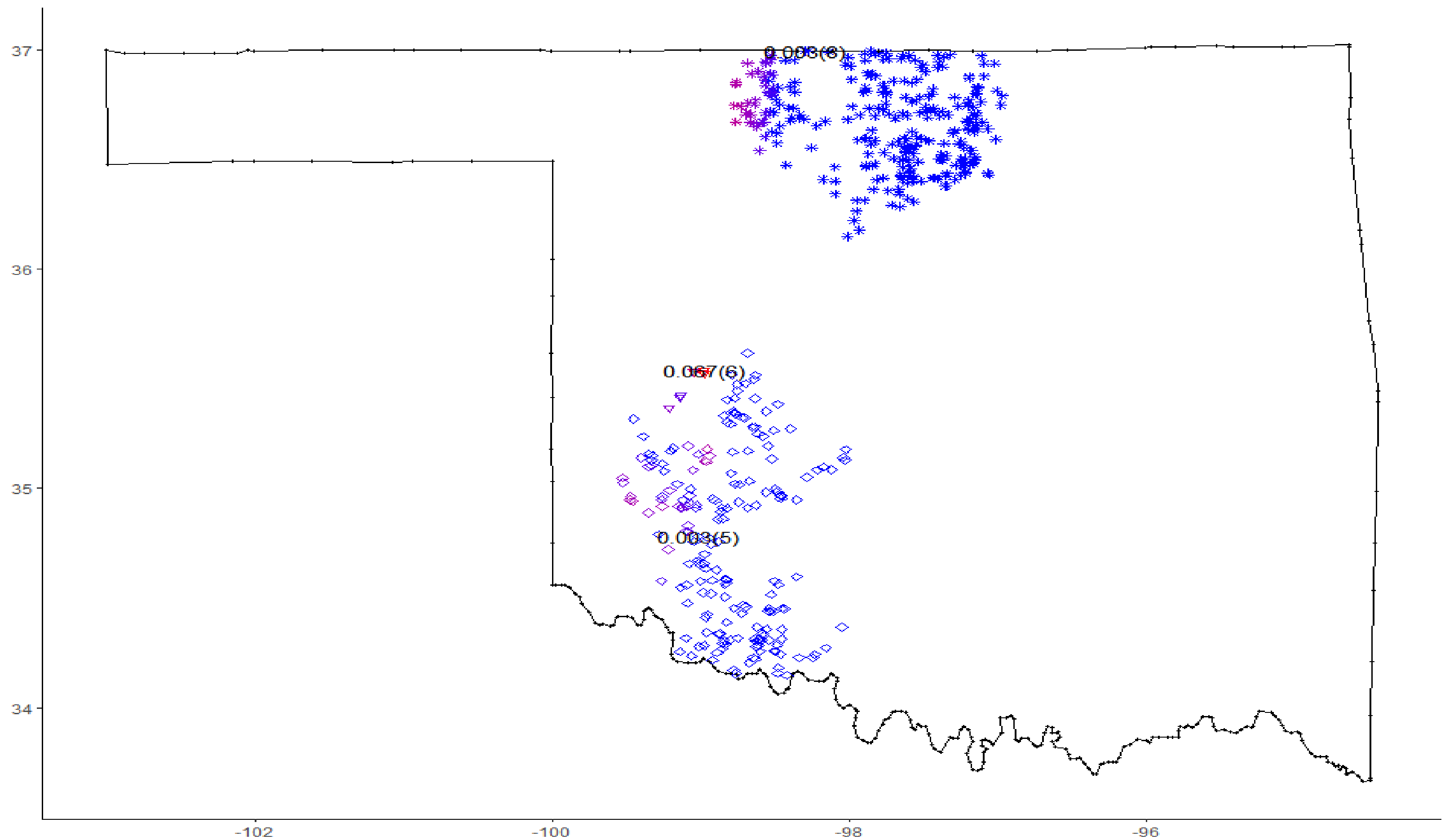


**Figure 1-12. A Cluster of SCA Movement Probability for Fields, 2019**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

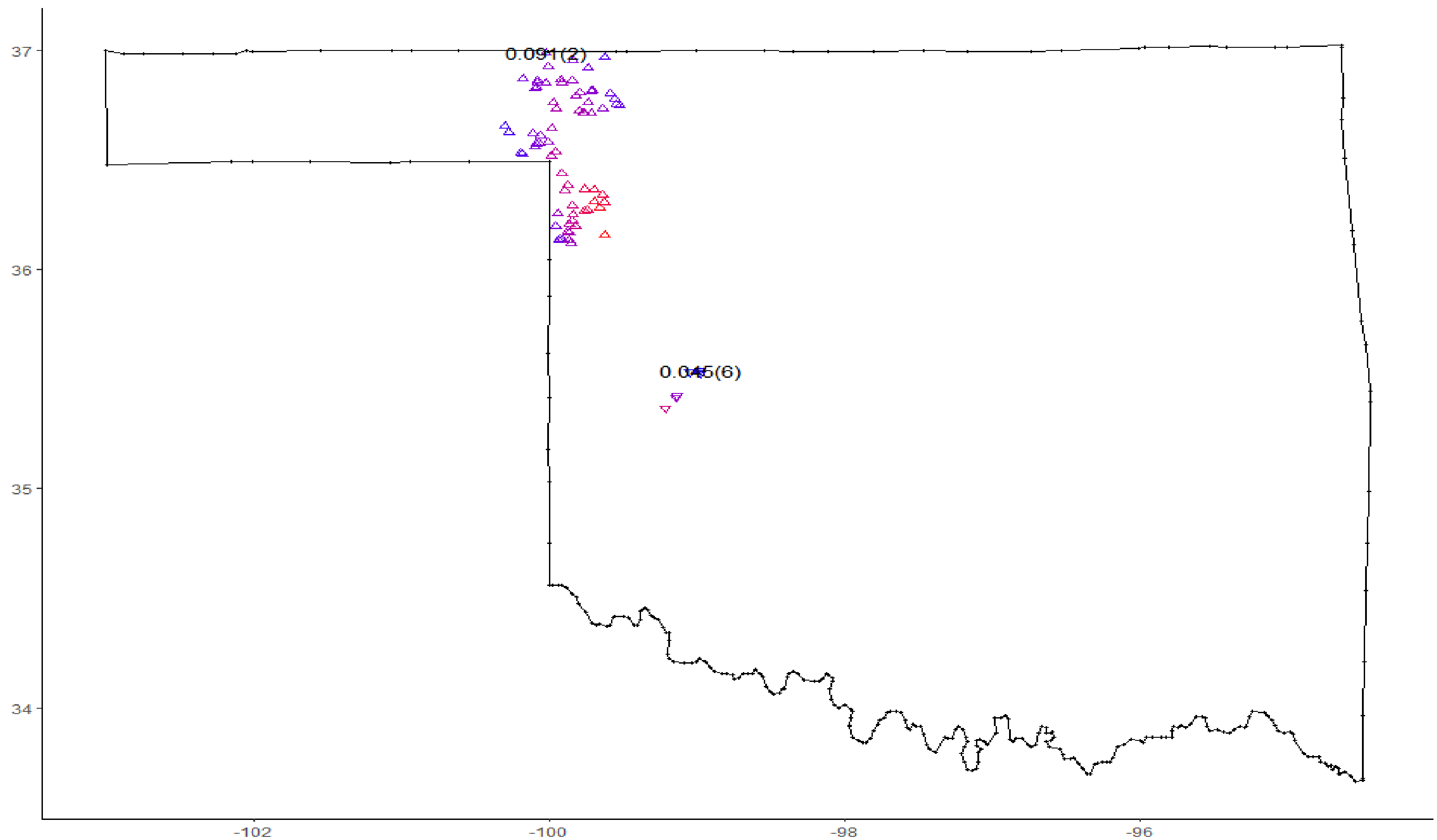


**Figure 1-13. A Cluster of SCA Movement Probability for Fields, 2020**

**Note: The shape of the fields means they belong to different clusters. The closer to red, the higher the probability. The number means the average cumulative movement probability for each cluster.**

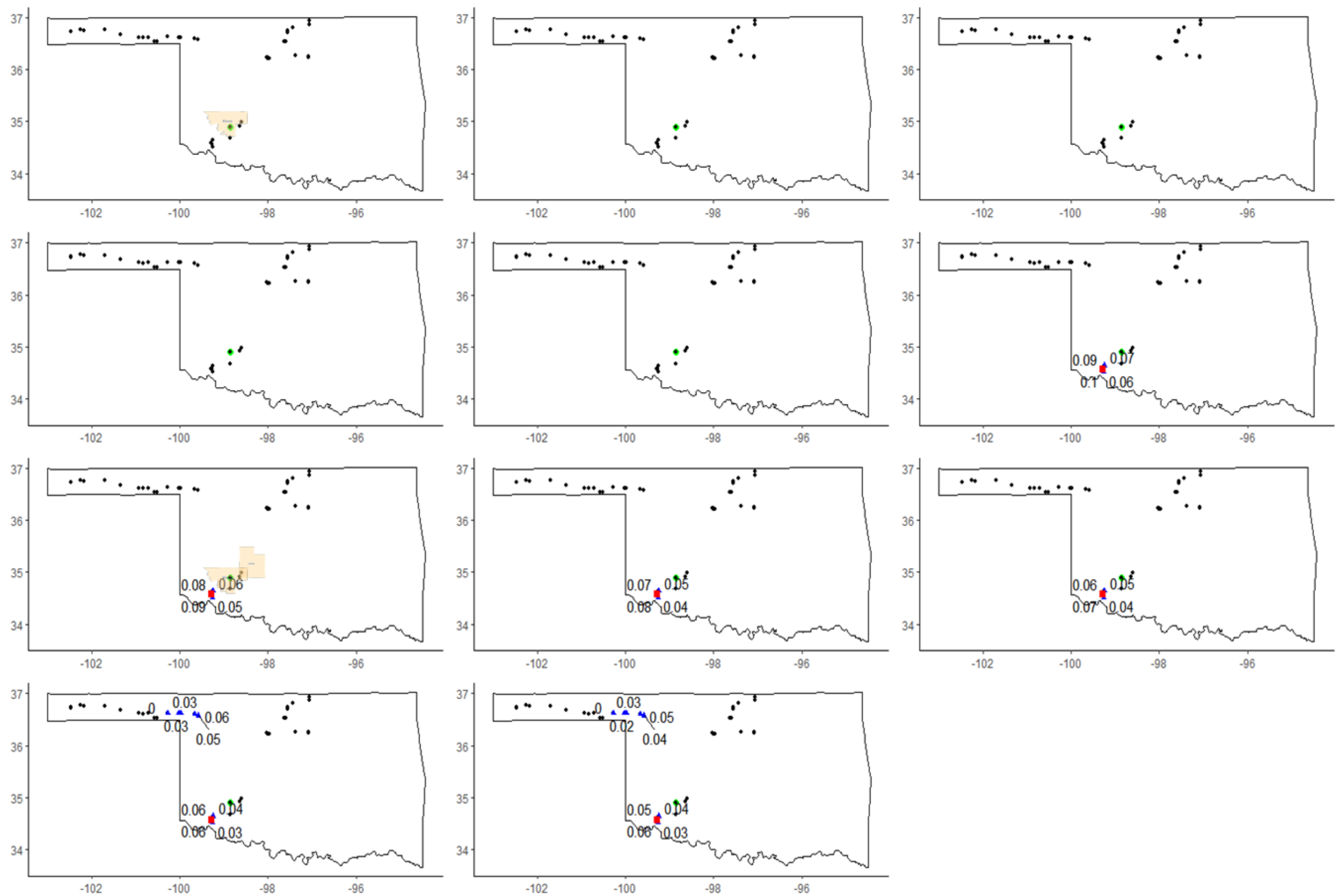

**Figure 1-14. Predicted Movement Probability and Location using Weather Variables on May 31-June 10 2017**

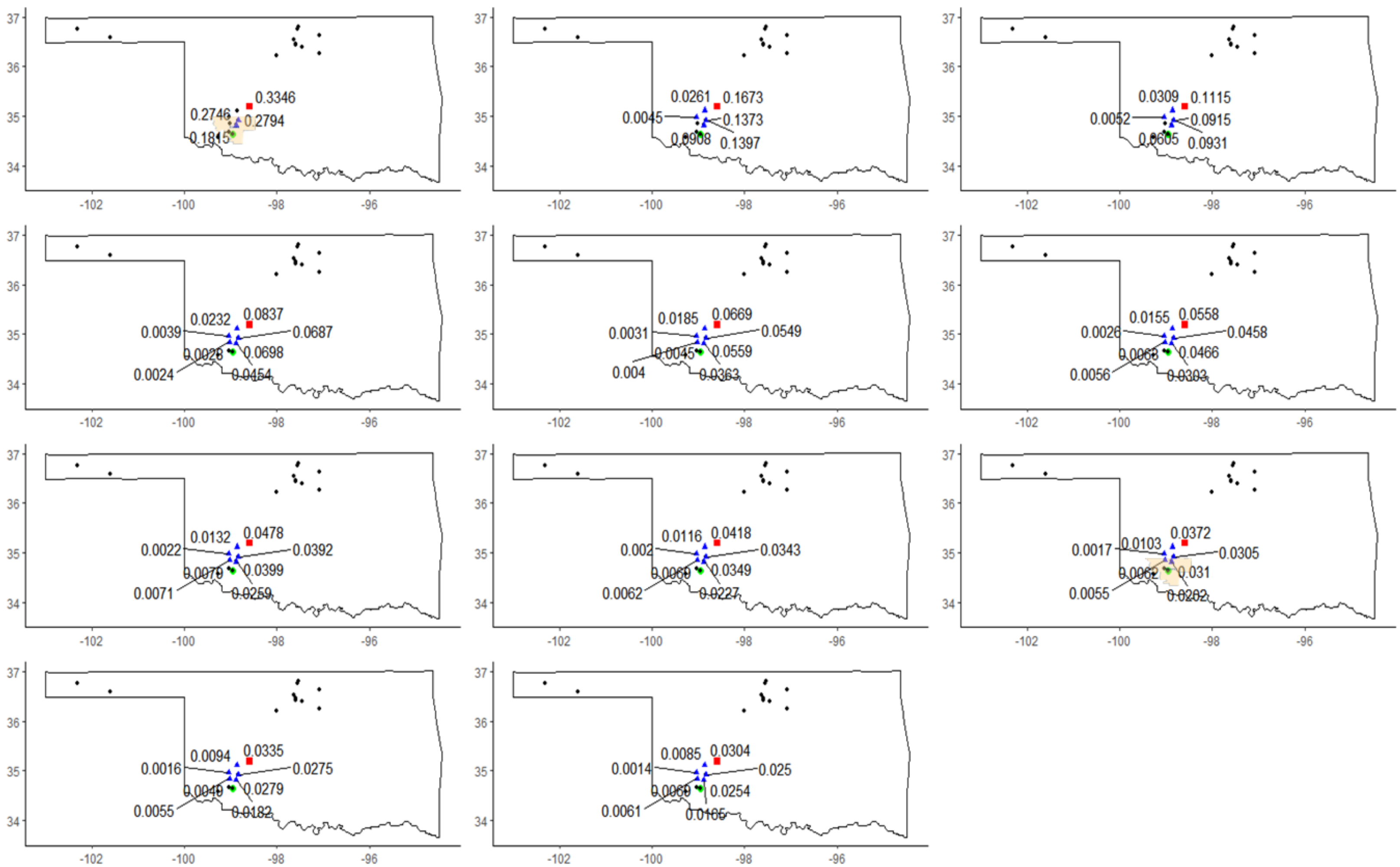

**Figure 1-15. Predicted Movement Probability and Location using Weather Variables on June 13- June 23 2018**

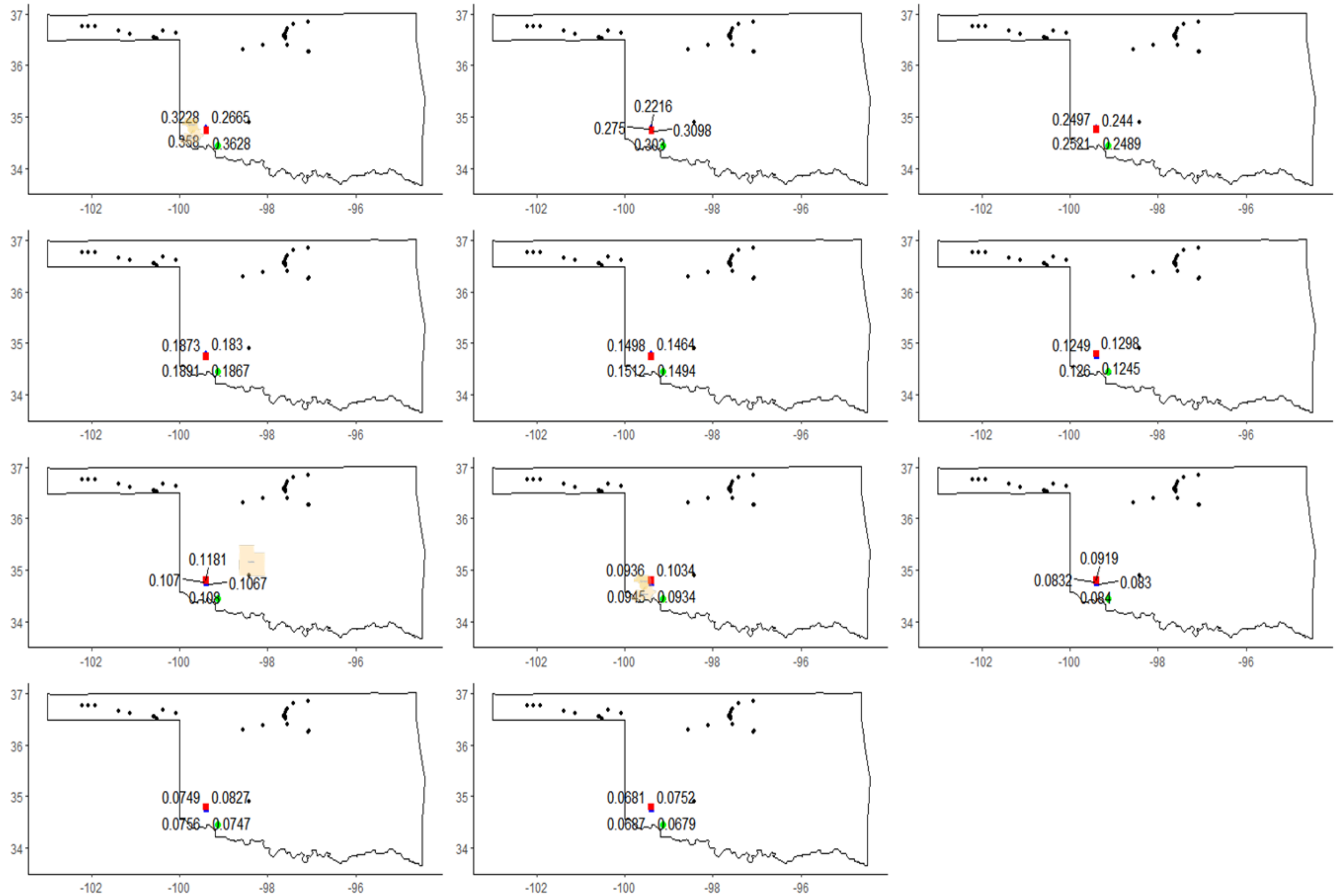


**Figure 1-16. Predicted Movement Probability and Location using Weather Variables on July 3- July 13 2019**

CHAPTER II

# WEATHER AND GEOGRAPHIC EFFECTS ON SUGARCANE APHID MIGRATION IN OKLAHOMA

## Introduction

This study measures the effect of weather variables and geographic features on the infestation of Oklahoma sorghum production by Sugarcane aphid (SCA). Sugarcane aphid (*Melanaphis sacchari* (Zehntner), Hemiptera: Aphididae) is a pest found in more than 30 countries that feeds on 20 species of rice and other commodity crops (Singh et al., 2004). In the United States, sorghum, sweet sorghum, and some millet varieties are SCA's main hosts since its introduction in the 1970s (Singh et al., 2004). SCA transmits sugarcane yellow leaf virus, causing yield loss in sorghum (Gonçalves, 2005; Lopes et al., 1997; Paray et al., 2011).

Damage to sorghum caused by SCA depends on factors such as aphid/plant density and the duration of infestation (Singh et al., 2004). Weather variables also affect SCA population density, their migration, and survival. Studies measuring the relationship between temperature and SCA population density find that dispersal of aphids occurs within 6 to 10 days at temperatures between 18 and 30°C, with populations destroyed at 35°C and higher (Behura & Bohidar, 1983). SCA thrive and proliferate on sorghum as humidity increases (Mote, 1983). Warm temperatures have a positive effect on aphid populations until a critical temperature of 35°C is reached, while rainfall has a negative effect on survival (Mann et al., 1995). Rainfall also removes aphids from host plants and makes them more vulnerable to predators (Cocu et al., 2005; Klueken et al., 2009). SCA infestation of sorghum may be most severe during the late sorghum growth phase due to plant stress caused by drought. SCA colonies increase rapidly after the piping stage (the extrusion of tailpipe-like appendages) (Van Rensburg, 1973). SCA populations on a single plant may reach as much as 30,000 individuals (Setokuchi, 1977).

Weather variables including temperature, humidity, and precipitation, as well as the growth stage of sorghum, affect the susceptibility of sorghum plants to SCA and their proliferation. Documentation of the effects of environmental variables on insect life cycles is often performed in laboratories. This study is the first to examine the effect of geographic and weather variables on the migration of SCA at the sorghum-field level using a survival model, which is also called a proportional hazard model (PHM). PHM is a flexible, regression-based modeling approach capable of incorporating the effects of weather, location, and temporal covariates for determining the likelihood of field-to-field SCA migration.

The study uses field-level observations of SCA infestations on Oklahoma sorghum fields to determine how weather variables and field proximity influence the time until a given field is infested. Infestation dynamics and population colonization vary, depending on factors such as native vegetation, food density, host plant species abundance, timing and rate of migration and dispersal, and geographic location (Singh et al., 2004). The covariates used in the SCA migration model include distance between sorghum fields, temperature, precipitation, wind direction, and the date when SCA were observed on a given field. Findings may be useful for developing predictive models for SCA migration and time until infestation.

**Data**

Data were obtained from the United States Department of Agricultural Research Station in Stillwater, Oklahoma[2]. Data were collected in 2017 from 47 Oklahoma sorghum fields. Observations were made on multiple days at an irregular frequency, ranging from one to 14 day intervals. Data collection resulted in 433 observations for the 47 sorghum fields surveyed. The number of observation days is 139, spanning from May 31 to October 16, 2017.

A field was considered infested if winged or wingless SCA were identified on sorghum plants. The period from May 31 to the observed infestation date marks the time ($t$, days) until a field was infested. If the field was never infested from May 31 to October 16, then it received a '0'. The '0' indicator censors un-infested fields. In this study, the date until the first infestation of the field is used, not the date when the field is infested. The use of the date the field was infested can lead to systematic bias in estimating the effect of multiple covariates. Also, the date until the first infestation occurs is used in the data structure for the survival model (Allison, 2010). For example, an infestation reported on August 16 corresponds with a period of 77 un-infested days. Thirty-four of Oklahoma's 47 sorghum fields were infested during the data collection period, with the remaining 12 fields classified as un-infested on October 16 (Table 2).

[2] Dr. Norman Elliott, United States Department of Agriculture-Agricultural Research Service (USDA-ARS) SDA-ARS Plant Science Research Laboratory, 1301 N. Western Rd., Stillwater, OK 74075.

The covariates included in the SCA migration model fall into three categories: geographic relationships between sampled sorghum fields, weather variables, and SCA population characteristics (Table 1). A description of the covariates belonging to each category follows.

<< Table 1 >>

*Geographic Relationships between Sorghum Fields*

Geographic proximity between sorghum fields are modeled using regional dummy variables and distances between sorghum fields. SCA infestations typically begin in the southern region of Oklahoma. SCA then migrate northward. Infestation occurred first in the southwest region (first sighting, May 31), followed by the north-central region, and finally the northwest region. Reporting dates for the north-central region occurred between June 1 to October 3. Reporting dates for the northwest region were from July 7 to October 16 (Figure 1).

The division of Oklahoma's sorghum-producing areas into three regions controls for regional differences in landscape, soil quality, and other agroecological features. The 47 sorghum fields were situated in three distinct growing regions of Oklahoma: 11 sorghum fields were in the southwest region, 18 sorghum fields were in the north-central region, and 18 sorghum fields were in the northwest region of Oklahoma (Figure 1). Field-to-field distances were calculated using the Haversine formula[3]. Distance variables are in miles.

---

[3] The Haversine formula is $d_h = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\theta_2-\theta_1}{2}\right) + \cos(\theta_1)\cos(\theta_2)\sin^2\left(\frac{\delta_2-\delta_1}{2}\right)}\right)$, where $r$ is the radius of the sphere, $\theta_2, \theta_1$ are latitude points in radians, $\delta_2, \delta_1$ are longitude points in radians.

Four distances were constructed. The first distance variable is the distance between the first field infested in a region to other fields in the same region.

<< Figure 1 >>

The second distance variable is the distance between the last infested sorghum field in Texas to all study fields in Oklahoma. This variable controls for the northward movement of SCA from Texas to Oklahoma sorghum fields as a function of geographic distance.

The third distance variable is the distance between the first sorghum field infested by SCA in Oklahoma to all other Oklahoma sorghum fields. This variable also controls for the northward movement of SCA as a function of geographic distance.

The fourth distance variable is the distance between infested fields in terms of temporal priority. For example, the distance between the field where the first infestation occurred (Field ID = 24, date: May 31, southwestern Oklahoma, Figure 1) to the field where the second infestation occurred (Field ID = 21, data: June 14, southwestern Oklahoma) was 167.1 miles. Sequential field-to-field infestations distances were calculated for each surveyed field.

*Weather Variables*

Weather variables include temperature, precipitation, and wind direction. Studies on the biological effects of temperature and precipitation on SCA conclude that these two factors directly affect aphid mobility, reproduction, and survival (Souza & Davis, 2020; Ukoroije & Abalis, 2020). Aphid movement is driven largely by atmospheric conditions (Irwin et al., 2007). Winged SCA alatae are more likely to travel longer distances when atmospheric conditions are stable. Aphids cannot migrate over long distances during adverse, or unstable, weather conditions when temperature and precipitation are in flux (Isard et al., 1994; Isard & Gage, 2001).

Weather variables include average air temperature (TAVG), precipitation (RAIN), and the dominant wind direction (PDIR). TAVG is the average of all temperature observations at 5-minute interval each day (degrees Celsius). RAIN is precipitation measured each day (mm). PDIR is the most common wind direction recorded for a day, and is based on 16-point compass heading with a 16-point cardinal direction (Mesonet, 2020) All weather variables were obtained from Oklahoma's Mesonet system (Mesonet, 2020).

Weather variables were imputed for each sorghum field by generating kriged surfaces from data collected at each of Oklahoma's 119 Mesonet stations. Kriging is a geostatistical interpolation technique that generates predictions for unobserved locations using distance and information between known data locations (Cressie, 1993; Paramasivam & Venkatramanan, 2019). Kriging was necessary because the weather variables were not measured at the study fields. Weather data were kriged for each of 139 days using a spherical semivariogram function, $\gamma(h) = C_0 + C \cdot \left(1.5\frac{h}{a} - 0.5(\frac{h}{a})^3\right) if\ h < a,$ otherwise $\gamma = C_0 + C$, where $\gamma$ is the semivariogram value, $a$ is the effective range (distance), the distance at which the function reaches its maximum value, $C$ is the variogram sill, $C_0$ is the variogram nugget (variance remaining after accounting for space), $h$ is the lag of separating distances between fields (Brooker, 1986; Burgess & Webster, 1980).[4] For each weather covariate, parameters were derived and used for each 139 days. For temperature, the mean of the range, sill, and nugget for 139 days was

[4] Anisotropy was not detected for any weather variables.

2676.71, 9.09, and 0.1. The mean of the range, sill, and nugget for precipitation was 510.76, 0.19, and 0.04. The averages for the range, sill, and nugget over 139 days for wind direction are 1011.10, 4.67, and 2.32, respectively.

The (x,y)-coordinates of the kriged surfaces were matched with the (x,y)-coordinates of each of the 47 study fields for the period of 31 May to 16 October. The median values of a field's kriged weather variables over 139 days were used as field-level observations to minimize the effect of outliers and to arrive at a single, representative record for a field as required by the PHM data structure. For example, 210 of the 556 kriged precipitation values (139 days) were 0 for field ID 23, resulting in a highly skewed distribution (Skewness: 2.97) [5]. If there was an infestation reported on a study field, then the median of the field's weather variables for the period from the first day of observations (31 May) to the infestation date was used as data. If the field was not infested between 31 May to 16 October, the then median of the weather variable of the field from the first observation date to the last observation date of the field was used. The resulting procedure generated a single vector of median kriged values for each weather variable and each of the 47 fields.

Wind direction is a qualitative variable that ranges from 1 to 16. The 16-point directions were converted to an angle, with east set to 0 degrees. For example, for the southernmost sorghum field in Oklahoma (field ID 23), the median of the kriged wind direction is 6.44 and the angle is 124.99 degrees (a prevailing southwest to northwest direction) (Appendix Figure 4).

[5] There are 556 (139 days × 4 kriging values) of kriging results for 139 days in the ID 23 field.

Figure 2 shows the kriged medians for each weather variable. The average of the kriged median value for temperature is 24.96°C, with minimum and maximum values of 22.9 and 27°C, respectively. The average precipitation of Oklahoma for the kriged median value is 0.02 mm. The average of the kriged wind direction medians is 120.47 degrees, which correspond with a 16-point cardinal direction of 6.65 that orients toward the northwest (minimum (maximum) value, 7.13 (6.16) in the 16-point cardinal direction). A change from yellow to red in Figure 2 indicates an increase in a variable's value. In terms of temperature, the southern part of Oklahoma is relatively warm compared to other regions. For precipitation, eastern Oklahoma is comparatively wetter.

For the 16-point cardinal heading system of wind direction, '0' is north ( = 90 degrees), '4' is east ( = 0 degrees), '8' is south ( = 270 degrees), and '12' is west ( = 180 degrees) (appendix Figure 4). As the angle increases, the wind direction orients towards the west. The minimum and maximum wind direction values of 109.5 and 131.4 degrees indicate that the wind direction tended to orient toward the northwest during the study.

<< Figure 2 >>

*Sugarcane Aphid Characteristics*

Characteristics of SCA populations include the number of individuals per plant and the ratio of the number of winged SCA (alatae) to un-winged SCA (apterae) observed on a plant. The number of SCA individuals and the winged/un-winged ratio were matched to the next-infested field. That is, the winged/un-winged ratio is a temporally and spatially lagged variable.

For example, the proportion of winged to un-winged individuals in field 21 (infestation date, 14 June) was 0.33. The next infestation was observed at field 35 on June 15. Thus, the winged/un-winged ratio of 0.33 (from field 21) is assigned to field 35. This is an important variable since winged individuals can use thermal updrafts to travel to distant fields.

<< Table 2 >>

## Methods and Procedures

### *Infestation Likelihood Curves*

Time-to-infestation curves depict the probability of a sorghum field becoming infested over some period. There are no previous studies analyzing the relationship between winged and unwinged SCA populations and the likelihood of a field becoming infested. Winged individuals are relatively more resistant to starvation (Noda, 1960), have a longer reproductive period, and live longer (Tsuji & Kawada, 1987; Tsumuki et al., 1990). These factors positively affect the likelihood of migration and field infestation. The expected result is that there will be a difference in the infestation curve according to the winged/un-winged ratios. The type of infestation (or 'hazard') curves used here require assigning outcomes to discrete groups (Allison, 2010). The winged/un-winged ratio is categorized into a high "High" and "Low" categories based on the median value of 1.4%. For reference, the number of fields with the number of winged alatae less than 1 was 34 (97.1%), and the number of fields with the number of wingless alatae was 28 (80%). The number of fields with a ratio of 0 was 14 (40%), and the number of fields with a ratio of 0.1 or less was 24 (68.6%).

Regional differences in growing conditions may also accelerate or deter SCA infestation. The southwestern region is a direct route for SCA migration from Texas northward. The likelihood a field becomes infested more quickly is expected to be relatively higher in the southwestern region compared to the north-central and northwestern regions of Oklahoma. Thus, time-to-infestation curves are estimated for each region.

Finally, inter-field distance is also expected to influence the likelihood a field is infested by SCA. A time-to-infestation curve is also generated for each of the four distance variables. Inter-field distance was classified into ‘closer’ and ‘more distant’ fields using the mean values of the distances as an arbitrary cut-off. The fact that SCA (especially winged SCA) migrate over long distances and can cause rapid diffusion (Suarez et al., 2001) supports the assumption that there is a difference in the hazard of infestation according to field proximity.

The mean values of each distance covariate are 207.6 miles (distance between the last infested sorghum field in Texas and all Oklahoma sorghum fields), 116.86 miles (distance between the first sorghum field in Oklahoma and to all other fields in Oklahoma), 39.67 miles (distance between the first infested field in a region to other fields in the same region), and 43.76 miles (distance between temporally consecutive infested fields) are used to group observations into ‘Low’ and ‘High’ classes. The null hypothesis is that the distributions of the infestation curves for each group (‘Low’ and ‘High’) are not different. The null hypotheses are evaluated using the log-rank (Mantel, 1966) and Wilcoxon statistic (Wilcoxon, 1992). In this study, the PROC LIFETEST syntax of sas software is used for the infestation likelihood curve.

### *Proportional Hazard Model*

In addition to the univariate comparisons with time-to-infestation curves, a proportional hazard model (PHM, Cox, 1972) is used to determine the *ceteris paribus* effects of covariates on the likelihood of a field being infested by SCA (Allison, 2010).

PHM model the likelihood of a hazardous event occurring as a censored outcome (George et al., 2014). Censoring occurs when an event is unobserved before the study terminates. For the data collection period of this research, the first observation date (May 31) is the left censoring date, and October 16 the right censoring date. There were 12 right-censored (un-infested) fields observed over the 139 days of data collection. In Oklahoma, sorghum planting occurs between April to July with harvest following in September to November (Hawkins et al., n.d.). Infestation can occur on multiple sorghum fields during this period. This means that uninfected fields in the north-central and northwestern regions may be left- or right-censored because SCA migrate south to north. Infestation is also less likely to occur earlier than the left-censoring date of May 31. Therefore, fields where infestation did not occur during the data collection period are right-censored. Thirty-four fields were infested between 31 May and 16 October, and are recorded as interval-censored observations.

The PHM measures the hazard of a field becoming infested by SCA on day $t$:

(1) $$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

where $h(t)$ is the infestation hazard and is a function of time. The PHM assumes the effect of a covariate is multiplicative with respect to the hazard rate. The PHM, including the weather and geographic instruments, is:

(2) $$h_i(t) = h_0(t) \cdot \exp(\beta_1 \cdot l_i^p + \beta_2 \cdot l_i^n + \beta_3 \cdot \left(\frac{d_i^1}{100}\right) + \beta_4 \cdot \left(\frac{d_i^2}{100}\right) + \beta_5 \cdot \left(\frac{d_i^3}{100}\right) + \beta_6 \cdot \left(\frac{d_i^4}{100}\right) + \beta_7 \cdot ta_i + \beta_8 \cdot ra_i + \beta_9 \cdot an_i + \beta_{10} \cdot tp_i + \beta_{11} \cdot w_i)$$

where $h_0(t)$ is a baseline hazard function; $\boldsymbol{\beta}$ are parameters to be estimated; and $l_i^p$ and $l_i^n$ are regional dummy variables. The southwestern region is the reference group. The expected sign of the coefficient on the northern and western region dummy variables is negative because SCA typically move from south to north.

The variable $d_i^1$ is the distance between sorghum fields in Oklahoma and the last observed field in Texas infested by SCA. This covariate controls for northward movement of SCA from its origin as a function of distance. The variable $d_i^2$ is the distance between the first infested field in the southwest region of Oklahoma and all other Oklahoma sorghum fields. The variable $d_i^3$ is the distance between the sorghum field where SCA infestation was first reported in a region (as indicated by the dummy variables; southwest, north-central, north-east), and all other sorghum fields in the region. The variable $d_i^4$ is the distance between an infested field in period $t$ and the most recently infested field in $t-1$. The distance variables were scaled by 100 for interpretability. The expected signs of the coefficients for all distance covariates are negative, meaning that the hazard of infestation is expected to decrease as the distance between fields increases.

The variables $ta_i$, $ra_i$, and $an_i$ are temperature, precipitation, and the wind direction angle, respectively, for field $i$. These are the median values of the kriged surfaces for each field, with the median value taken over $t$ = 0 up to the infestation date. For uninfected fields, the median values are over the 139-day period.

The expected sign of temperature ($ta$) on the time until a field was infested is negative. Research finds that SCA only reproduce in the temperature range of 10 to 30°C, and that fertility and longevity decrease outside this temperature range (De Souza et al., 2019). The median field temperature over the study period was 24°C (minimum (maximum) of 19 (27) °C). This range is ideal for SCA reproduction. It is hypothesized that temperatures outside this range will negatively affect SCA survival (De Souza et al., 2019). Precipitation ($ra$) negatively affects the survivability of SCA. The expected sign of precipitation on field infestation is negative.

The expected sign of the wind direction angle ($an$) on SCA infestation is positive. An increase in the wind direction angle indicates a shift in wind direction from the northeast to northwest. This shift is expected to positively correlate with SCA migration northward.

SCA population characteristics include the temporally lagged population per sorghum plant and the temporally lagged ratio winged-total individuals per sorghum plant. The expected sign for the temporally lagged total population ($tp$) per plant is positive. Fields with a larger number of sugarcane aphid populations at $t$ – 1 are hypothesized to increase the likelihood of proximate fields becoming infested. The expected sign of the variable for the winged/un-winged ratio variable ($w$) is also positive, suggesting that proximity to a field that is infested with relatively more winged SCA in $t$ – 1 will increase the likelihood of a field becoming infested. Winged SCA are mobile over longer, field-to-field distances compared to un-winged individuals.

The hazard associated with a covariate is a log odds ratio. The ratio is calculated as the exponential of the estimated coefficient. For dummy variables, the log odds is interpreted as the ratio of the hazard for a variable with a value of '1' to the hazard for a variable with a value of '0'. That is, the hazard ratio is the difference in the likelihood of becoming infested (Motulsky, 2014). For continuous covariates, the hazard ratio is the change in infestation hazard for an increase of 1-unit in a covariate (Zwiener et al., 2011).

The percentage change in infestation for every 1-unit increase in a covariate is the exponentiated coefficient less ‘1’, times 100 (Allison, 2010).

*Proportional Hazard Assumption*

The underlying assumption of the PHM is that the relative hazard of becoming infested is constant over time. This assumption is not always justifiable, given the biological properties of SCA and geographic variability in weather (Kuitunen et al., 2021). If the effect of a covariate on a hazard varies over time, then the proportional hazard assumption is violated and statistical inference may be compromised (Nakamura, 1992).

The regional dummy variables and distance covariates are invariant with respect to time. However, weather covariates can change quickly day to day. There is also likely temporal correlation between weather today and tomorrow’s weather. In addition, growing conditions change from early summer to autumn (late May to mid-October). The day-to-day correlation of weather may be associated with changes in the infestation hazard.

The non-proportional hazard assumption is tested to determine if the time-dependent covariates of weather and population characteristics should be included in the PHM. The null hypothesis is that the effect of weather variables and SCA population characteristics on the infestation hazard is invariant over time. The alternative hypothesis is that the effects of weather and SCA population characteristics covariates on the likelihood of infestation vary over time. Time-dependent covariates are added to the regression model to test this assumption using Lin et al. (1993)’s Monte Carlo procedure. The p-value for this test was simulated using 1,000 replications.

*Ties and Exact Proportional Hazards Models*

Tied events mean that multiple fields were reported to be infected at, or around, the same time (Xin, 2011). The likelihood a coefficient estimate is biased downward increases as ties become more frequent (Allison, 2010). Of the 35 infested fields, the number of tied fields with reports of infestation occurring on the same day was 31 (88.6%). Specifically, the infestation date, the number of infested fields observed on that date, and the field ID are May 31 (number of infested fields: 2; ID: 24,27), June 16 (3; 40,44,45), June 19 (2; 25). ,26), July 10 (2; 29,39), 13 July (2; 19,22), 25 July (2; 42,47), 3 August (3; 36,37,41), 16 August (10 2,3,4,6,7,9,12,14,16,18), 29 August (2; 8,10), 30 August (3; 23,28,34) (Appendix Table 6 reports the dates, field ID, and the total SCA population per plant).

Two partial likelihood methods proposed by Efron (1977), exact and discrete, provide a better approximation of the hazard function when data are tied. The exact partial likelihood method assumes there is an unknown ordering with respect to infestation events, whereas the discrete method assumes that events occurred at exactly the same time (Allison, 2010). During the study period (May 31 to October 16; 139 days), 47 sorghum fields were surveyed at an irregular frequency for each field. For infested fields, SCA could have migrated there before the date fields were surveyed. For example, an infestation report was made on August 30 for field 23. The fields on which infestation was reported for August 30 are 19, 23, 24, 27, 28, and 34. It assumed there is an unknown ordering in these infestation events since SCA infestation may have occurred on the same date, or at an earlier date, for these six fields. The exact partial likelihood method attends to this issue and it is used here. In this study, the PROC PHREG syntax of sas software is used for the proportional hazard model.

## Results

### *Infestation Likelihood Curves*

Infestation likelihood curves demonstrate the univariate relationship between a covariate and the likelihood SCA infest a field (Figure 3). There is no statistical difference in the hazard of infestation between the "High" and "Low" variable groupings, except for the lagged ratio of the winged/un-winged ratio.

Fields in southwestern of Oklahoma region appears to have a greater chance of infestation after eight days compared to the other regions (Figure 3-A). However, the log-rank and Wilcoxon statistics are 1.58 and 1.33, respectively, which means that the null hypothesis that the infestation probability between the three regions is the same cannot be rejected. In addition, there was no statistical difference in infestation curves between the southwestern and northwestern regions (log-rank statistic: 0.05, Wilcoxon statistic: 1.32), and the southwestern and north-central regions (log-rank statistic: 1.36, Wilcoxon statistic: 0.63). These results differ from the expectation that fields in the southwestern region of Oklahoma are relatively more susceptible to infestation. This finding is consistent with previous research that finds SCA can migrate over great distances during variable weather conditions (Suarez et al., 2001).

For the lagged winged/un-winged ratio covariate, the log-rank and Wilcoxon statistics are 4.3 and 3.96, respectively. The null hypothesis that there is no difference in the infestation probability for the "High" and "Low" groups is rejected ($P < 0.05$). As shown in Figure 3-B, the infestation probability of the "High" group is higher than that of the “Low” group. Populations with lower wing/un-winged ratios have lower mobility. This result is consistent with the expectation that the higher the proportion of winged to un-winged individuals on a field, the higher the probability the next field is infested.

For the distance between the last infested sorghum field in Texas and the sorghum field in Oklahoma covariate ($d^1$), there is no difference in the probability of infestation between the “High” and “Low” groups (log-rank statistic: 0.12, Wilcoxon statistic: 0.43, P > 0.05) (Figure 3-C). There is no difference in the infestation probability between the “High” and “Low” groups for the distance between the first sorghum field infested in Oklahoma to other sorghum fields in Oklahoma ($d^2$) (log-rank and Wilcoxon statistic, 0.289 and 0.16, P > 0.05), the distance between the first infested field and other fields in the region ($d^3$) (log-rank and Wilcoxon statistics are 0.03 and 0.17, P > 0.05), and the distance between temporally consecutive infested fields covariate ($d^4$) (log-rank and Wilcoxon statistics are 1.58 and 0.16, P > 0.05) (Figure 3-D, 3-E, 3-F). The infestation probability curves calculated with each of the four distance covariates suggest that the difference in distance between fields in Oklahoma does not influence the time until a field is infested. This finding suggests that it may be difficult for sorghum producers to establish insect control plans based on field proximity. Knowing when SCA appeared in north Texas may be the most important source of information since SCA generally migrate northward. However, information on the wing/un-winged populations in the previously infested field may be useful information for establishing insect control plans and predicting where SCA will migrate.

*Proportional Hazard Regression*

Results of the proportional hazard assumption test indicate that time-dependent covariates are unnecessary (Table 3). Since the region wherein a field is located and the distance between the fields is time-invariant, violation of the proportional hazard assumption for the northwest dummy covariate is unlikely.

<< Table 3 >>

The likelihood ratio value and the Wald chi-square statistic are large enough to reject the null hypothesis that the variables included in the regression have a joint zero effect on the infestation hazard at the 1% significance level (Table 4). All covariates in the geographic relationship category were uncorrelated with the infestation hazard. The regional dummy covariates and the distance between the sorghum fields were unassociated with the infestation hazard. Studies on the long-distance migration and dispersion activity of America aphid may be one explanation that may support the results of the location and distance of the field were uncorrelated with infestation.

<< Table 4>>

Temperature, precipitation and wind direction affect the hazard of infestation at the 10%, 5% and 10% levels of significance, respectively (weather category in table 4). An increase in temperature decreases the likelihood of infestation by SCA. For a 1°C increase in temperature, the infestation hazard decreases by 50.1% ( = 100 × [exp(-0.696) – 1]). This result is consistent with the expected sign of this variable. The temperature range over the study period was 18 to 27 °C, and the median temperature was 24 °C, all of which are favorable temperatures for SCA reproduction. Studies on SCA fertility and longevity show a decrease in population outside the temperature range of 10 to 30 °C (De Souza et al., 2019). The probability of infestation for the temperature covariate was 0.33 (= odds ratio of $ta$/(1 + odds ratio of $ta$); 0.499 /(1 + 0.499)). An increase in precipitation reduces the probability of SCA infestation. A 1-mm increase in precipitation reduces the hazard of infestation by 99.99%. This suggests that precipitation reduces the hazard of infestation due to a decrease in SCA numbers caused by rain.

The more frequently wind direction changed from northeast to northwest (0 degrees, then to the east at 90 degrees), the lower was the infestation hazard. A 1-degree change in wind direction decreased the infestation hazard by 12.1% ( = 100 × [exp(-0.129) – 1]). The infestation probability for the wind direction covariate was 0.468 ( = odds ratio of $an$/(1 + odds ratio $an$); 0.868/(1 + 0.868)).

Total SCA populations were uncorrelated, ceteris paribus, with the hazard of infestation ($P > 0.05$), but the proportion of winged SCA in the most recent field was negatively associated with the likelihood of infestation ($P < 0.05$). These findings differ from the *a priori* expectations, but are consistent with the notion that SCA transmission can occur over longer distances, given favorable weather conditions.

Studies on pests and disease management strategies suggest the need for prediction models based on weather and climate (Bhagwan et al., 2022; Marini et al., 2022). The results of this study reaffirm the value of including weather data on temperature, precipitation, and wind direction to predict the movement of SCA. Brown et al. (2022) notes that the use of geospatial data in early warning systems for pest and disease risk can be used to estimate potential risks. Findings from this study suggest that geospatial data (field regions and distances between fields) may not be as important as previously suspected for the infestation of sorghum fields.

## Conclusions

This study is the first to evaluate the effects of field proximity, weather, and SCA biology on infestation at the field level for Southern Great Plains sorghum producers. The purpose of this study was to measure the effects of weather, geographic, and biological characteristics of SCA on the infestation of sorghum fields in Oklahoma. Temperature, precipitation, and wind direction were hypothesized to be closely related to the survival and migration of SCA. Geographical characteristics used in this study include between-field distance and locational information. Geographical characteristics were uncorrelated with SCA survival and migration. Total SCA population per plant was correlated with the migration of SCA and field infestation.

Infestation likelihood curves suggest there is a difference in the infestation probability between the high- and low-density wing/un-winged SCA groups. The larger the ratio of winged to un-winged SCA, the more likely their movement to other sorghum fields.

Temperature, precipitation and wind direction were negatively correlated with SCA infestation hazards. A decrease in the likelihood of infestation due to an increase in precipitation was associated with a decrease in the likelihood of a field becoming infested. Infestation hazards are lower when prevailing winds change from northeast to northwest. As mentioned in previous studies on pest infestation prediction modeling, weather is an important determinant of infestation hazard. However, field location and the distance between fields was not a significant factor in the present study.

Despite the limited number of field observations available for the analysis, a contribution of this study is the field-level analysis of SCA infestation of sorghum fields.

As mentioned by previous studies, depending on data availability, it may be possible to measure the influence of other factors that affect SCA infestation, such as infestation dynamics and food density. In addition, the infestation curve and PHM results provide field-level information about more detailed infestation hazard. These aids may improve field-level decision making for pest control planning, such as coordination of pesticide use and harvest timing.

**Table 2-1. Covariate Categories and Descriptions**

| Category | Covariate | Description | Acronym |
|---|---|---|---|
| Geographic relationship | Field location | North-central (*ncok*), North-west (*phok*), South-west (*swok*) | $l^p, l^n$ |
| | Distance | Distance (miles) between the first infested field and other fields in the region by region | $d^1$ |
| | | Distance (miles) between the last infested sorghum field in Texas and the sorghum field in Oklahoma | $d^2$ |
| | | Distance (miles) between the first sorghum field in Oklahoma and other fields in Oklahoma | $d^3$ |
| | | Distance (miles) between sorghum fields infested in sequential time order | $d^4$ |
| Weather | Temperature | Temperature (°C) in Sorghum Field (average of all 5-minute averaged temperature observations each day) | $ta$ |
| | Precipitation | Precipitation (mm) in Sorghum Field (liquid precipitation measured each day) | $ra$ |
| | Wind direction | Wind direction angle (East: 0 degrees) in Sorghum Field (most common wind direction converted to degrees) | $an$ |
| Sugarcane aphid population characteristics | Total population of SCA | Number of sugarcane aphids per plant | $tp$ |
| | Winged SCA ratio | Proportion of winged sugarcane aphid to total population of sugarcane aphid per plant | $w$ |

Source: Dr. Norman Elliott, USDA-ARS (personal conversation, 2021), Oklahoma Mesonet (https://www.mesonet.org)

**Table 2-2. Covariate Descriptive Statistics (N = 47 fields)**

| | Mean | Standard deviation | Min | Max | Range |
|---|---|---|---|---|---|
| Infestation | 0.72 | 0.45 | 0 | 1 | 1 |
| Period (days) | 49.38 | 28.01 | 0 | 92 | 92 |
| Distance to last field infested in Texas (miles) | 207.61 | 72.14 | 105.46 | 317.89 | 212.43 |
| Distance to first field infested in Oklahoma (miles) | 116.87 | 66.27 | 0 | 260.77 | 260.77 |
| Distance between the first infested field in a region and other fields in that region (miles) | 39.67 | 30.41 | 0 | 106.47 | 106.47 |
| Distance between temporally consecutive infested fields (miles) | 43.76 | 66.25 | 0 | 290.44 | 290.44 |
| Population per plant | 11.66 | 75.60 | 0 | 518.78 | 518.78 |
| Winged SCA ratio | 0.05 | 0.09 | 0 | 0.33 | 0.33 |
| TAVG (°C) | 24.96 | 1.06 | 22.90 | 27.00 | 4.10 |
| RAIN (mm) | 0.02 | 0.01 | 0 | 0.04 | 0.04 |
| PDIR (degrees) | 120.47 | 6.02 | 109.50 | 131.40 | 21.90 |

**Note**: Infestation is indicated as '1' if the field is infested (the number of sugarcane aphid populations per plant is '0' or more) during the infestation investigation period, otherwise '0'. TAVG, RAIN, wind direction are the median of temperature, rainfall, and angle of wind direction variables from the first infestation date in Oklahoma to the first infestation date.

Sources: Dr. Norman Elliott, USDA-ARS (personal communication, 2021), Oklahoma Mesonet (https://www.mesonet.org)

**Table 2-3. Proportional Hazards Test Results (N = 47)**

| Category | Covariate | Coefficient Estimate | P-value |
|---|---|---|---|
| Geographic relationship | North-west region | 2.36 | 0.04 |
| | North-central region | 4.06 | 0.18 |
| | Distance between the last infested sorghum field in Texas and the sorghum field in Oklahoma | 4.80 | 0.17 |
| | Distance between the first sorghum field in Oklahoma and other fields in Oklahoma | 5.10 | 0.32 |
| | Distance between the first infested field and other fields in the region | 2.52 | 0.45 |
| | Distance between temporally consecutive infested fields | 1.26 | 0.15 |
| Weather | Temperature in Sorghum Field | 1.49 | 0.42 |
| | Precipitation in Sorghum Field | 1.65 | 0.13 |
| | Wind direction in sorghum field | 1.55 | 0.17 |
| Sugarcane aphid population characteristics | Total population of SCA | 0.16 | 0.35 |
| | Winged SCA ratio | 1.32 | 0.20 |

Table 2-4. Cox Proportional Hazards Model Results (N = 47)

| Category | Covariate | Estimate | Hazard ratio | Infestation Probability |
|---|---|---|---|---|
| Geographic relationship | North-west region | -1.126 | 0.324 | 0.245 |
| | North-central region | 0.536 | 1.710 | 0.631 |
| | Distance between the last infested sorghum field in Texas and the sorghum field in Oklahoma | -0.942 | 0.390 | 0.281 |
| | Distance between the first sorghum field in Oklahoma and other fields in Oklahoma | 0.510 | 1.665 | 0.625 |
| | Distance between the first infested field and other fields in the region | -1.936 | 0.144 | 0.126 |
| | Distance between temporally consecutive infested fields | 0.398 | 1.489 | 0.598 |
| Weather | Temperature | -0.696* | 0.499 | 0.333 |
| | Precipitation | -10.570** | 0.000 | 0.000 |
| | Wind direction | -0.129* | 0.879 | 0.468 |
| Sugarcane aphid population characteristics | Total SCA population | -0.039* | 0.962 | 0.490 |
| | Winged SCA ratio | -4.186 | 0.015 | 0.015 |
| Likelihood Ratio | | 75.62*** | | |
| Wald $\chi^2$ | | 117.83*** | | |
| Likelihood | | 66.57 | | |
| AIC | | 88.58 | | |
| SBC | | 105.69 | | |

**Note**: ***, **, and *, significant at the 1%, 5%, and 10% significance levels, respectively.

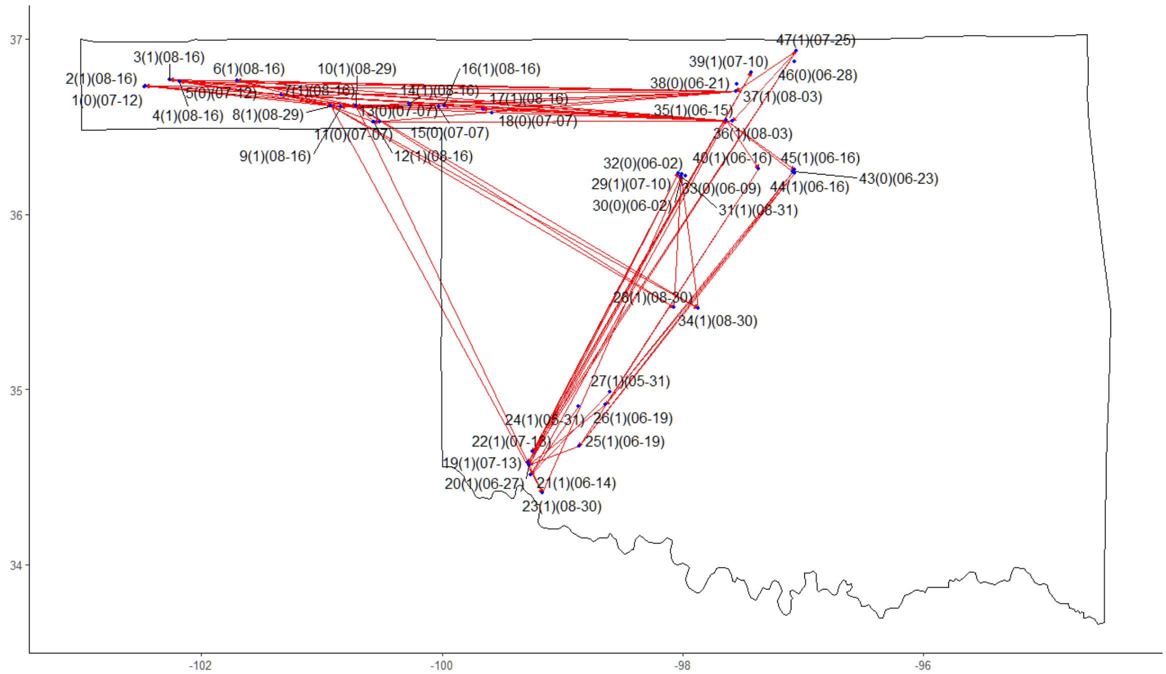


**Figure 2-1. Oklahoma Sorghum Field Location used for Analysis, Infestation Status, and Infestation Date (or last reported date).**

**Note:** The number in each field indicates the field's identifier, if the field was infested ('1' if SCA present in the field, '0' otherwise), and the date of infestation. The red line indicates the chronological connection of the fields infested by SCA.

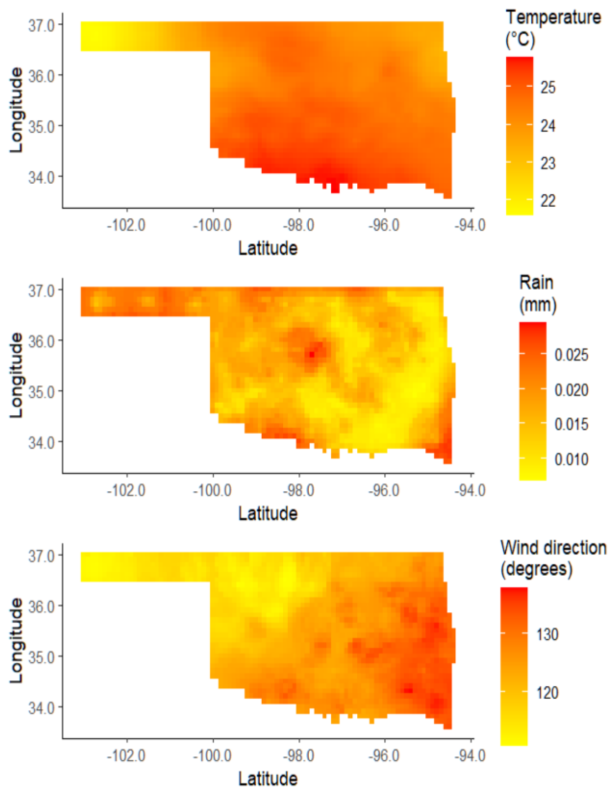


**Figure 2-2. Median Kriged Values for Temperature, Precipitation, and Wind Direction.**

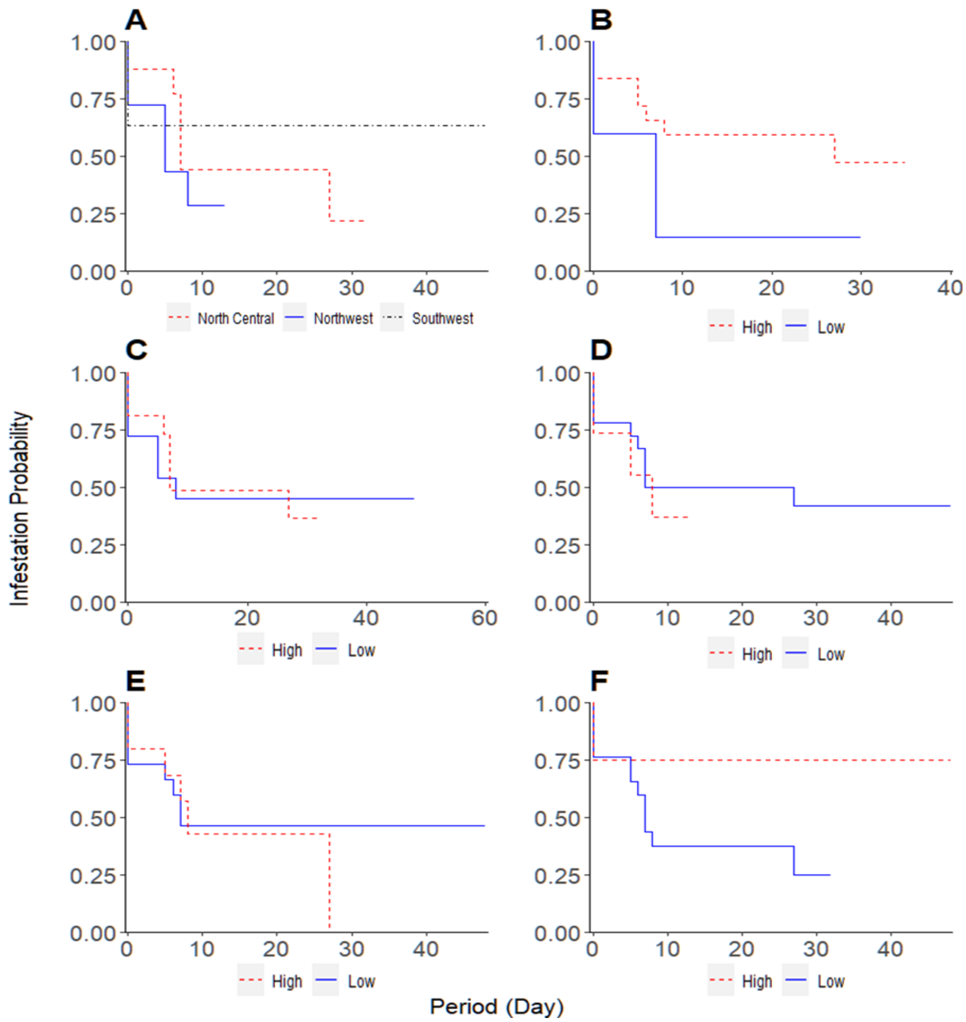


**Figure 2-3. Sugarcane Aphid Infestation Curves Between Groups of Region Dummy Variable (A), Lagged Wing Ratio (B), Distance between the Last Infested Sorghum Field in Texas and the Sorghum Field in Oklahoma (C), Distance between the First Sorghum Field Infested in Oklahoma to Other Sorghum Fields in Oklahoma (D), Distance between the First Infested Field and Other Fields in the Region (E), Distance between Temporally Consecutive Infested Fields (F) Covariates**

# REFERENCES


Acreman, S. J., & Dixon, A. F. G. (1989). The Effects of Temperature and Host Quality on the Rate of Increase of the Grain Aphid (sitobion Avenae) on Wheat. *Annals of Applied Biology*, *115*(1), 3–9. https://doi.org/10.1111/j.1744-7348.1989.tb06805.x

Allison, P. D. (2010). *Survival analysis using SAS: a practical guide*. Sas Institute.

Angilletta, Jr., M. J., & Dunham, A. E. (2003). The Temperature-Size Rule in Ectotherms: Simple Evolutionary Explanations May Not Be General. *The American Naturalist*, *162*(3), 332–342. https://doi.org/10.1086/377187

Angilletta, M. J. (2004). Temperature, Growth Rate, and Body Size in Ectotherms: Fitting Pieces of a Life-History Puzzle. *Integrative and Comparative Biology*, *44*(6), 498–509. https://doi.org/10.1093/icb/44.6.498

Asin, L., & Pons, X. (2001). Effect of High Temperature on the Growth and Reproduction of Corn Aphids (Homoptera: Aphididae) and Implications for Their Population Dynamics on the Northeastern Iberian Peninsula. *Environmental Entomology*, *30*(6), 1127–1134. https://doi.org/10.1603/0046-225X-30.6.1127

Auad, A. M., Alves, S. O., Carvalho, C. A., Silva, D. M., Resende, T. T., & Veríssimo, B. A. (2009). The Impact of Temperature on Biological Aspects and Life Table of *Rhopalosiphum padi* (Hemiptera: Aphididae) Fed with Signal Grass. *Florida Entomologist*, *92*(4), 569–577. https://doi.org/10.1653/024.092.0406

Bale, J. S., Masters, G. J., Hodkinson, I. D., Awmack, C., Bezemer, T. M., Brown, V. K., Butterfield, J., Buse, A., Coulson, J. C., Farrar, J., Good, J. E. G., Harrington, R., Hartley, S., Jones, T. H., Lindroth, R. L., Press, M. C., Symrnioudis, I., Watt, A. D., & Whittaker, J. B. (2002). Herbivory in Global Climate Change Research: Direct Effects of Rising Temperature on Insect Herbivores. *Global Change Biology*, *8*(1), 1–16. https://doi.org/10.1046/j.1365-2486.2002.00451.x

Behura, B. K., & Bohidar, K. (1983). Effect of Temperature on the Fecundity of Five Species of Aphids. *Pranikee*, *4*, 23–27.

Bhagwan, D., Chahal, K., Minal, S., & Parashurama, Dr. T. R. (2022). *Advances in Plant Science Volume IV*.

Bowling, R., Brewer, M. J., Biles, S., & Gordy, J. (2015). Occurrence of Sugarcane Aphid in the United States and Mexico with Reference to Occurrence in 2013 and 2014. *27th Annual Texas Plant Protection Conference, Bryan, TX. 8-9 Dec*.

Brooker, P. I. (1986). A Parametric Study of Robustness of Kriging Variance as a Function of Range and Relative Nugget Effect for a Spherical Semivariogram. *Mathematical Geology*, *18*(5), 477–488.

Brown, M. E., Mugo, S., Petersen, S., & Klauser, D. (2022). Designing a Pest and Disease Outbreak Warning System for Farmers, Agronomists and Agricultural Input Distributors in East Africa. *Insects*, *13*(3), 232.

Burgess, T. M., & Webster, R. (1980). Optimal Interpolation and Isarithmic Mapping of Soil Properties: I the Semi-Variogram and Punctual Kriging. *Journal of Soil Science*, *31*(2), 315–331.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (Vol. 7819, pp. 160–172). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37456-2_14

Chen, C., Harvey, J. A., Biere, A., & Gols, R. (2019). Rain Downpours Affect Survival and Development of Insect Herbivores: The Specter of Climate Change? *Ecology*, *100*(11). https://doi.org/10.1002/ecy.2819

Cocu, N., Harrington, R., Rounsevell, M. D. A., Worner, S. P., Hulle, M., & Participants, E. P. (2005). Geographical Location, Climate and Land Use Influences on the Phenology and Numbers of the Aphid, Myzus Persicae, in Europe. *Journal of Biogeography*, *32*(4), 615–632.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

De Souza, M. A., Armstrong, J. S., Hoback, W. W., Mulder, P. G., Paudyal, S., Foster, J. E., Payton, M. E., & Akosa, J. (2019a). *Temperature dependent development of sugarcane aphids Melanaphis Sacchari,(Hemiptera: Aphididae) on three different host plants with estimates of the lower and upper threshold for fecundity*.

De Souza, M. A., Armstrong, J. S., Hoback, W. W., Mulder, P. G., Paudyal, S., Foster, J. E., Payton, M. E., & Akosa, J. (2019b). *Temperature Dependent Development of Sugarcane Aphids Melanaphis Sacchari,(hemiptera: Aphididae) on Three Different Host Plants with Estimates of the Lower and Upper Threshold for Fecundity*.

*EDDMapS*. (2020). EDDMapS.Org. https://www.eddmaps.org/index.cfm

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, *72*(359), 557–565.

George, B., Seals, S., & Aban, I. (2014). Survival Analysis and Regression Models. *Journal of Nuclear Cardiology : Official Publication of the American Society of Nuclear Cardiology*, *21*(4), 686–694. https://doi.org/10.1007/s12350-014-9908-2

Gonçalves, M. C. (2005). *Principais viroses da cana-de-açúcar*.

Hawkins, S., Williams, E., Massey, B., & Hodges, M. (n.d.). Grain Sorghum Production Calendar. *Oklahoma Cooperative Extension Service*, 4.

Huberty, A. F., & Denno, R. F. (2004). Plant Water Stress and Its Consequences for Herbivorous Insects: A New Synthesis. *Ecology*, *85*(5), 1383–1398. https://doi.org/10.1890/03-0352

Irwin, M. E., Kampmeier, G. E., & Weisser, W. W. (2007). Aphid Movement: Process and Consequences. In H. F. van Emden & R. Harrington (Eds.), *Aphids as crop pests* (pp. 153–186). CABI. https://doi.org/10.1079/9780851998190.0153

Isard, S. A., & Gage, S. H. (2001). *Flow of Life in the Atmosphere*. Michigan State University Press.

Isard, S. A., Irwin, M. E., Carter, M., & Holtzer, T. O. (1994). Temperature Stratification and Insect Layer Concentrations: A Preliminary Analysis of Atmospheric Measurements and Concurrent Aerial Insect Collections from Northeastern Colorado and East Central Illinois. *Preprint Volume of the 21st Conference on Agricultural and Forest Meteorology and the 11th Conference on Biometeorology and Aerobiology*, 407–410.

Klueken, A. M., Hau, B., Ulber, B., & Poehling, H.-M. (2009). Forecasting Migration of Cereal Aphids (hemiptera: Aphididae) in Autumn and Spring. *Journal of Applied Entomology*, *133*(5), 328–344. https://doi.org/10.1111/j.1439-0418.2009.01387.x

Knutson, A., Bowling, R., Brewer, M., Bynum, E., & Porter, P. (2016a). The Sugarcane Aphid: Management Guidelines for Grain and Forage Sorghum in Texas. *NTO-035, Texas A&M AgriLife Extension, College Station, TX*.

Knutson, A., Bowling, R., Brewer, M., Bynum, E., & Porter, P. (2016b). Management Guidelines for Grain and Forage Sorghum in Texas. *Texas A&M Agrilife Extension*, 8.

Kobori, Y., & Amano, H. (2003). Effect of Rainfall on a Population of the Diamondback Moth, Plutella Xylostella (lepidoptera: Plutellidae). *Applied Entomology and Zoology*, *38*(2), 249–253. https://doi.org/10.1303/aez.2003.249

Kuitunen, I., Ponkilainen, V. T., Uimonen, M. M., Eskelinen, A., & Reito, A. (2021). Testing the Proportional Hazards Assumption in Cox Regression and Dealing with Possible Non-Proportionality in Total Joint Arthroplasty Research: Methodological Perspectives and Review. *BMC Musculoskeletal Disorders*, *22*(1), 489. https://doi.org/10.1186/s12891-021-04379-2

Kukal, M., & Irmak, S. (2016). Long-Term Patterns of Air Temperatures, Daily Temperature Range, Precipitation, Grass-Reference Evapotranspiration and Aridity Index in the Usa Great Plains: Part Ii. Temporal Trends. *Journal of Hydrology*, *542*, 978–1001. https://doi.org/10.1016/j.jhydrol.2016.06.008

Lin, D. Y., WEI, L. J., & YING, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika*, *80*(3), 557–572. https://doi.org/10.1093/biomet/80.3.557

Lopes, J. R. S., Vega, J., Gonçalves, M. C., Krügner, R., & Navas, S. M. (1997). Aphid Transmission of a Virus Associated with Sugarcane Yellow Leaf Disease. *Fitopatologia Brasileira*, *22*, 335.

Mann, J. A., Tatchell, G. M., Dupuch, M. J., Harrington, R., Clark, S. J., & McCARTNEY, H. A. (1995). Movement of Apterous Sitobion Avenae (homoptera: Aphididae) in Response to Leaf Disturbances Caused by Wind and Rain. *Annals of Applied Biology*, *126*(3), 417–427. https://doi.org/10.1111/j.1744-7348.1995.tb05376.x

Mantel, N. (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration. *Cancer Chemother Rep*, *50*(3), 163–170.

Marini, G., Rizzoli, A., & Tagliapietra, V. (2022). Predicting Rodent Population Dynamics as Early Warning for Zoonotic Disease Transmission. *International Journal of Infectious Diseases*, *116*, S70.

*Mesonet*. (2020). https://www.mesonet.org/index.php

Michael A., D. S. (2019). Temperature Dependent Development of Sugarcane Aphids Melanaphis Sacchari, (Hemiptera: Aphididae) on Three Different Host Plants with Estimates of the Lower and Upper Threshold for Fecundity. *Current Trends in Entomology and Zoological Studies*, *2*(1). https://doi.org/10.29011/2690-0114.100011

Michael A., D. S., & J. Scott, A. (2019). Temperature Dependent Development of Sugarcane Aphids Melanaphis Sacchari, (Hemiptera: Aphididae) on Three Different Host Plants with Estimates of the Lower and Upper Threshold for Fecundity. *Current Trends in Entomology and Zoological Studies*, *2*(1). https://doi.org/10.29011/2690-0114.100011

Mote, U. N. (1983). Epidemic of Delphacids and Aphids on Winter Sorghum. *Sorghum Newsl*, *26*, 76.

Motulsky, H. (2014). *Intuitive biostatistics: A nonmathematical guide to statistical thinking*. Oxford University Press, USA.

Nakamura, T. (1992). Proportional Hazards Model with Covariates Subject to Measurement Error. *Biometrics*, *48*(3), 829–838. https://doi.org/10.2307/2532348

Nibouche, S., Costet, L., Holt, J. R., Jacobson, A., Pekarcik, A., Sadeyen, J., Armstrong, J. S., Peterson, G. C., McLaren, N., & Medina, R. F. (2018). Invasion of sorghum in the Americas by a new sugarcane aphid (Melanaphis sacchari) superclone. *PLOS ONE*, *13*(4), e0196124. https://doi.org/10.1371/journal.pone.0196124

Noda, I. (1960). The Emergence of Winged Viviparous Female in Aphid Vi. Difference in the Rate of Development Between the Winged and the Unwinged Forms. *Japanese Journal of Ecology*, *10*(3), 97–102.

Paramasivam, C. R., & Venkatramanan, S. (2019). An Introduction to Various Spatial Analysis Techniques. *GIS and Geostatistical Techniques for Groundwater Science*, 23–30.

Paray, N. B., Khoodoo, M. H. R., Saumtally, A. S., & Ganeshan, S. (2011). Vector-Virus Relationship for Melanaphis Sacchari (zehnt.)(hemiptera: Aphididae) Transmitting Sugarcane Yellow Leaf Luteovirus in Mauritius. *Sugar Tech*, *13*(1), 77–80.

Rodríguez-del-Bosque, L. A., Silva-Serna, M. M., & Aranda-Lara, U. (2020). Effect of Natural and Simulated Rainfall and Wind on Melanaphis Sacchari1 on Sorghum. *Southwestern Entomologist*, *45*(2), 357. https://doi.org/10.3958/059.045.0204

Setokuchi, O. (1977). Ecology of Longiunguis Sacchari Zehntner (aphididae) Infesting Sorghums, 5: Influence of Harvesting Time and Plant Population on the Aphid Occurrence. *Proceedings of the Association for Plant Protection of Kyushu*.

Shao, X., Zhang, Q., Liu, Y., & Yang, X. (2020). Effects of Wind Speed on Background Herbivory of an Insect Herbivore. *Écoscience*, *27*(1), 71–76. https://doi.org/10.1080/11956860.2019.1666549

Singh, B. U., Padmaja, P. G., & Seetharama, N. (2004). Biology and Management of the Sugarcane Aphid, Melanaphis Sacchari (zehntner) (homoptera: Aphididae), in Sorghum: A Review. *Crop Protection*, *23*(9), 739–755. https://doi.org/10.1016/j.cropro.2004.01.004

Souza, M. F., & Davis, J. A. (2020). Potential Population Growth of Melanaphis Sacchari (zehntner) (hemiptera: Aphididae) Under Six Constant Temperatures on Grain Sorghum (sorghum Bicolor L.). *Florida Entomologist*, *103*(1), 116. https://doi.org/10.1653/024.103.0419

Suarez, A. V., Holway, D. A., & Case, T. J. (2001). Patterns of Spread in Biological Invasions Dominated by Long-Distance Jump Dispersal: Insights from Argentine Ants. *Proceedings of the National Academy of Sciences*, *98*(3), 1095–1100.

*Texas A&M AgriLife Extension*. (2020). Texas A&M AgriLife Extension Service. https://agrilifeextension.tamu.edu/

Tsuji, H., & Kawada, K. (1987). Effects of Starvation on Life Span and Embryo Development of Four Morphs of Pea Aphid (acyrthosiphon Pisum (harris)). *Japanese Journal of Applied Entomology and Zoology (Japan)*.

Tsumuki, H., Nagatsuka, H., Kawada, K., & Kanehisa, K. (1990). Comparison of Nutrient Reservation in Apterous and Alate Pea Aphids, Acyrthosiphon Pisum (harris): 1. Developmental Time and Sugar Content. *Applied Entomology and Zoology*, *25*(2), 215–221.

Ukoroije, R. B., & Abalis, R. O. (2020). Population Dynamics and Seasonal Abundance of the Beetle Oryctes owariensis Beavois, 1807 (Coleoptera: Scarabaeidae: Dynastinae) in Raphia Palms in Bayelsa State, Nigeria. *Asian Journal of Biology*, 14–27. https://doi.org/10.9734/ajob/2020/v9i330087

*USDA NASS*. (2020). https://www.nass.usda.gov/

Van Rensburg, N. J. (1973). Population Fluctuations of the Sorghum Aphid Melanaphis (longiunguis) Pyrarius (passerini) Forma Sacchari (zehntner). *Phytophylactica*, *5*(4), 127–134.

Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in statistics* (pp. 196–202). Springer.

Wulder, M. A., White, J. C., Bentz, B., Alvarez, M. F., & Coops, N. C. (2006). Estimating the Probability of Mountain Pine Beetle Red-Attack Damage. *Remote Sensing of Environment*, *101*(2), 150–166. https://doi.org/10.1016/j.rse.2005.12.010

Xin, X. (2011). *A study of ties and time-varying covariates in cox proportional hazards model* [PhD Thesis]. University of Guelph.

Zapata, S. D., Dudensing, R., Sekula, D., Esparza-Díaz, G., & Villanueva, R. (2018). Economic Impact of the Sugarcane Aphid Outbreak in South Texas. *Journal of Agricultural and Applied Economics*, *50*(1), 104–128. https://doi.org/10.1017/aae.2017.24

Zhang, Z., Zhang, J., & Xue, H. (2008). Improved K-Means Clustering Algorithm. *2008 Congress on Image and Signal Processing*, 169–172. https://doi.org/10.1109/CISP.2008.350

Zwiener, I., Blettner, M., & Hommel, G. (2011). Survival analysis: Part 15 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, *108*(10), 163.

Norman Elliott. (2021). Personal communication, 2021

**Table 1-4. Summary Statistics for Weather in Oklahoma, 2013-2020**

| Variable | Statistics | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|
| TAVG (°F) | Mean | 79.86 | 78.24 | 78.76 | 83.02 | 78 | 78.75 | 76.56 | 77.47 |
| | S.D. | 4.07 | 3.37 | 3.83 | 2.36 | 4.14 | 4.45 | 4.82 | 3.32 |
| | Min | 68.28 | 68.55 | 68.62 | 73.45 | 66.67 | 66.25 | 64.1 | 67.52 |
| | Max | 90.28 | 87.79 | 86.63 | 89.57 | 88.14 | 89.38 | 90.5 | 86.08 |
| PDIR (16-point) | Mean | 6.94 | 7.57 | 7.79 | 7.45 | 5.65 | 7.41 | 7.78 | 6.66 |
| | S.D. | 1.97 | 1.91 | 2.79 | 1.53 | 2.54 | 2.72 | 2.72 | 2.08 |
| | Min | 0 | 0 | 0 | 0.5 | 0 | 0 | 0 | 0 |
| | Max | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| WSPD (miles/hours) | Mean | 10.52 | 9.95 | 8.95 | 8.36 | 7.91 | 9.34 | 8.15 | 8.52 |
| | S.D. | 4.71 | 4.65 | 3.63 | 3.04 | 3.29 | 3.8 | 3.78 | 3.97 |
| | Min | 2.85 | 1.92 | 2.49 | 2.49 | 1.6 | 2.47 | 1.58 | 2.2 |
| | Max | 24.55 | 23.54 | 20.22 | 17.71 | 18.29 | 21.72 | 22.23 | 23.16 |
| RAIN (inches) | Mean | 0.13 | 0.16 | 0.29 | 0.04 | 0.09 | 0.26 | 0.26 | 0.16 |
| | S.D. | 0.32 | 0.4 | 0.79 | 0.23 | 0.29 | 0.55 | 0.55 | 0.39 |
| | Min | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Max | 2.48 | 3.34 | 7.98 | 4.43 | 2.37 | 4.74 | 5.34 | 3.14 |

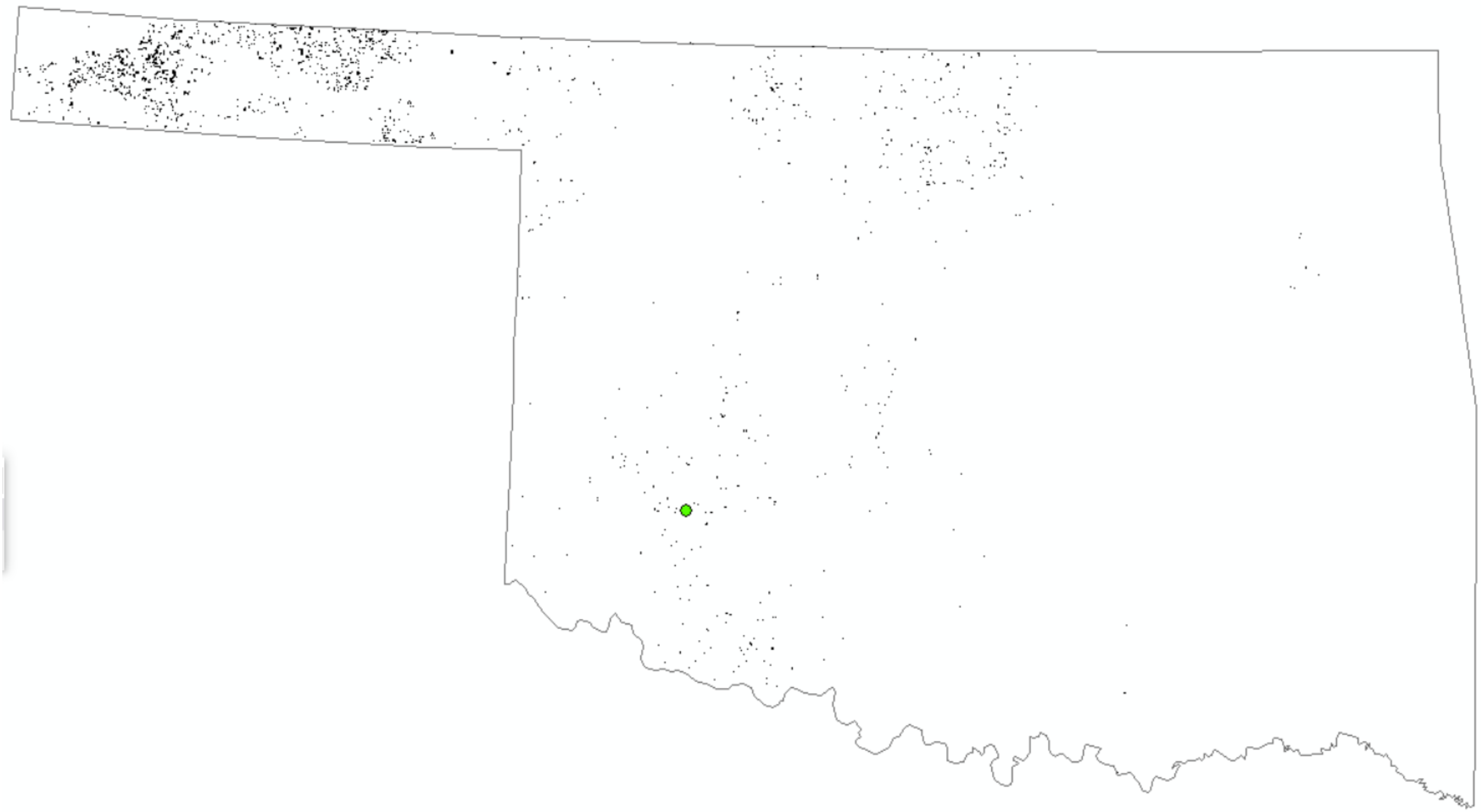

**Figure 1-17. Oklahoma's Sorghum Area Distribution (the larger the area, the lower the brightness) and Kiowa County's Centroid (green dot)**

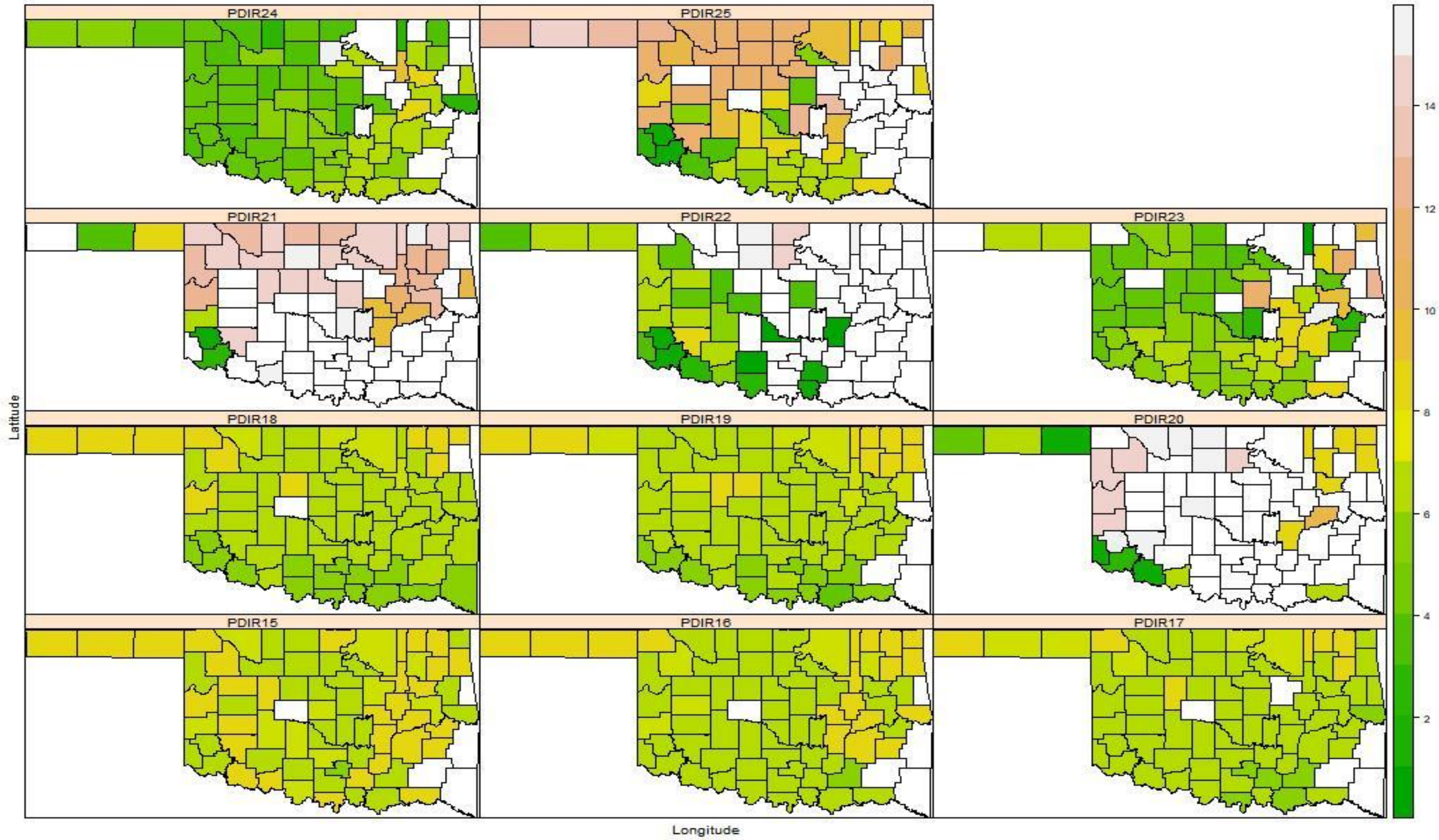


**Figure 1-18. PDIR *D*istribution for 10 days after June 15, 2018 in Oklahoma (white area is missing data)**

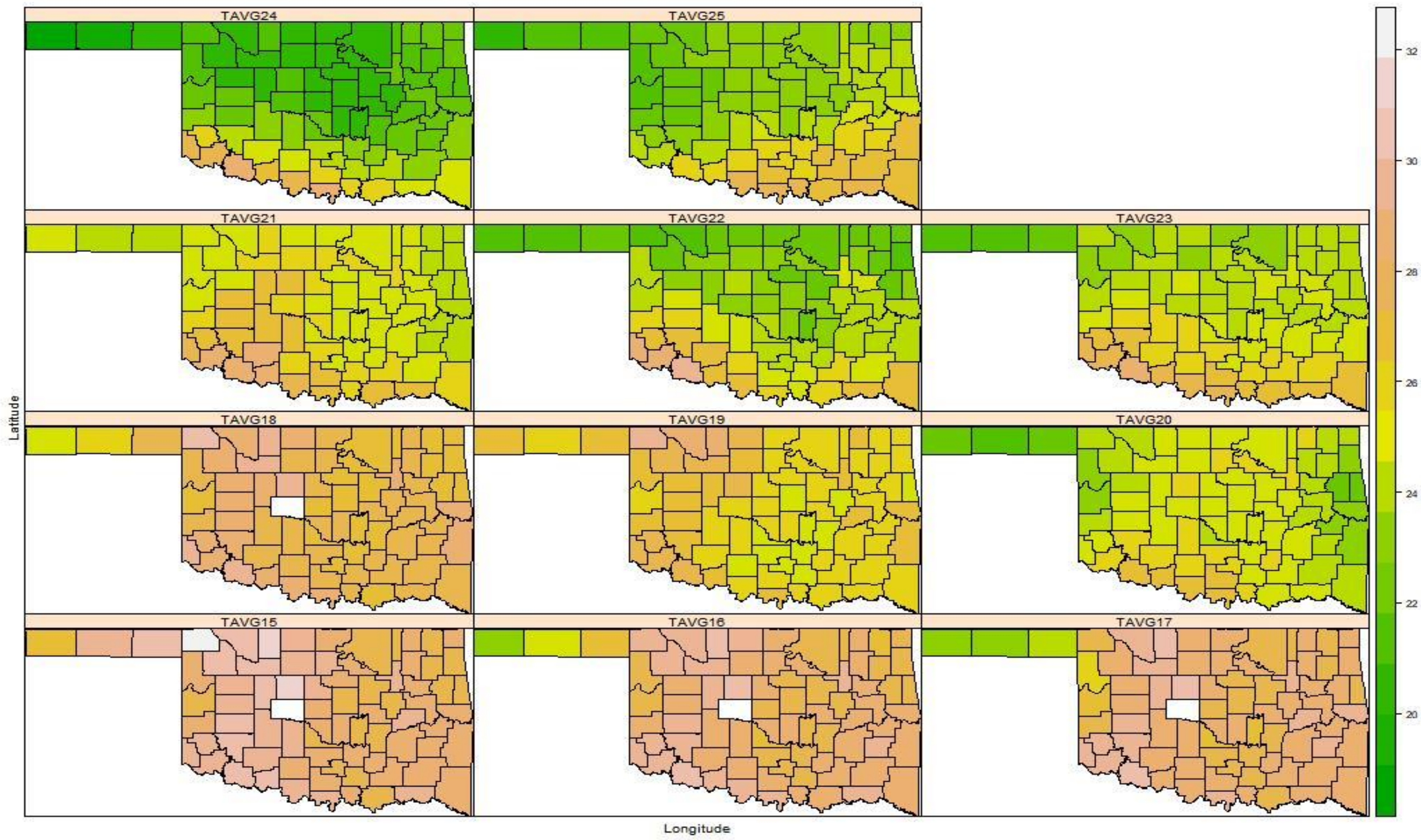


**Figure 1-19. TAVG Distribution for 10 days after June 15, 2018 in Oklahoma (white area is missing data)**

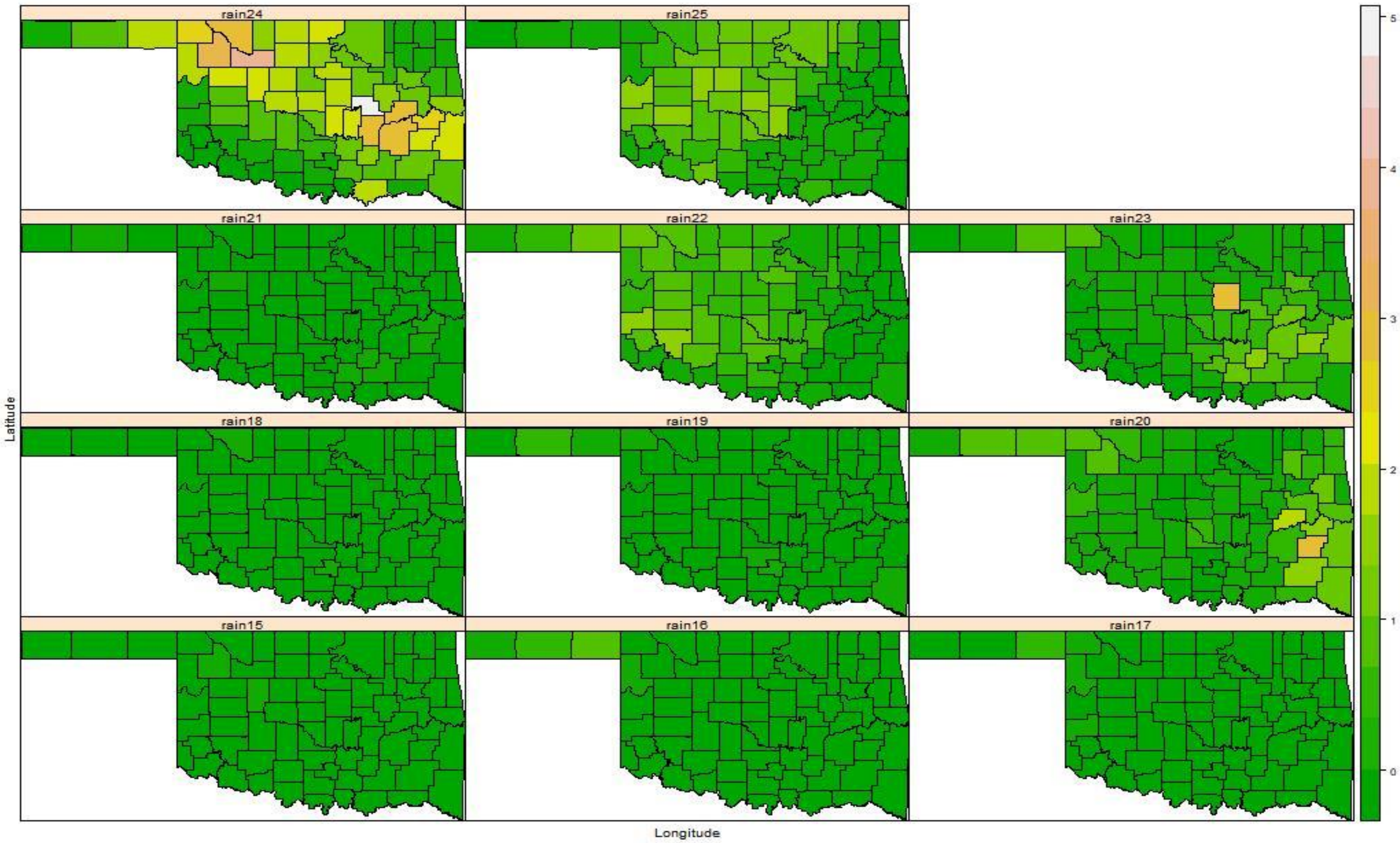


**Figure 1-20. RAIN Distribution for 10 days after June 15, 2018 in Oklahoma (white area is missing data)**

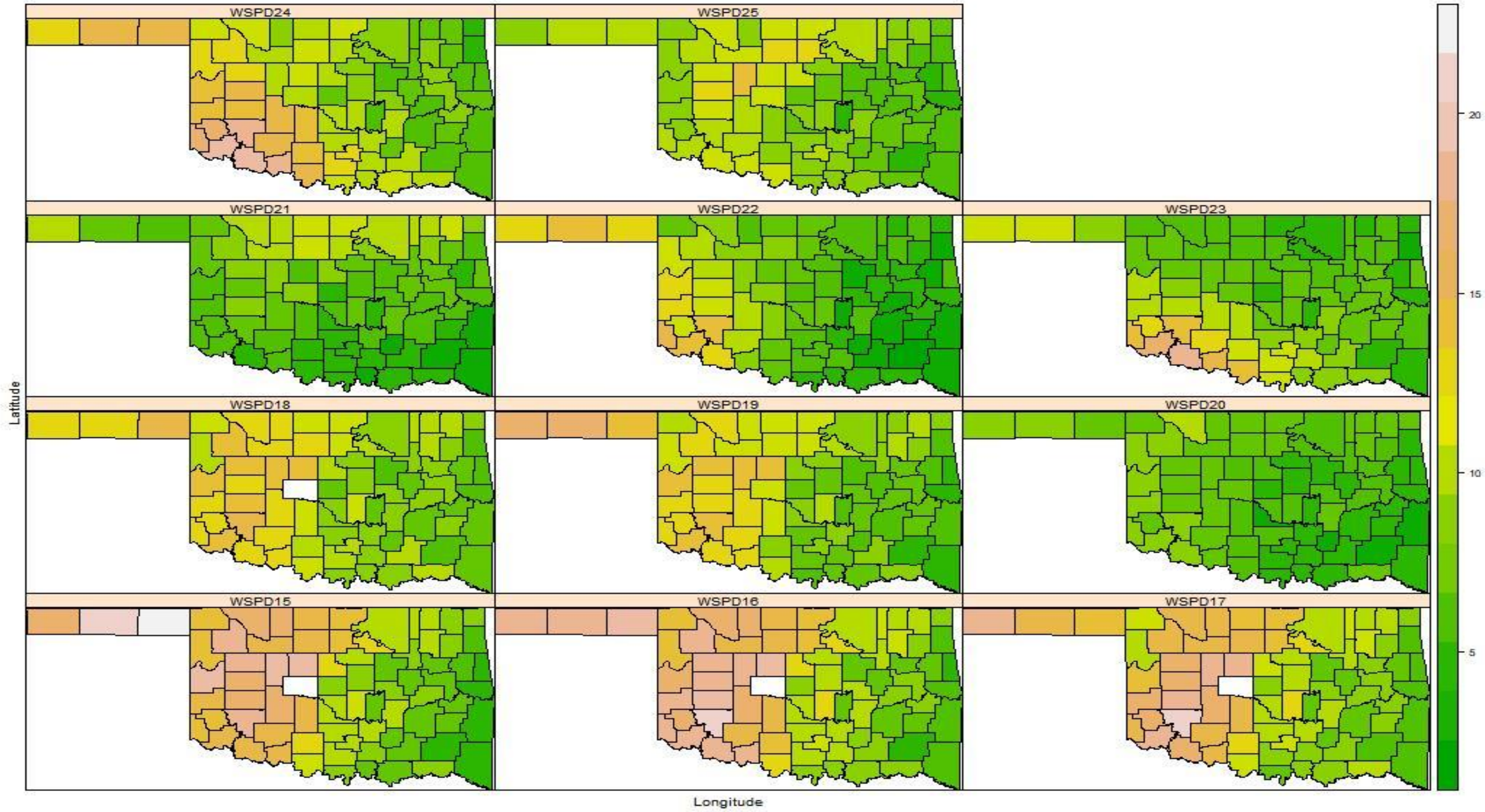


**Figure 1-21. TAVG Distribution for 10 days after June 15, 2018 in Oklahoma (white area is missing data)**

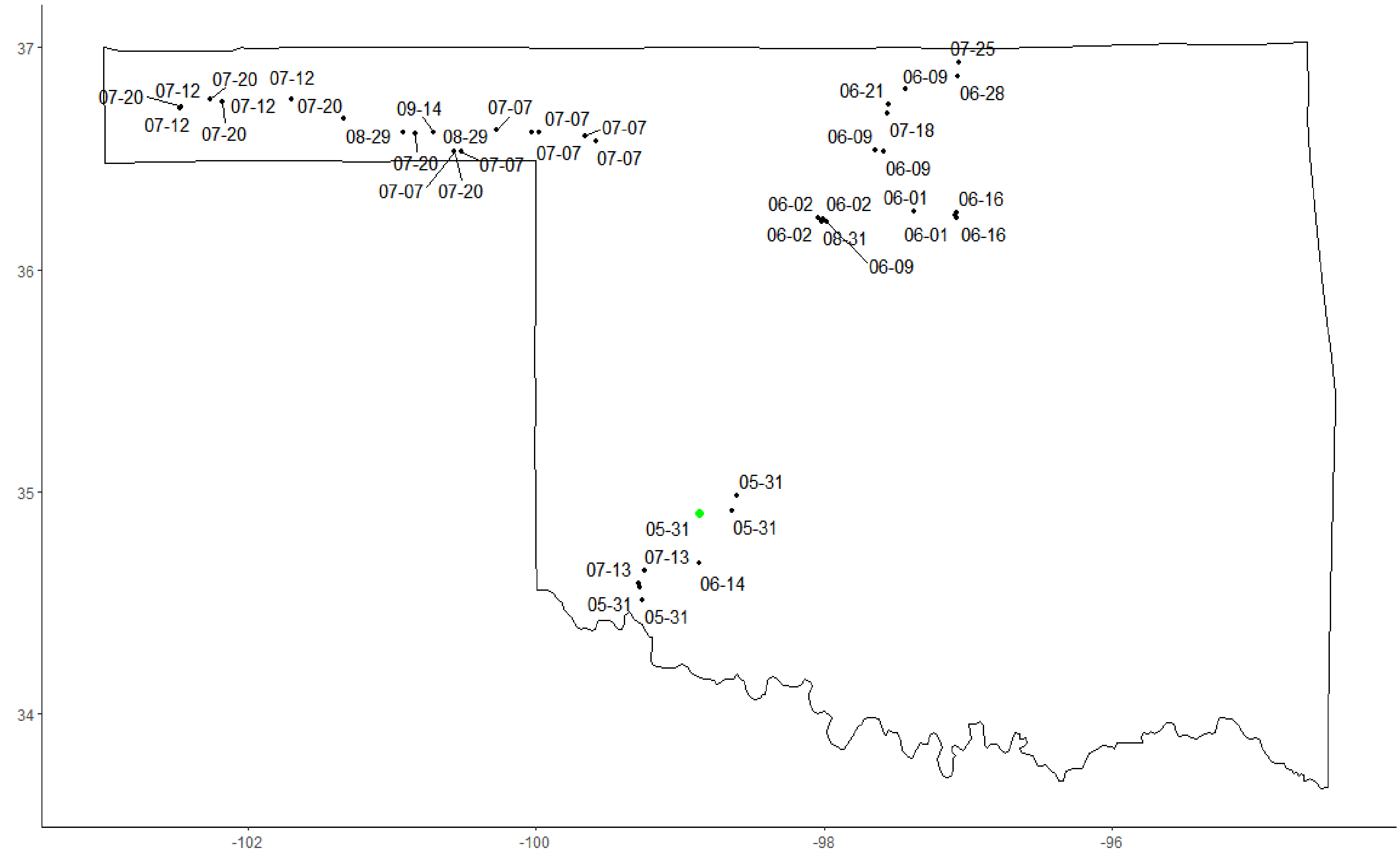


**Figure 1-22. Oklahoma's 2017 Infection Survey Field and First Survey Date**

**Note: The green dot is the initial location used in the moving probability model.**

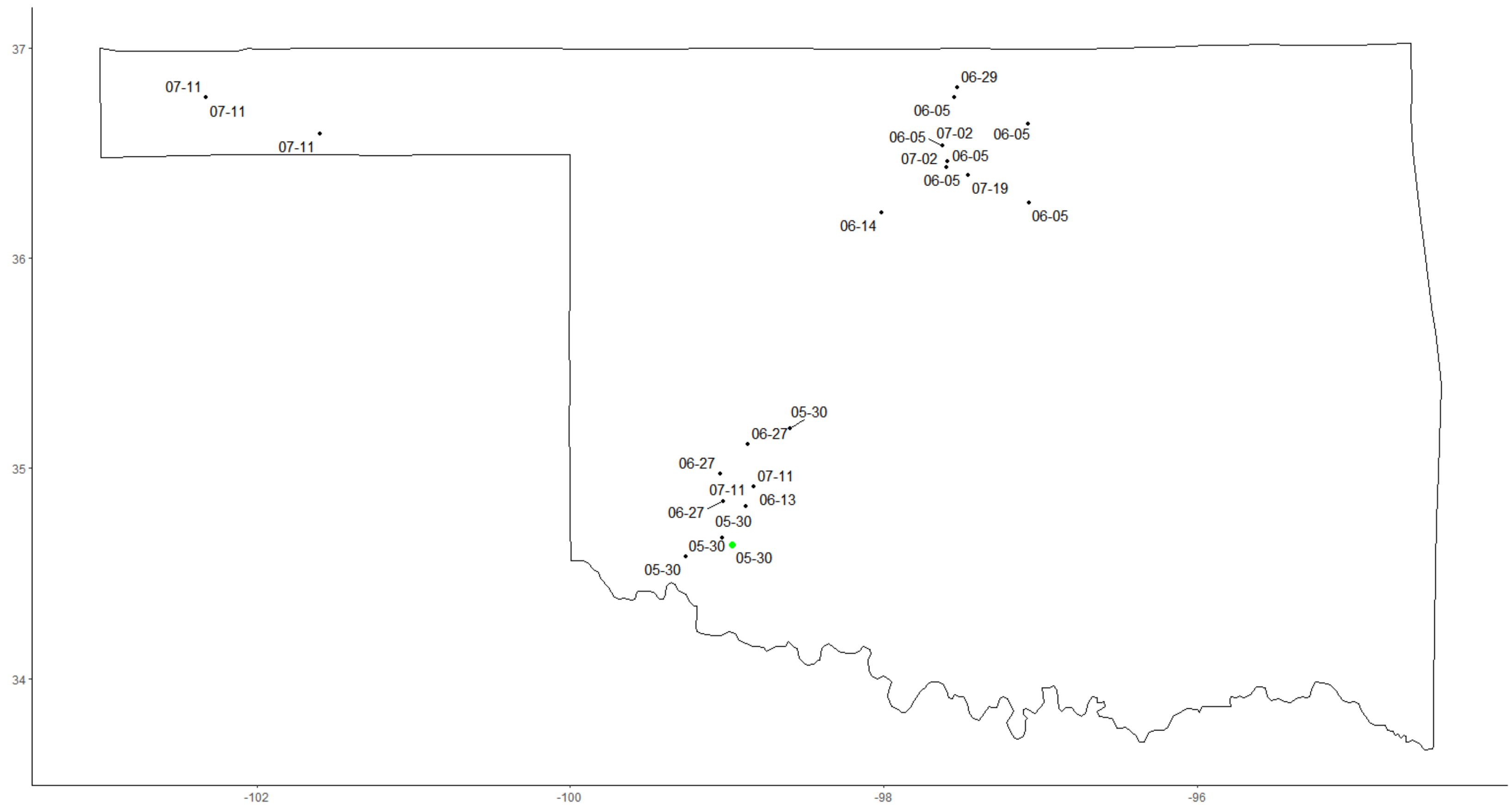


**Figure 1-23. Oklahoma's 2018 Infection Survey Field and First Survey Date**

**Note: The green dot is the initial location used in the moving probability model.**

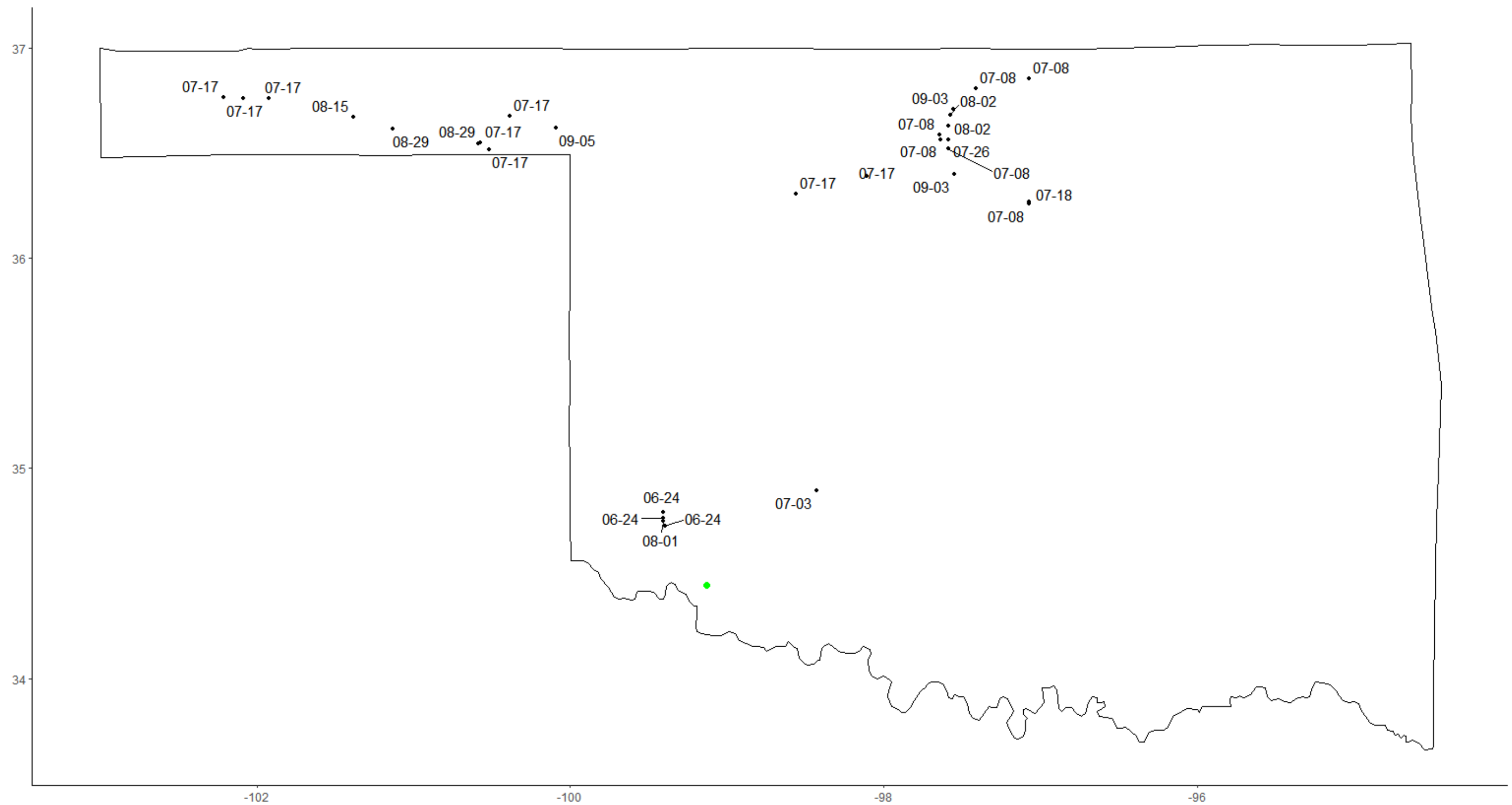


**Figure 1-24. Oklahoma's 2019 Infection Survey Field and First Survey Date**

**Note: The green dot is the initial location used in the moving probability model.**

R code

```
#Angle of centroid and fields in Kiowa County

x <- vector("double", length(field$Id ))

y <- vector("double", length(field$Id ))

B <- vector("double", length(field$Id ))

for (i in 1:length(field$Id )) {

  x[[i]] <-
as.numeric(((cos(deg2rad(field$Lati[length(field$Lati)])))*(sin(deg2rad(field$Lati[i]))))-
((sin(deg2rad(field$Lati[length(field$Lati)])))*(cos(deg2rad(field$Lati[i])))*(cos(deg2rad(field$
Long[i]-field$Long[length(field$Long)])))))

  y[[i]] <- as.numeric(((sin(deg2rad(field$Long[i]-
field$Long[length(field$Long)]))))*(cos(deg2rad(field$Lati[i]))))

  B[[i]]<- rad2deg(atan2(x[[i]],y[[i]]))}

field$angle<- B

field$angle <- ifelse(field$angle<0, field$angle+360,field$angle)


#Bearing of centroid and fields in Kiowa County (North:0)

Bear <- vector("double", length(field$Id ))

for (i in 1:length(field$Id )) {

  Bear[[i]] <- bearing(c(field$Long[length(field$Long)],field$Lati[length(field$Lati)]),
c(field$Long[i],field$Lati[i]), a=6378137, f=1/298.257223563)}

field$bearing<- Bear


# Kiowa County's daily weather information and PDIR converted to angles

A <- read_excel("G:/Group/Common/Seokil
Lee/Kiowa/Kiowatodesti/data_kiowa.xlsx",sheet='Sheet1')

A$angle <- ifelse(A$angle<0, A$angle+360,A$angle)

#Angle range for PDIR: PDIR angle +-11.25

A$angle_low <- A$angle-(22.5/2)

A$angle_up <- A$angle+(22.5/2)



#Identify the field in the PDIR angular range for each day

Z <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
```

```
for (i in 1:length(A$pdir)){
  for(j in 1:length(field$Id)){
    Z[j,i] <- ifelse((A$angle_low[i] <= field$angle[j] & field$angle[j]
<=A$angle_up[i]),"true","false")}}
head(Z)
colnames(Z) <-c(paste0("indicator",1:length(A$pdir)))
field <- data.frame(cbind(field,Z))


#calculate probability for each day using triangle distribution;1-1*(abs(angle-pdirangle)/11.25);
parameter:1
C <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
for (i in 1:length(A$pdir)){  C[,i] <- 1-1*(abs(field$angle-A$angle[i])/11.25)}
C[C<0] <- 0
colnames(C) <-c(paste0("P_PDIR",1:length(A$pdir)))
data.frame(C)
summary(C)
field <- data.frame(cbind(field,C))


# calculate distance between centroid and fields
a <- data.frame(matrix(nrow=(length(field$Id)),ncol=5))
colnames(a) <-c("Long_cen","Lati_cen","Long","Lati","dis")
a$Long_cen <- field$Long[length(field$Long)]
a$Lati_cen <- field$Lati[length(field$Lati)]
a$Long <-field$Long
a$Lati <-field$Lati
for (i in 1:length(a$Long_cen)){  a$dis[i]<- distm(c(a$Long_cen[i], a$Lati_cen[i]), c(a$Long[i],
a$Lati[i]))}


#change the unit from meter to miles:
a$dis <- a$dis*0.000621371
field$dis <- a$dis
```

```
#calculate probability for each day using triangle distribution
T <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
for (i in 1:length(A$pdir)){
  T[,i] <- 1-(1*((abs(field$dis-5*A$wspd[i]))*(1/(5*A$wspd[i]))))}
head(T)
T[T<0] <- 0
colnames(T) <-c(paste0("P_WSPD",1:length(A$pdir)))
field <- data.frame(cbind(field,T))

#combine P of PDIR, P of WSPD
P_PW <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
for (i in 1:length(A$pdir)){  P_PW[,i] <- C[,i]*T[,i]}
colnames(P_PW) <-c(paste0("P_PW",1:length(A$pdir)))
field <- data.frame(cbind(field,P_PW))

#identify field location
ok_field <- read_excel("G:/Group/Common/Seokil Lee/Kiowa/State/OK.xlsx", sheet = "OK")
ok_field <- data.frame(ok_field)
ok_field<-ok_field[!(ok_field$CLASS_NAME=="Station"),]
ok <- st_as_sf(maps::map("county","oklahoma", plot = FALSE, fill = TRUE))
testPoints <- data.frame(x=c(field$Long), y=c(field$Lati))
testPoints <- st_as_sf(testPoints, coords = c("x", "y"), crs = st_crs(ok))
ok<-st_join(testPoints, ok)
field[,c("state","county")] <- tstrsplit(ok$ID,",")
field$county <- str_to_title(field$county)

#get TAVG, RAIN for each county, day
#for TAVG
weather <- read_excel("G:/Group/Common/Seokil Lee/weather after all county___final.xlsx",
sheet = "Interpolation")
weather <-weather[!with(weather,is.na(TAVG)),]
weather1$TAVG_C <- (weather1$TAVG-32)*5/9
```

```
weather1 <- data.frame(weather1)

weather1_t <- weather1 %>% group_by(DAY,county) %>%
summarize_at(vars(TAVG_C),list(TAVG_C=mean))


#divide tavg,rain for each day

tv <- data.frame(split(weather1_t, weather1_t$DAY))

colnames(tv)[seq(1,ncol(tv),3)] <- c("DAY")

colnames(tv)[seq(2,ncol(tv),3)] <- c("county")

colnames(tv)[seq(3,ncol(tv),3)] <- c("TAVG_C")

tv <-split.default(tv, rep(1:nrow(A), each =3))


#combine data

for (i in 1:length(tv)) {

  field=inner_join(field, pivot_wider(tv[[i]], names_from=DAY, names_prefix="TAVG_C",
values_from=TAVG_C), by="county")}


##P of TAVG

P_T <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))

for (i in 1:length(A$pdir)){

  P_T[,i] <- (0.1*(field[ncol(field)-nrow(A)+i])) - (0.0026*((field[ncol(field)-nrow(A)+i])^2))}

colnames(P_T) <-c(paste0("P_TAVG",1:length(A$pdir)))

P_T[P_T<0] <- 0

field <- data.frame(cbind(field,P_T))


#for Rain

weather <- read_excel("G:/Group/Common/Seokil Lee/weather after all county___final.xlsx",
sheet = "Interpolation")

weather$RAIN[is.na(weather$RAIN)]=0

weather1 <- data.frame(weather)

weather1_r <- weather1 %>% group_by(DAY,county) %>%
summarize_at(vars(RAIN),list(RAIN=mean))


#divide tavg,rain for each day
```

```
m <- data.frame(split(weather1_r, weather1_r$DAY))
colnames(m)[seq(1,ncol(m),3)] <- c("DAY")
colnames(m)[seq(2,ncol(m),3)] <- c("county")
colnames(m)[seq(3,ncol(m),3)] <- c("RAIN")
m <-split.default(m, rep(1:nrow(A), each =3))

#combine data
field=data.frame(field)
for (i in 1:length(m)) {
  field=inner_join(field, pivot_wider(m[[i]], names_from=DAY, names_prefix="RAIN",
values_from=RAIN), by="county")}

##P of RAIN:
P_R <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
for (i in 1:length(A$pdir)){  P_R[,i] <- 1/exp(3*field[ncol(field)-nrow(A)+i])}
colnames(P_R) <-c(paste0("P_RAIN",1:length(A$pdir)))
P_R[P_R<0] <- 0
field <- data.frame(cbind(field,P_R))

#P of PDIR,WSPD,TAVG,RAIN
P_total <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
for (i in 1:length(A$pdir)){  P_total[,i] <- (field %>%
dplyr::select(starts_with('P_PW')))[i]*P_T[,i]*P_R[,i]}
colnames(P_total) <-c(paste0("p_to",1:length(A$pdir)))
field <- data.frame(cbind(field,P_total))
field<-field[!(field$CLASS_NAME=="centroid"),]
col<-(colnames((field %>%  dplyr::select(starts_with('p_to')))))

#cumulative probability by date
P_total<-head(P_total,-1)
P_cul <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
for (i in 1:length(A$pdir)){  P_cul[,i] <- rowSums(P_total[1:i])/i}
```

```
colnames(P_cul) <-c(paste0("P_cul",1:length(A$pdir)))
field <- data.frame(cbind(field,P_cul))

#for 25th
field$cilow<- as.numeric(cilow)
field$cihigh<- as.numeric(cihigh)

#95 percentile
b <- subset(field,field$P_cul1>0)
quantile(b$P_cul1,probs = 0.95)
per <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
quantile((subset(field,field$P_cul1>0))$P_cul1,probs = 0.95)
per[,1] <- ifelse((field$P_cul1>=(quantile((subset(field,field$P_cul1>0))$P_cul1,probs =
0.95))),"T","F")

for (i in 1:nrow(A)) {
  q<-data.frame(((subset(field,(field %>%  dplyr::select(starts_with('P_cul')))[i]>0))[ncol(field)-
13+i]))
  quantile(q[,1], probs = 0.95)
  per[,i] <- ifelse((field$P_cul1>=(quantile(q[,1], probs = 0.95))),"T","F")}
colnames(per) <-c(paste0("percen_indi",1:length(A$pdir)))
percen <- data.frame(ifelse(per=="T",P_cul,0))
#_____________________________________________________________
#only p greater than 95% percentile for each day
percen <- data.frame(matrix(nrow=(length(field$Id)),ncol=nrow(A)))
for (i in 1:length(A$pdir)){
  for(j in 1:length(field$Id)){
    percen[j,i] <- ifelse((per[j,i]=="T"),P_cul[j,i],"0")    }}
```

**Table 2-5. Field Statistics**

| Id | Region | Infestation | Last Report (m/d) | Period (day) | Wing | Distance to Field in Region (mile) | Distance to Last Field Infested in Texas (mile) | Distance to First South Field Infested in Oklahoma (mile) | Distance between Temporally Consecutive Infested fields (mile) | Population per Plant | TAVG (°C) | RAIN (mm) | Wind Direction Angle (degree) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ph | 0 | 7/12 | 42 | 0 | 106 | 105 | 261 | 0 | 0 | 24.1 | 0.026 | 109.5 |
| 2 | ph | 1 | 8/16 | 77 | 1 | 106 | 106 | 260 | 11 | 0.25 | 23.5 | 0.028 | 122.2 |
| 3 | ph | 1 | 8/16 | 77 | 0 | 95 | 108 | 251 | 5 | 0.02 | 23.6 | 0.034 | 122.9 |
| 4 | ph | 1 | 8/16 | 77 | 1 | 91 | 108 | 246 | 27 | 0.03 | 23.7 | 0.035 | 122.9 |
| 5 | ph | 0 | 7/12 | 42 | 0 | 91 | 108 | 246 | 27 | 0 | 24.2 | 0.030 | 113.7 |
| 6 | ph | 1 | 8/16 | 77 | 1 | 65 | 113 | 222 | 21 | 0.25 | 24.2 | 0.034 | 122.4 |
| 7 | ph | 1 | 8/16 | 77 | 1 | 44 | 116 | 201 | 28 | 3.28 | 24.3 | 0.031 | 124.4 |
| 8 | ph | 1 | 8/29 | 90 | 1 | 21 | 125 | 178 | 12 | 0.84 | 24.2 | 0.035 | 126.8 |
| 9 | ph | 1 | 8/16 | 77 | 0 | 16 | 128 | 174 | 16 | 0.06 | 25.2 | 0.037 | 127.9 |
| 10 | ph | 1 | 8/29 | 90 | 1 | 10 | 133 | 168 | 176 | 1.61 | 24.5 | 0.034 | 126.1 |
| 11 | ph | 0 | 7/12 | 42 | 0 | 0 | 134 | 158 | 0 | 0 | 25.4 | 0.030 | 123.5 |
| 12 | ph | 1 | 8/16 | 77 | 0 | 0 | 135 | 158 | 18 | 0.02 | 25.5 | 0.034 | 126.9 |
| 13 | ph | 0 | 7/7/ | 37 | 0 | 3 | 136 | 155 | 0 | 0 | 24.1 | 0.028 | 123.6 |
| 14 | ph | 1 | 8/16 | 77 | 1 | 18 | 151 | 147 | 16 | 0.19 | 25.8 | 0.026 | 127.8 |
| 15 | ph | 0 | 7/7/ | 37 | 0 | 31 | 162 | 135 | 0 | 0 | 24.3 | 0.028 | 121.6 |
| 16 | ph | 1 | 8/16 | 77 | 1 | 33 | 164 | 133 | 18 | 0.29 | 25.9 | 0.031 | 128.4 |
| 17 | ph | 1 | 8/16 | 77 | 0 | 51 | 178 | 118 | 70 | 0.05 | 25.9 | 0.031 | 123.5 |
| 18 | ph | 0 | 7/12 | 42 | 0 | 55 | 181 | 114 | 0 | 0 | 25.5 | 0.031 | 118.8 |
| 19 | sw | 1 | 7/13 | 43 | 1 | 32 | 178 | 92 | 5 | 2.08 | 27.0 | 0.014 | 123.5 |
| 20 | sw | 1 | 6/27 | 27 | 0 | 33 | 179 | 92 | 135 | 0.01 | 26.0 | 0.006 | 131.4 |
| 21 | sw | 1 | 6/14 | 14 | 1 | 35 | 181 | 94 | 167 | 0.03 | 24.1 | 0.009 | 125.1 |
| 22 | sw | 1 | 7/13 | 43 | 1 | 28 | 179 | 87 | 203 | 6.99 | 26.8 | 0.018 | 124.5 |
| 23 | sw | 1 | 8/30 | 91 | 1 | 38 | 188 | 96 | 96 | 0.03 | 26.7 | 0.028 | 128.7 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | sw | 1 | 5/31 | 0 | 1 | 0 | 197 | 59 | 16 | 0.08 | 23.3 | 0.000 | 109.5 |
| 25 | sw | 1 | 6/19 | 19 | 0 | 15 | 200 | 70 | 21 | 0.04 | 24.5 | 0.006 | 121.9 |
| 26 | sw | 1 | 6/19 | 19 | 1 | 13 | 210 | 50 | 43 | 0.03 | 24.1 | 0.008 | 118.0 |
| 27 | sw | 1 | 5/31 | 0 | 1 | 16 | 211 | 45 | 50 | 0.06 | 22.9 | 0.001 | 109.5 |
| 28 | sw | 1 | 8/30 | 91 | 1 | 59 | 241 | 0 | 11 | 9.1 | 25.9 | 0.029 | 127.0 |
| 29 | nc | 1 | 7/10 | 40 | 0 | 2 | 251 | 53 | 52 | 0.06 | 25.5 | 0.033 | 120.6 |
| 30 | nc | 0 | 6/21 | 21 | 0 | 0 | 252 | 52 | 0 | 0 | 24.6 | 0.032 | 112.5 |
| 31 | nc | 1 | 8/31 | 92 | 0 | 0 | 252 | 52 | 0 | 0.59 | 26.0 | 0.030 | 121.6 |
| 32 | nc | 0 | 6/21 | 21 | 0 | 1 | 253 | 53 | 0 | 0 | 24.6 | 0.032 | 112.5 |
| 33 | nc | 0 | 6/21 | 21 | 0 | 2 | 254 | 52 | 0 | 0 | 24.6 | 0.032 | 112.5 |
| 34 | sw | 1 | 8/30 | 91 | 1 | 68 | 252 | 11 | 53 | 518.78 | 25.8 | 0.031 | 129.6 |
| 35 | nc | 1 | 6/15 | 15 | 0 | 31 | 278 | 78 | 24 | 0.08 | 24.2 | 0.013 | 111.2 |
| 36 | nc | 1 | 8/3/ | 64 | 1 | 32 | 281 | 79 | 12 | 0.02 | 26.9 | 0.014 | 118.4 |
| 37 | nc | 1 | 8/3/ | 64 | 1 | 42 | 287 | 90 | 30 | 0.18 | 26.8 | 0.016 | 115.6 |
| 38 | nc | 0 | 7/10 | 40 | 0 | 45 | 288 | 93 | 0 | 0 | 25.0 | 0.018 | 116.7 |
| 39 | nc | 1 | 7/10 | 40 | 0 | 52 | 295 | 100 | 186 | 0.06 | 24.9 | 0.017 | 116.4 |
| 40 | nc | 1 | 6/16 | 16 | 0 | 36 | 288 | 68 | 17 | 0.02 | 24.1 | 0.008 | 109.5 |
| 41 | nc | 1 | 8/3/ | 64 | 0 | 73 | 312 | 120 | 290 | 2.69 | 26.1 | 0.023 | 122.3 |
| 42 | nc | 1 | 7/25 | 55 | 0 | 74 | 315 | 121 | 11 | 0.01 | 26.1 | 0.019 | 120.2 |
| 43 | nc | 0 | 6/23 | 23 | 0 | 52 | 303 | 77 | 0 | 0 | 24.6 | 0.006 | 121.3 |
| 44 | nc | 1 | 6/16 | 16 | 1 | 53 | 304 | 77 | 2 | 0.01 | 24.0 | 0.005 | 114.6 |
| 45 | nc | 1 | 6/16 | 16 | 0 | 53 | 304 | 78 | 149 | 0.02 | 24.0 | 0.005 | 113.9 |
| 46 | nc | 0 | 6/28 | 28 | 0 | 69 | 316 | 112 | 0 | 0 | 24.3 | 0.013 | 119.5 |
| 47 | nc | 0 | 7/25 | 55 | 1 | 73 | 318 | 116 | 40 | 0.05 | 26.0 | 0.016 | 121.1 |

Note: region means the northwest (ph), southwest (sw), and north-central (nc) regions, respectively. Infestation means that the sugarcane aphid population per plant is positive and the field is infested during the infestation investigation period. The last report means the date of the first infestation if there is an infestation, and the date of the last investigation if there is no infestation. Period means the difference between the first infestation (May 31) and the date of infestation or the last investigation in each field. wing means that the field is infested with alatae with wings. Distance to field in region is the distance between the first infested field for each local region and other fields in the region. Distance to last field infested in Texas is the distance between the last infested sorghum field in Texas and the sorghum field in

Oklahoma. Distance to first south field infested in Oklahoma is distance between the sorghum field where the first infestation occurred in Oklahoma and other fields in Oklahoma. The population per plant is the number of wing/wingless sugarcane aphid per sorghum plant. TAVG, RAIN, Angle are the median of temperature, rainfall, and angle of wind direction variables from the first infestation date in Oklahoma to the first infestation date.

**Table 2-6. Field Ties by Date of Infestation Report**

| Date (M/D) | Number of Tie Fields | ID | Total Population per Plant | Region | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 05/31 | 2 | 24 | 0.08 | swok | 34.900 | -98.864 |
| | | 27 | 0.06 | swok | 34.985 | -98.603 |
| 06/16 | 3 | 40 | 0.02 | ncok | 36.262 | -97.369 |
| | | 44 | 0.01 | ncok | 36.232 | -97.073 |
| | | 45 | 0.02 | ncok | 36.257 | -97.073 |
| 06/19 | 2 | 25 | 0.04 | swok | 34.677 | -98.862 |
| | | 26 | 0.03 | swok | 34.912 | -98.639 |
| 07/10 | 2 | 29 | 0.06 | ncok | 36.233 | -98.036 |
| | | 39 | 0.06 | ncok | 36.810 | -97.433 |
| 07/13 | 2 | 19 | 2.08 | swok | 34.585 | -99.283 |
| | | 22 | 6.99 | swok | 34.644 | -99.246 |
| 07/25 | 2 | 42 | 0.01 | ncok | 37.059 | -97.181 |
| | | 47 | 0.05 | ncok | 36.934 | -97.058 |
| 08/03 | 3 | 36 | 0.02 | ncok | 36.532 | -97.586 |
| | | 37 | 0.18 | ncok | 36.702 | -97.554 |
| | | 41 | 2.69 | ncok | 37.056 | -97.231 |
| 08/16 | 10 | 2 | 0.25 | phok | 36.733 | -102.466 |
| | | 3 | 0.02 | phok | 36.769 | -102.264 |
| | | 4 | 0.03 | phok | 36.758 | -102.182 |
| | | 6 | 0.25 | phok | 36.765 | -101.703 |
| | | 7 | 3.28 | phok | 36.679 | -101.339 |
| | | 9 | 0.06 | phok | 36.615 | -100.841 |
| | | 12 | 0.02 | phok | 36.529 | -100.568 |
| | | 14 | 0.19 | phok | 36.630 | -100.277 |
| | | 16 | 0.29 | phok | 36.621 | -99.982 |
| | | 17 | 0.05 | phok | 36.601 | -99.660 |
| 08/29 | 2 | 8 | 0.84 | phok | 36.617 | -100.925 |
| | | 10 | 1.61 | phok | 36.620 | -100.714 |
| 08/30 | 3 | 23 | 0.03 | swok | 34.407 | -99.168 |
| | | 28 | 9.1 | swok | 35.467 | -98.076 |
| | | 34 | 518.78 | swok | 35.465 | -97.874 |

Note: region means the northwest (ph), southwest (sw), and north-central (nc) regions, respectively.

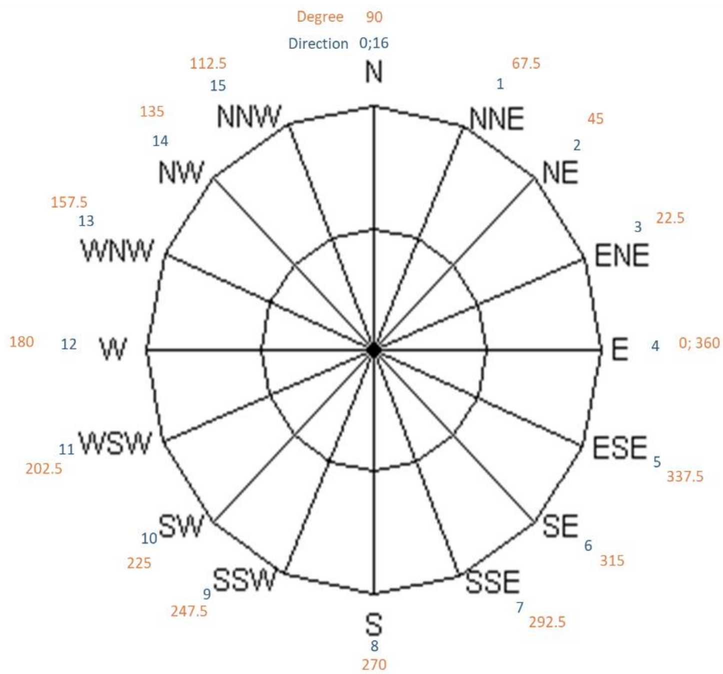


**Figure 2-4. Direction and Angle of the 16-wind Compass**

VITA

Seokil Lee

Candidate for the Degree of

Doctor of Philosophy

Dissertation: SUGARCANE APHID MOVEMENT MODEL FOR OKLAHOMA, WEATHER AND GEOGRAPHIC EFFECTS ON SUGARCANE APHID MIGRATION IN OKLAHOMA

Major Field: Agricultural Economics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Agricultural Economics at Oklahoma State University, Stillwater, Oklahoma in July, 2022.

Completed the requirements for the Master of Science in Agricultural Economics at Chungbuk National University, CheongJu, South Korea in 2016.

Completed the requirements for the Bachelor of Science in Agricultural Economics at Chungbuk National University, CheongJu, South Korea in 2013.

Experience:

Employed as a Lecturer at Chungbuk National University, Department of Agricultural Economics, March 2022 – Present

Employed as a Graduate Research Assistant at Oklahoma State University, Department of Agricultural Economics, August 2018 – September 2020.

Employed as a Researcher at Chungbuk National University, Institute of Agricultural Science and Technology, May 2016 – May 2018.

Professional Memberships:
Agricultural and Applied Economics Association