

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2004 - 16th Annual Conference Proceedings

PREDICTION OF YELLOW STARHISTLE SURVIVAL AND MOVEMENT OVER TIME AND SPACE

Fei Tian

Bahman Shafii


Christopher J. Williams

Timothy S. Prather

William J. Price

See next page for additional authors

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

 Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Tian, Fei; Shafii, Bahman; Williams, Christopher J.; Prather, Timothy S.; Price, William J.; and Lass, Lawrence W. (2004). "PREDICTION OF YELLOW STARHISTLE SURVIVAL AND MOVEMENT OVER TIME AND SPACE," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1152>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

Author Information

Fei Tian, Bahman Shafii, Christopher J. Williams, Timothy S. Prather, William J. Price, and Lawrence W. Lass

PREDICTION OF YELLOW STARHISTLE SURVIVAL AND MOVEMENT OVER TIME AND SPACE

Fei Tian¹, Bahman Shafii¹, Christopher J. Williams²
Timothy S. Prather³, William J. Price¹, and Lawrence W. Lass³

¹Statistical Programs, ²Division of Statistics, ³Division of Crop and Weed Science
University of Idaho
Moscow, Idaho

Abstract

Yellow starthistle is a noxious weed that has become a serious plant pest with devastating impact on ranching operation and natural resources in western states. Early detection of yellow starthistle and predicting its spread has important managerial implications and greatly reduce the economic losses due to this weed. The dispersal of yellow starthistle consists of two main components, plant survival and seed movement. Resources and direct factors relating to these components are not typically available or are difficult to obtain. Alternatively, topographic factors, such as slope, aspect and elevation, are readily available and can be related to plant survival and seed movement. In this study, several GIS network models incorporating these topographic factors are considered for the prediction of yellow starthistle spread. The models differed in their assessment of the costs of movement derived from these factors. Models were evaluated based on their predictive ability and residual analysis. The optimal model gave an accurate estimate of the dispersal boundary for the study area. Further validation of the estimated model using an independent data set from a larger area also verified its predictive capability.

I. INTRODUCTION

It is believed that yellow starthistle (*Centaurea solstitialis* L.) originally came from Eurasia, Middle East and south central Europe (Maddox 1981). Its introduction to North America probably occurred sometime after 1849 as a seed contaminant. Today in the U.S., yellow starthistle (YST) can be found in 23 of the 48 contiguous states, extending as far east as New York (Maddox et al. 1985). Although no economic assessments have been conducted for YST, millions of dollars in losses may occur from interference with livestock grazing and forage harvesting procedures as well as lower yield and forage quality of rangelands (Callihan et al. 1982, Roché and Roché 1988). Numerous reports also have characterized the toxic effect of yellow starthistle on horses (Cordy, D.R. 1954, Cheeke and Shull 1985, Lass et al. 1999). The best management of yellow starthistle is obtained by identifying the potential weed problems early, controlling them before they reproduce and spread, and monitoring the sites regularly to maintain adequate follow-up control. One successful method for the initial step of locating YST infestations is to inventory areas using on site

surveys or remote sensing techniques. The pattern and scope of their future occurrences can then be predicted.

The prediction of plant distributions in space based on sample data has been a popular research topic. Generalized regression models (Lehmann et al,2002), classifications, regression trees (Miller and Franklin, 2001), Bayesian approaches (Fisher and Langley,1990) and neural networks (Fitzgerald and Lees 1992), are some of the techniques investigated. Some researchers have considered the prediction of plant migration over time based on changes in resources and predictive variables such as climate (Schwartz 1993). However, for YST dispersal prediction, both the space and the time, or more specifically, plant survival (time) and seed movement (space), should be considered. The process of YST plant dispersal may be described as follows: in the first year, a plant grows to maturity and produces seeds. These seeds may be carried away from the parent plants, and potentially grow the following year to produce a new generation of seeds. This cycle is repeated each year. Thus, the dispersal of YST is dependent on plant survival to seed production as well as how far the seed is carried away from the parent plant. YST survival to seed production is influenced by three major factors: available resources (water, nutrients, air and light), direct factors (temperature, precipitation, soil and radiation) and indirect factors (slope, aspect and elevation) (Guisan and Zimmermann 2000). Resource information for individual plants is difficult to obtain. In addition, the small areas used for modeling purpose typically show little variation in the direct factors. Therefore, topographic variables are selected for modeling plant survival. These are related to resource factors and can usually be easily obtained. As with survival, different factors may affect YST seed movement. For example, wind, animal/bird movement or human activities can move seeds. Wind typically leads only to short distance movement for YST and is inconsequential here (Roché 1992). Tracking of animals and/or birds may be important, but is expensive and time consuming. Tracking all human activities is also impractical. Hence, as with survival, topographic characteristics such as elevation, slope, and aspect are attractive for modeling seed movement.

The purpose of this article is to explore potential models for predicting yellow starthistle dispersal over time and space in a localized area using topographic information with geography information system (GIS) techniques. Four prediction models are introduced and compared. While these models are based on GIS network analysis methods, they differ in their survival and movement assumptions. The optimal prediction for an area in central Idaho is shown for the dispersal of YST in 1987 given a documented 1981 infestation. Further validation of the model with a larger independent data set is also demonstrated.

II. METHODS

1. Networks and Resource Allocation

A GIS network consists of nodes and links. Links are usually straight line segments. Each node has at least one link and each link has two nodes. Figure 1 shows a small and simple network with 9 nodes and 12 links in an arrangement referred to as a square regular network (SRN) where links have equal lengths. A SRN node is an intersection of links. Nodes in the network are characterized by their turn impedance.

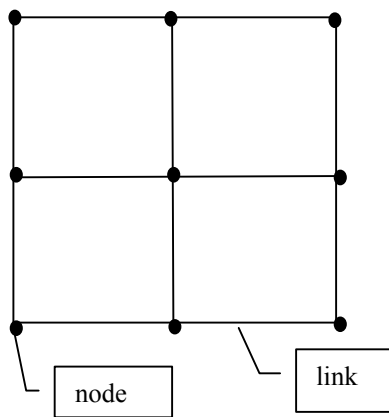


Figure 1. The structure of SRN

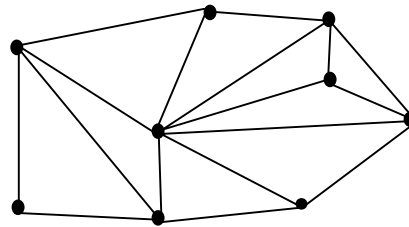


Figure 2. The structure of TIN

A turn is a directional choice i.e. left, right or straight and turn impedance may be defined as the time it takes to complete the choice. The attributes of links are their costs. Cost, in this case, can be thought of as the expense of moving an object through the link. Each link may have two costs, corresponding to the two possible directions along the link. While the length of the link can be directly considered as a cost, the time and the resistance of an object passing through the link may also be reasonable values for cost depending on the application.

Another common network type is the triangulated irregular network (TIN). A TIN is a 3-dimensional network consisting of triangles that represent planes or surfaces. The TIN link is the intersection of two planes except for the network boundary. TIN nodes are join points of more than three planes (Figure 2). TIN arrangements can be used for slope and aspect computations as well as prediction networks.

GIS applications based on network analyses commonly involve allocation (Chang 2002). Allocation considers the spatial distribution of resources through a network. Here, “resources” usually refer to public facilities, such as fire stations or schools, whereas the distribution of resources defines their service zones. The main objective of spatial allocation analysis in the service example is to measure the efficiency of public facilities such as the response time of an emergency service. Allocation analysis methods may also be useful for YST dispersal prediction. If each YST plant can be considered as a resource, spatial allocation analysis would then measure the efficiency of YST dispersal through time. Numerically, the spreading time of YST is the total cost in the network and is defined by the sum of all costs, across all links that YST seed must pass through. In allocation analysis one or more YST moves along the most efficient paths to all potential destinations and stop at all possible positions on links or nodes at when a predefined threshold total cost (TTC) is reached. Connecting all stop points with straight line segments forms the predicted boundary for the YST infestation. For example, in figure 3, if TTC is 5, then the location of the boundary is determined by connecting all the end points.

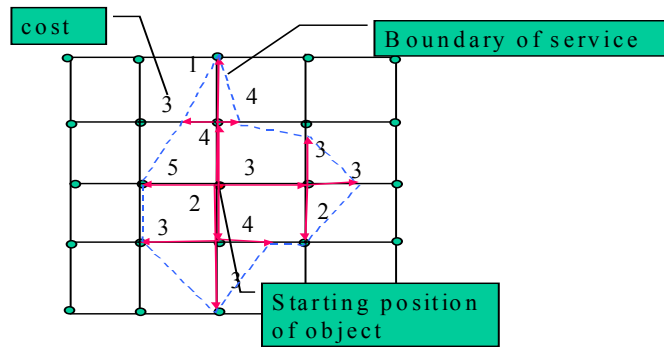


Figure 3. Allocation analysis

In order to use network allocation analysis methods for YST dispersal, however, the costs or cost function relating to movement through each link must be determined.

2. Cost Functions

YST dispersal can be viewed as a process based on two factors: plant survival and seed movement. Therefore, the cost of a link in a YST dispersal network should be related to these two factors. Cost functions may be given in general by:

$$Cost = cs \times cm \tag{1}$$

where *cs* is a component for YST plant survival and *cm* is a component for YST seed movement. Here survival and movement are assumed to be independent. The values of these variables will be determined through three regression models: 1) an occurrence model related to the plant survival component; 2) a transition model and 3) a velocity model, both related to the seed movement component. Once cost functions are determined, the prediction of YST dispersal can be carried out using network allocation analysis methods.

Four prediction models based on GIS network analysis are explained below. The models differ in the construction of their networks and the definition of the cost functions for a link. The differences and similarities of the four models are listed in Table 1.

Model	Survival component	Movement component	Cost function	Network structure
1	Yes	No	$(1-P_s) \cdot 1$	TIN
2	Yes	Yes	$1 - P_s \cdot P_t$	SRN
3	No	Yes	$1 \cdot (1/v)$	SRN
4	Yes	Yes	$(1 - P_s) \cdot (1/v)$	SRN

Table 1. The difference and similarities of the four GIS prediction models.

3. Prediction Models

- **Prediction model 1:**

The links and nodes of this network are directly created by TIN (Figure 2). In this model, only the survival component is considered and the movement component is assumed to be constant.

The following YST occurrence model is used for the survival component:

$$\ln \frac{Ps_{ij}}{1 - Ps_{ij}} = \beta_0 + \beta_1 u + \beta_2 u^2 + \beta_3 v + \beta_4 v^2 + \beta_5 \ln(u^2 + v^2) + \varepsilon_{ij} \quad (2)$$

This model (Shafii et. al 2003) is used to calculate the probability of yellow starthistle presence based on the landscape variables slope and aspect. The proportion of yellow starthistle present, Ps_{ij} , is computed at the i^{th} and j^{th} classification levels of aspect and slope, respectively, $i=1,2,\dots,16$, $j=1,2,\dots,16$, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are regression parameters, and u and v are polar transformations of slope and aspect given as $u=\text{slope}*\cos(\text{aspect})$ and $v=\text{slope}*\sin(\text{aspect})$. Slope is classified into 16 levels (1st-4th: 1-10%, every 2%; 5th-15th: every 5% from 10% to 60%; 16th: >60%). Aspect is also classified into 16 levels incremented every 22.5° , and ε_{ij} is an error term assumed to be $NID(0, \sigma^2)$.

Figure 4 illustrates the observed and predicted surfaces of this model for rangeland (Shafii et. al 2003) and demonstrates the changes evident in yellow starthistle occurrence or survival over various slope and aspect combinations.

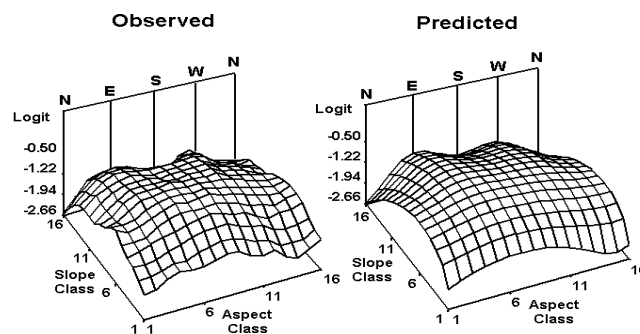


Figure 4. Comparison of observed and predicted yellow starthistle occurrence.

Using equation (1) and assuming a constant cost for seed movement, the cost function for this prediction model is defined as:

$$Cost_i = (1 - Ps_i) \times 1 \quad (3)$$

where $cost_i$ is the cost of the i^{th} link and Ps_i is the predicted probability of yellow starthistle occurrence at the i^{th} link. This is calculated as the average probability of YST occurrence from equation (2) computed at the two nodes of the link.

• **Prediction model 2:**

This model uses a SRN network and considers seed movement in addition to plant survival for computing the ability of YST seed to move through a fixed distance for a given combination of topographic factors.

The additional movement component may be obtained from the transition model and assumes that 1) only seeds from parent YST at the original observed infestation boundary develop into a new infestation boundary, and 2) YST seed moves along a straight line (radius) originating at the center of the existing infestation. Each radius is defined by its spatial position and length which the infestation center is given by:

$$(\bar{x}_{mc}, \bar{y}_{mc}) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right) \tag{4}$$

where $\bar{x}_{mc}, \bar{y}_{mc}$ are the coordinates of the mean center, x_i and y_i are the coordinates of the i^{th} point, and n is the number of observations in the polygon formed by the infestation boundary (Figure 5).

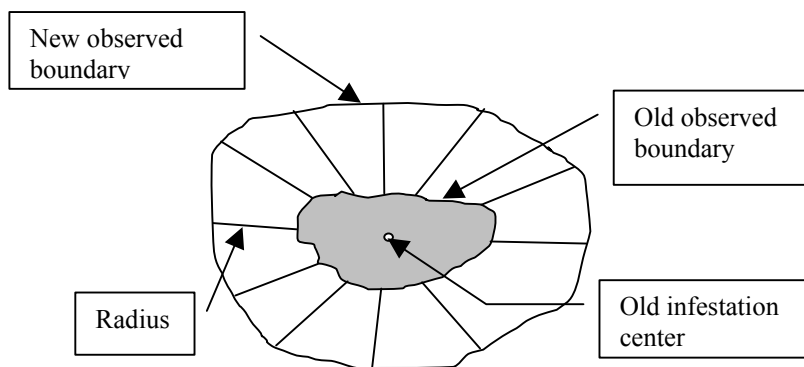


Figure 5. Transition model components.

The probability of YST seed movement (P_t) can be computed as proportional to the frequency of radii at a fixed length. Using this method, P_t is modeled as a function of topographic factors such as slope, elevation etc. In this case, slope and elevation are given by:

$$s = (\arctan(e_a - e_b) / L) \times 180 / \pi \tag{5}$$

$$e = (e_a + e_b) / 2 \tag{6}$$

where s is the slope of a radius, e_a and e_b are the elevations of the two radius ends (end A and end B), e is the average elevation of the radius and L is the projected length of the radius (Figure 6).

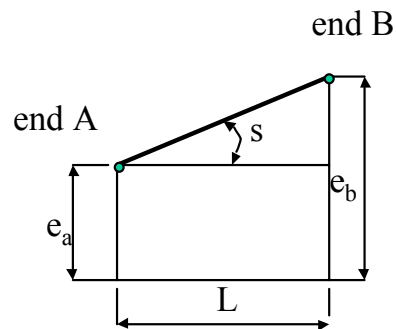


Figure 6. The relationship between the slope of a radius and its elevation.

As will be shown later, the transition probability can be given as:

$$\ln \frac{Pt_i}{1 - Pt_{ii}} = \beta_0 + \beta_1 s_i + \beta_2 s_i^2 + \beta_3 e_i + \beta_4 s_i * e_i + \varepsilon_i \quad (7)$$

where Pt_i is the probability of yellow starthistle seed movement for i^{th} radius, β_0 , β_1 , β_2 , β_3 and β_4 are regression parameters, and s_i and e_i are the slope and average elevation of i^{th} radius, respectively, as defined above, and ε_i is an error term assumed distributed as $NID(0, \sigma^2)$.

With the assumption of independence of P_s and P_t , the cost function in this case is defined as:

$$Cost_i = 1 - P_{s_i} \times P_{t_i} \quad (8)$$

where, $cost_i$ is the cost of the i^{th} link, P_{s_i} is the predicted probability of YST occurrence at the i^{th} link as in model 1, and P_{t_i} is the predicted probability of YST seed movement at i^{th} link .

- **Prediction model 3:**

For this prediction model, the network arrangement is the same as prediction model 2; however, only the seed movement component is considered.

A regression model, referred to as the velocity model, is used to compute the speed of YST movement along a straight-line distance. This model has the same assumptions as the transition model in equation (7).

Consider the following where distance is given as velocity multiplied by time:

$$d = v \times t \quad (9)$$

Here d is the distance an object moves, v is the velocity of the object, and t is the time the object takes to cover the distance d . For the YST movement problem, d would be the length of a straight line (radius), v the speed of YST movement through each radius, and t the time it takes YST to move.

As is shown later, velocity is estimated with the following polynomial model:

$$v_i = \beta_0 + \beta_1 \times e_i + \beta_2 \times e_i^2 + \varepsilon_i \quad (10)$$

where, v_i is the YST seed movement speed within a fixed period and corresponds to the length of i^{th} radius, e_i is the average elevation of the radius as defined in (6), β_0 , β_1 and β_2 are regression parameters, and ε_i is an i^{th} error term assumed NID $(0, \sigma^2)$.

Assuming a constant growth rate along all links, the cost function for this model is defined as:

$$Cost_i = 1 \times time_i = 1 \times (1/v_i) \tag{11}$$

where $cost_i$ is the cost of the i^{th} link, $time_i$ is the time required for a YST to pass through the link, and v_i is the velocity of YST seed movement at the i^{th} link of the network.

- **Prediction model 4:**

This model combines attributes of models 2 and 3. The cost function is defined as:

$$Cost_i = (1 - Ps_i) \times (1/v_i) \tag{12}$$

where, $cost_i$ is the cost of the i^{th} link, Ps_i is the predicted probability of yellow starthistle occurrence at the i^{th} link from equation (2) and v_i is the velocity of yellow starthistle seed movement at the i^{th} link of the network in equation (10).

4. Model Assessment Methods

Two basic methods are used for assessing the prediction results: 1) visual assessment and 2) residual assessment. Visual assessment is a qualitative and dynamic measure of how well a predicted boundary fits the observed boundary. While subjective, visual inspection allows for quick assessment of model predictions and is easily done by a potential user. For example, using four predicted boundaries based on differing TTC's, the predicted dynamic trend can be clearly shown. Larger or smaller TTC values will lead to larger or smaller predicted boundaries, respectively. A residual assessment method can be used to quantitatively assess how well the predicted boundary fits the observed data. Initially, rays are created from the mean polygon center within the new observed boundary with the mean center is given by:

$$(\bar{x}_{rc}, \bar{y}_{rc}) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right) \tag{13}$$

Here $\bar{x}_{rc}, \bar{y}_{rc}$ are the coordinates of the mean center of new infested area, x_i and y_i are the coordinates of the i^{th} point in the new observed area, and n is the number of points in the new observed area (Figure 7). The segments, defined as the distance along the radius between the observed and predicted boundaries, are considered as residuals, δ_i . These residuals may be positive or negative depending on the shape of the observed and predicted boundaries.

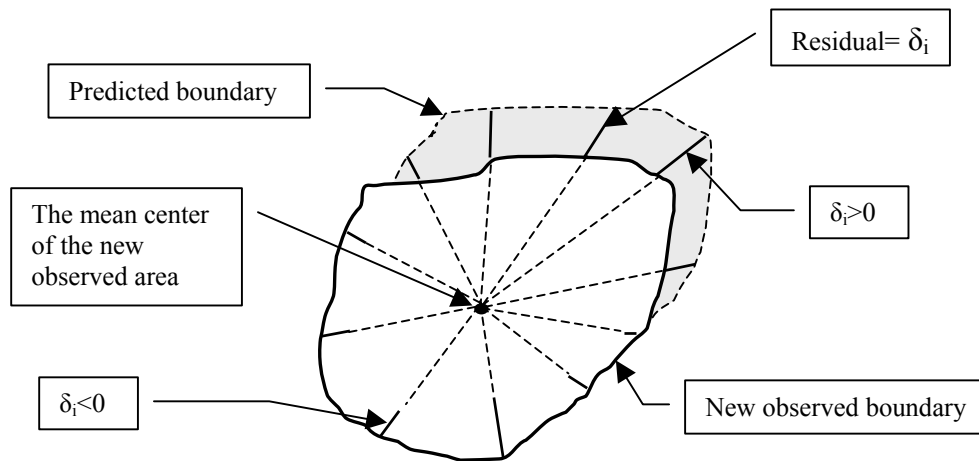


Figure 7. Definition of prediction residuals, δ_i .

A mean square error (MSE) can then be defined from the residuals as:

$$MSE = \frac{\sum \delta_i^2}{n - 1} \quad (14)$$

where MSE is the mean square error, δ_i is the length of the i^{th} residual segment and n is the number of segments. The number of rays defined depends on the size of new observed area and the available resolution of the coverage point file. In order to use all available information, each point at the outer boundary is used as an end of a residual segment.

The MSE, however, will change with various TTC values. In order to find an optimum prediction boundary which has the smallest MSE, one of two possible methods: the first, a numeric approach, would calculate the MSE for several different TTC levels until the MSE is less than a given minimum value. This method, however, may be too computationally intensive and time consuming if all operations are run in GIS software such as ArcView (ArcView 2002) and ArcGIS (ArcGIS 2002). A second method is more practical. Here, it is assumed that the MSE follows the deterministic quadratic form:

$$MSE_k = \beta_0 + \beta_1 \times TTC_k + \beta_2 \times TTC_k^2 \quad (15)$$

Where MSE_k is the mean square error when TTC is at the k^{th} level, and β_0 , β_1 and β_2 are regression parameters. By selecting a few TTC levels such that the minimum MSE can be defined, the optimum total cost (OTC) leading to the smallest MSE predicted boundary can be calculated by minimizing equation (15).

Once determined, the OTC can then be used to predict YST dispersal in a future time period for the same or similar area.

III. EMPIRICAL RESULTS

1. Description of data

Two types of data sets are used in this study. The first is YST data which records the presence or absence of YST. The second data set is topographic data which contains information on position and elevation.

Yellow starthistle data: YST occurrence has been previously surveyed for north central Idaho. A University of Idaho aerial survey was conducted to cover Idaho, Latah, Lewis and Nez Perce counties in the years 1981 and 1987. Each observation recorded the presence or absence of YST at a resolution of 30 by 30 meters. Figure 8 indicates the presence of YST for the two years in the region surveyed. The data were made available in the form of shape files which are the standard non-topological vector data form used in ArcView software.

Topographic data: The raw topographic data form was a USGS Digital Elevation Model (DEM) recorded at a resolution of 10*10 meters (see for example <http://data.geocomm.com/dem/demdownload.html>). DEM files record the spatial coordinates of a rectangular array of points in addition to their elevation. For use in modeling, the DEM files were converted to TIN format and point coverage files. The SRN nodes were acquired from the point coverage files. Slope and aspect information for nodes were derived from the TIN files.

2. Data Selection Criteria

The prediction area, for model development, was Moughmer Point in Idaho County, Idaho (Figure 8). Several areas on and around Moughmer Point were available for use as regression model training data. The different regression models, however, required different data selection criteria. For the occurrence model, training data with a large number of YST observations was preferred to ensure an adequate amount of data for all combinations of the topographic variables. The transition model training data also needed a large number YST observations for same reason, however, obstructing factors, such as forest, farmland, range and road, were avoided. Such areas are not preferable because they are affected by human activity. While the training data for the velocity model also avoided areas influenced by humans, it required more variation in topographic factors such that the model is applicable to a wider range of topographic conditions. Figure 8 also shows the training data selection results.

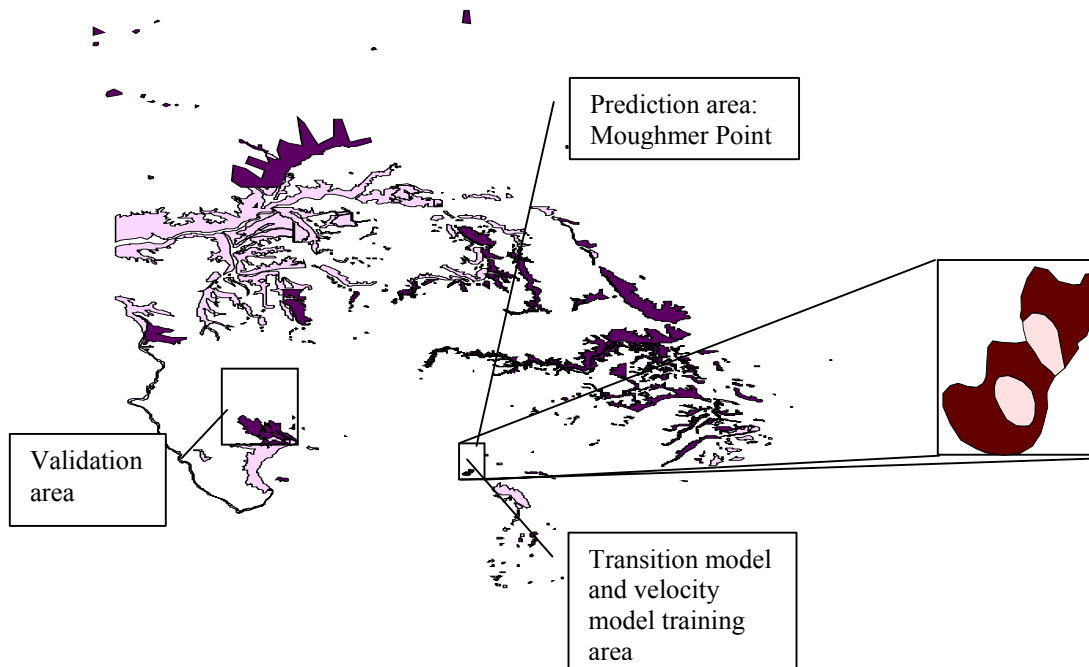


Figure 8. Surveyed YST data in the north central Idaho. Light gray refers to YST presence in 1981 and dark grey refers to new YST infestation between 1981 and 1987. Training data selection results are also shown.

3. Prediction Model Comparison

- **Model 1: TIN arrangement with the occurrence model**

P_s in cost function (3) was obtained from the predicted occurrence model in equation (2) (Shafii, et al. 2003). The estimated parameters reported from that model are given in Table 2.

Parameters	Estimate	Std Error
β_0	-0.19	0.027
β_1	-0.13	0.012
β_2	-0.64	0.023
β_3	-0.31	0.014
β_4	-0.88	0.022
β_5	0.34	0.011

Table 2. Estimated parameters and their associated standard errors for the yellow starthistle occurrence model.

The prediction model, based on the cost function in equation (3), was initially run with 4 values of TTC (Figure 9). Visual assessment of this model shows poor correspondence between predicted and observed boundaries. The rate of YST spread is approximately the same in all directions.

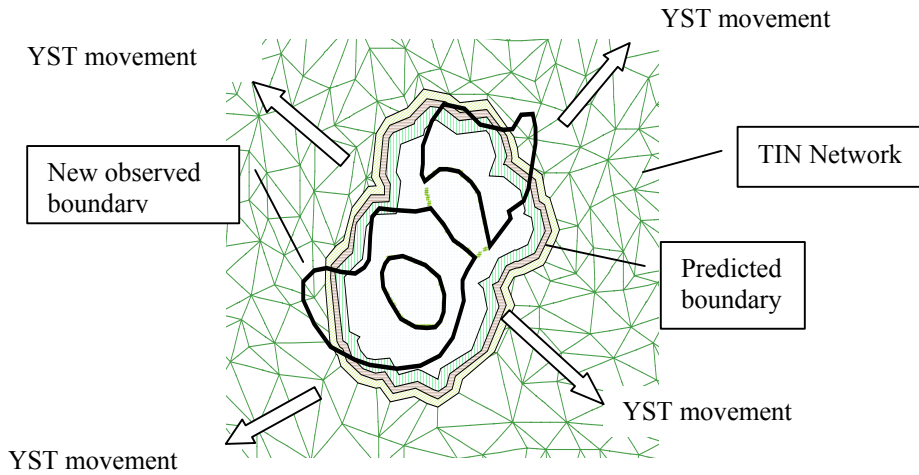


Figure 9. Model 1 Predicted boundaries shown for 4 increasing values of TTC (cross hatched areas) compared to the observed boundary (solid line).

Using 6 levels of TTC, six predicted boundaries were drawn and their associated MSEs computed from equation (14). Using the quadratic form in equation 15, calibration of the model gave an $OTC = 0.052$ with an associated MSE of 23158. The optimum predicted boundary and resulting prediction residuals assuming the OTC of 0.052 are shown in Figure 10A.

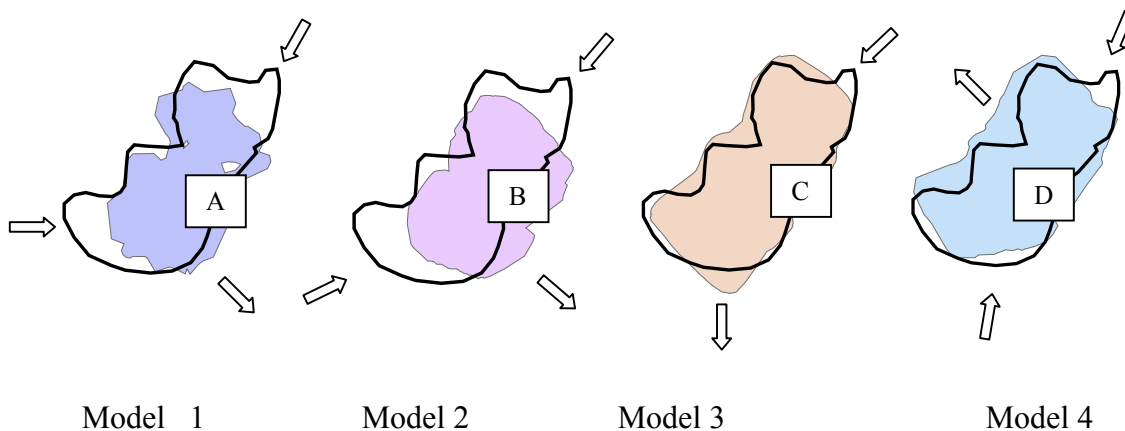


Figure 10. Visual comparison of the 4 yellow starthistle models. Optimal boundaries--grey, observed 1987 boundaries—solid lines. Arrows indicate under or over estimation.

In order to do a residual assessment, 720 residuals rays were created from the mean center of new (1987) observed area.

The mean and median residuals were -30 and -14, respectively, and all tests for $\mu_0=0$ were significant. Figure 11 indicates numerous negative residuals or model under estimation.

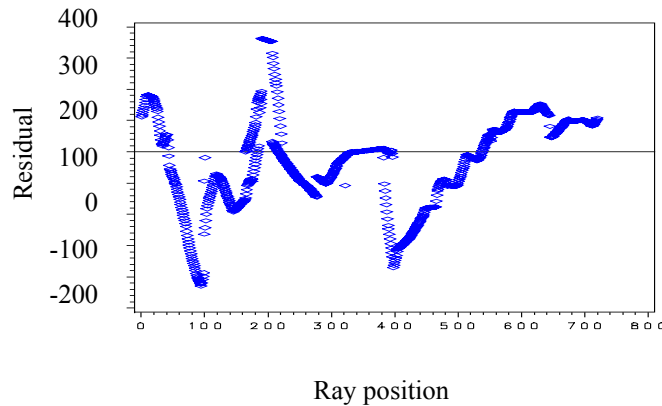


Figure 11. The residual plot for prediction Model 1.

The reasons for this prediction failure can be attributed to nodes and links derived directly from TIN. Because of this, it is hard to determine which plane the slope and aspect of each node came from. In addition, some areas in the network are sparse and result in too much information lost. Furthermore, the cost function (1-Ps) does not supply the necessary dynamic information about YST seed movement.

- **Model 2: SRN arrangement with the occurrence model and the transition model**

Pt in cost function (8) was obtained from the transition model developed using the Moughmer area as training data. This data consisted of two older infestation areas. Rays were created using the centers of the original infestations. Radii were defined to be the ray segment between the old and new infestation boundaries. Figure 12 shows the shape of the training area and its associated radii.

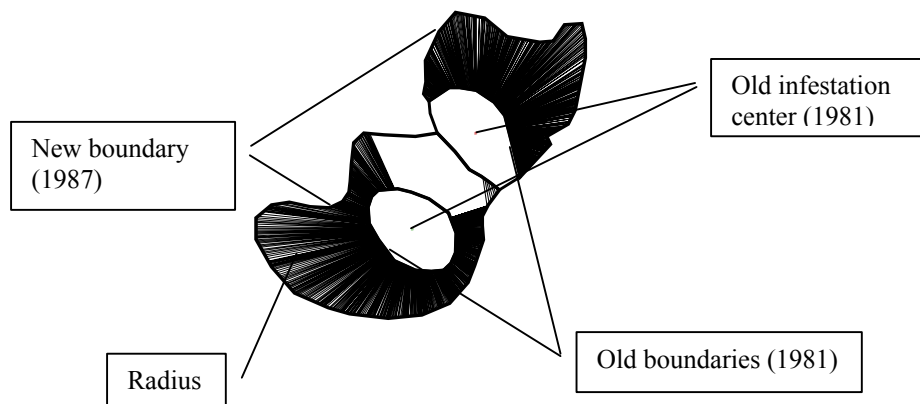


Figure 12. The transition model training data. Radii were formed by rays extending from old infestation centers to the new infestation boundary.

All available points in the new observed boundary were used as radii ends resulting in 917 observations. The length of each radius represents the distance YST moved over a 6 year period. The zone between the two old infestations areas was not used because it was not possible to determine which old infestation created the new boundary.

The frequency distribution of radii lengths was then examined. Thirty percent of the observations (270) had approximate lengths of 20 meters (ranging from 15 to 25 meters). Therefore, 20 meters was selected as a representative fixed length for estimating the probability of YST seed movement. This is also consistent with the length of links in the network. Estimation of Pt considered the topographic factors of slope, aspect and elevation within the 20 meter data. Initial investigation found limited effects of aspect; however, slope and elevation were indicated as prominent factors.

In order to model the probability of YST seed movement, it was necessary to summarize this landscape information. Therefore, the slope was classified into 7 levels (≤ -30 , $-30\sim-20$, $-20\sim-10$, $-10\sim 10$, $10\sim 20$, $20\sim 30$, and ≥ 30) and the elevation was classified into 5 levels (≤ 750 , $750\sim 800$, $800\sim 950$, $950\sim 1050$, and ≥ 1050). These class intervals were chosen to provide adequate resolution for slope and elevation while maintaining a sufficient number of observations within each class. The cross tabulation of these classes resulted in a 7 by 5 table (35 cells) of slope and elevation. Equation (7) was found to give the best fit for the proportion of radii length. Estimated parameters and standard errors for this model are given in Table 3. All parameters were significant and the predicted model fit the data well (Figure 13).

Parameters	Estimate	Std error
β_0	-2.02	0.2795
β_1	1.55	0.2852
β_2	-0.27	0.0490
β_3	0.04	0.0150
β_4	-0.42	0.1083
β_5	-0.74	0.0943

Table 3. Estimated parameters and their associated standard errors for the transition model. Residuals were minimal in magnitude and showed no trend or pattern.

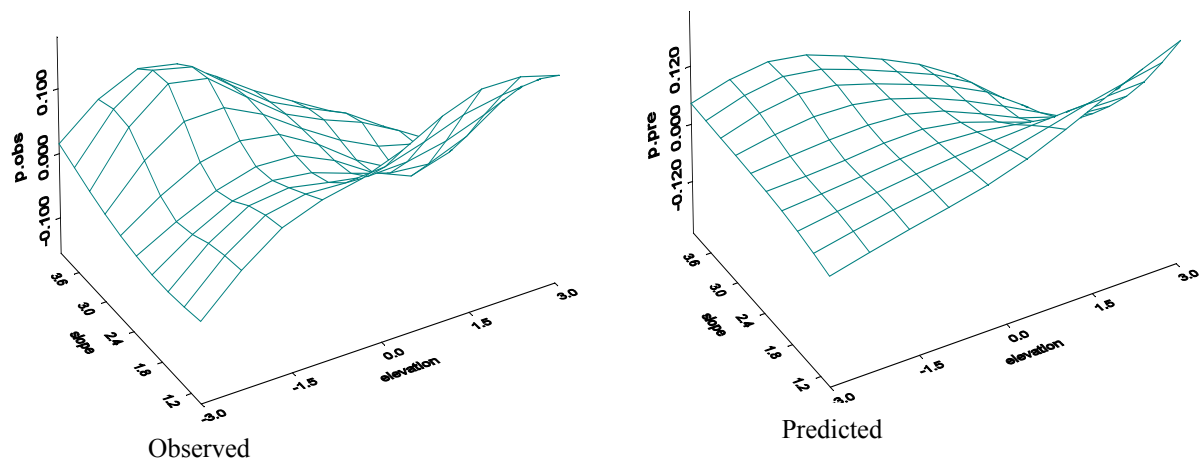


Figure 13. The comparison of observed and predicted surfaces for the velocity model.

Using the estimated transition model in cost function (8), this prediction model was first run with 3 values of TTC. Figure 14 shows that the predicted YST movement was faster in the upper left and lower right quadrants.

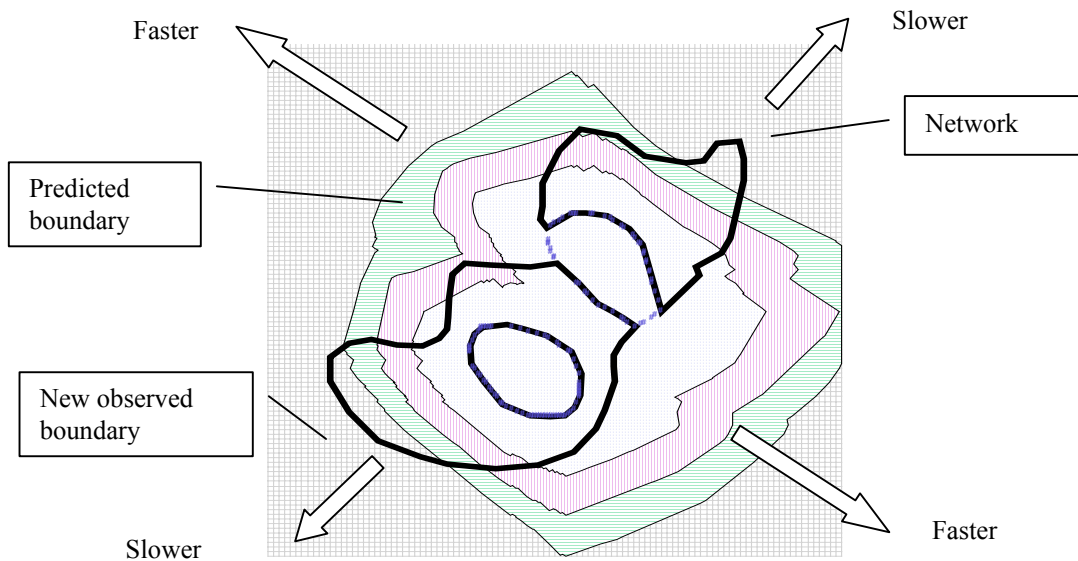


Figure 14. Model 2 prediction results for 3 levels of TTC (cross hatched areas) compared to the observed boundary (solid line).

Using 6 levels of TTC, the associated boundaries were drawn and the related MSEs were calculated. Based on equation (15), OTC was 0.05 and the associated MSE was 34749. The optimum predicted boundary and predicted residuals are shown in Figures 10B and 15, respectively.

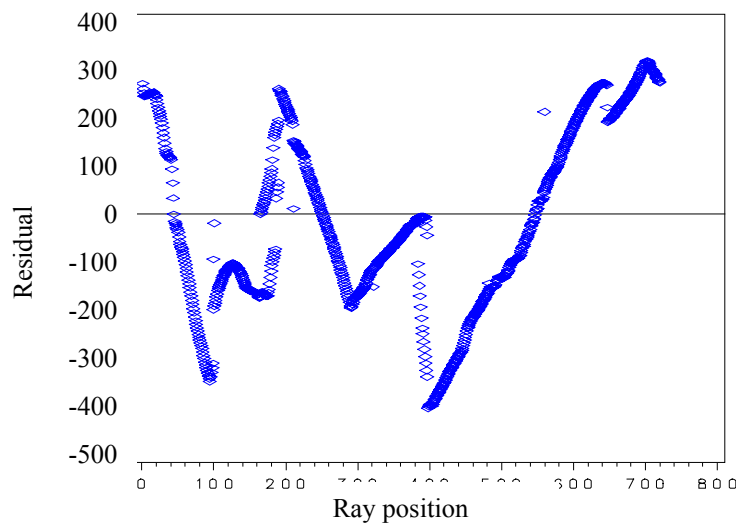


Figure 15. The residuals plot for prediction Model 2.

The mean residual was -19 and the median was -50. All tests for $\mu_0=0$ were significant. Based on both visual inspection and quantitative results, the prediction boundary did not fit the observed data. Upper right and lower left areas were under estimated while the right was over estimated. Figure 15 shows again that residuals are mostly negative, and hence, prediction Model 2 provided inaccurate results.

This is mainly due to the fact that the transition model had more 20 meter length lines appearing at the upper left and lower right sides of the training area. When Pt is large in these regions, then the cost, $(1-P_s*P_t)$, will be small. YST would, therefore, spread at the two lower cost sides faster. Thus, the transition model does not give correct information on YST seed movement.

- **Model 3: SRN arrangement with the velocity model**

Model 3 was similar to Model 2 with the exception that the cost function was replaced with velocity as given in equation (11). As shown above in Model 2, the same training area was used and radii were defined in the same manner.

A close relationship between the length of radii and their respective elevations was observed (Figure 16). The general parabolic shape suggested that a polynomial model may describe this data well.

Therefore, a quadratic model in elevation was estimated using velocity as described in Equation 10. The parameter estimates and standard errors for this model are given in Table 4. All parameters were significant. The predicted curve and 95% bounds are shown in Figure 16.

Parameters	Estimate	Std error
β_0	-1478.7	52.88
β_1	3.51	0.1244
β_2	0.02	0.00007

Table 4. Estimated parameters and their associated standard errors for the velocity model.

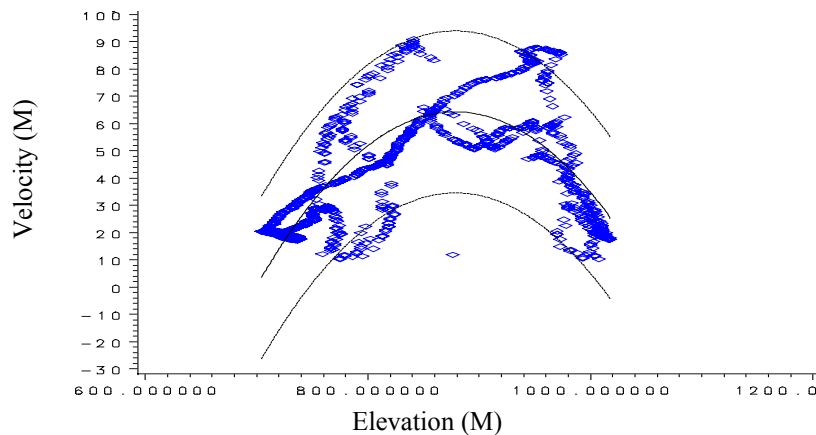


Figure 16. Scatter plot of the radius length versus elevation. Predicted quadratic model and 95% confidence bounds for velocity.

The residuals were acceptable in magnitude and trends indicated an adequate fit .

Using the estimated velocity model, the costs of the links were calculated assuming cost function (11). Three levels of TTC were run. The predicted patterns in Figure 17 are improved over previous models, however, the predicted boundaries are rectangular and do not fit the observed boundary (irregular polygon) well. The upper left and bottom areas are over predicted while the upper right area is under predicted.

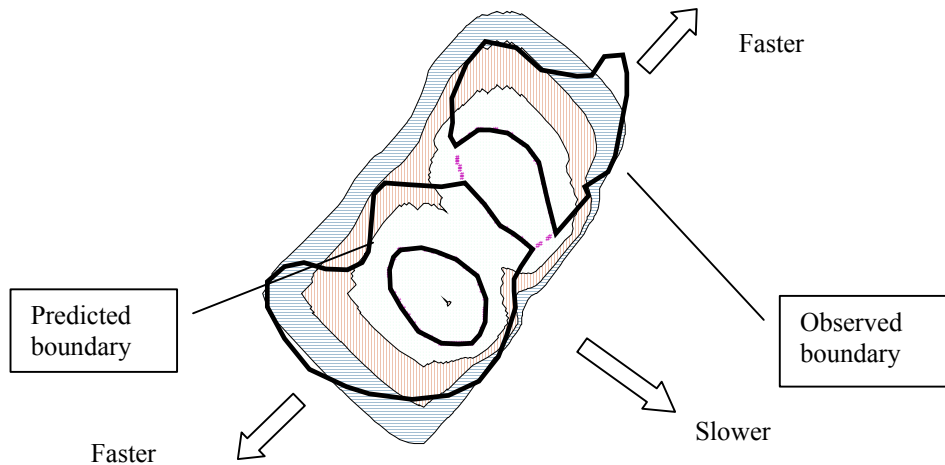


Figure 17. Model 3 predicted results (cross hatched areas) compared to the observed boundary (solid line).

A residual analysis using 6 levels of TTC was conducted. OTC was 0.038 and the associated MSE was 7745. The optimum predicted boundary and predicted residuals are shown in Figures 10C and 18, respectively.

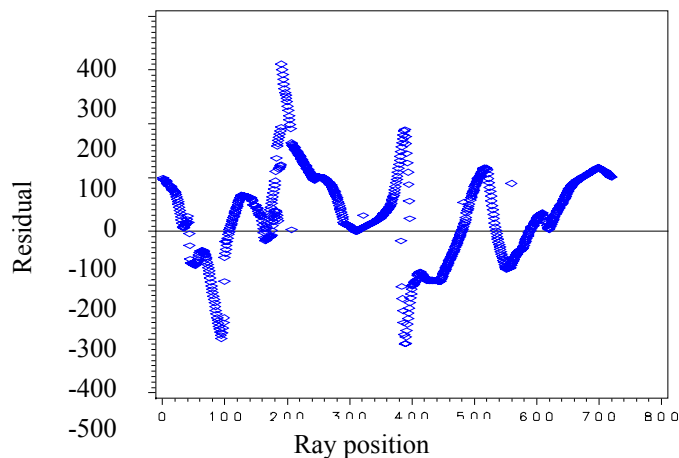


Figure 18. The residuals of prediction Model 3.

The mean residual was 28 and the median was 29, and all tests for $\mu_0=0$ were significant. Figure 18 demonstrates that the majority of residuals were positive. The residuals and visual inspection suggest that the model is over predicting the observed values.

One possible reason for this pattern is that, while the seed of YST can move long distances, it may not mature to produce seeds the next year. Therefore, a survival factor such as the occurrence probability (Ps) may be helpful in adjusting the link costs.

• **Model 4: SRN arrangement with the velocity model and the occurrence model**

Model 4 considers both YST survival and its movement. This was achieved by using the occurrence model in equation (2) as well as the velocity model from equation (10). Using the cost function (12), all costs of links in network were computed and the network was run with 3 levels of TTC (Figure 19). The prediction results were improved and the predicted boundary appeared more dynamic following the changes in the observed data well.

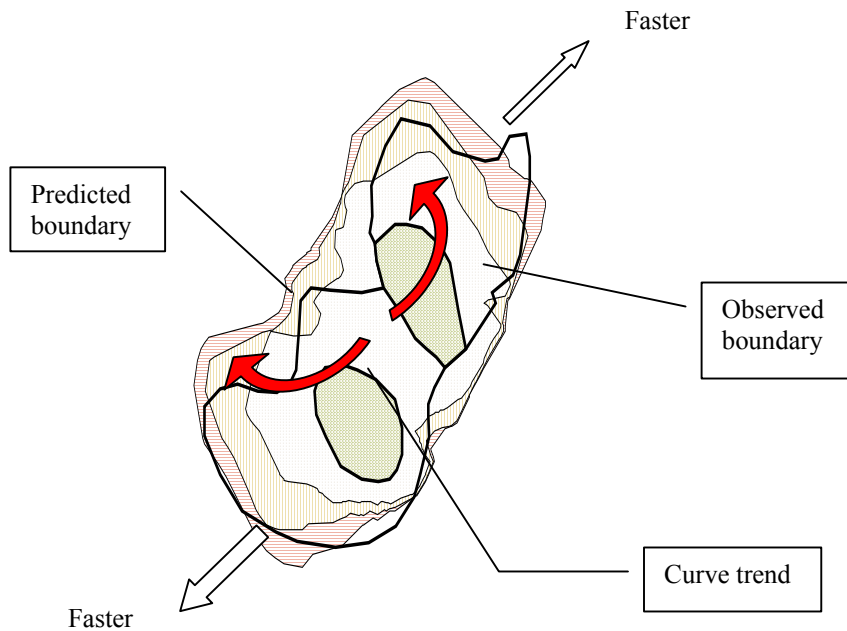


Figure 19. Model 4 predicted results using 3 levels of TTC (cross hatched areas) compared to the observed boundary (solid line).

Further consideration using residual analysis with 6 levels of TTC found an OTC of 0.05 and an associated MSE of 20477. The optimum predicted boundary and predicted residuals are shown in Figures 10D and 20, respectively.

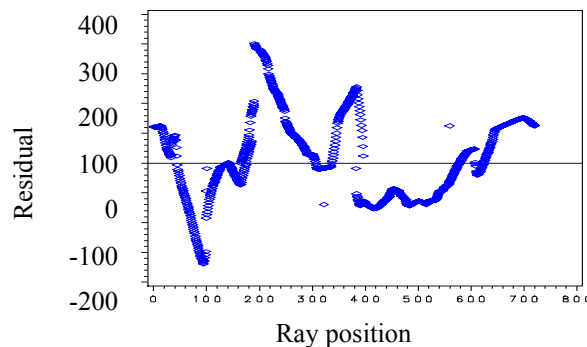


Figure 20. The residuals of prediction in Model 4.

The mean residual was 17 and the median was 4, and all tests for $\mu_0=0$ were significant. Figure 20 demonstrates that fewer residuals were positive compared to the previous model, suggesting less model bias.

A comparison of the four models using a visual assessment method (Figure 10) shows that the patterns of Model 3 and Model 4 appear to be the best. Though Model 3 fits the observed boundaries with less error, Model 4 has a curved shape and follows YST infested trend well. Whereas Model 3 has the smallest MSE, Model 4 has the smallest mean and median residual values and the second smallest MSE of the 4 models. Therefore, considering only data from Moughmer Point, Model 3 is statistically the best with the lowest MSE, however, Model 4 considered both plant survival and seed movement and had the second lowest MSE. This model has better biological relevance and may be preferred over Model 3.

5. Validation

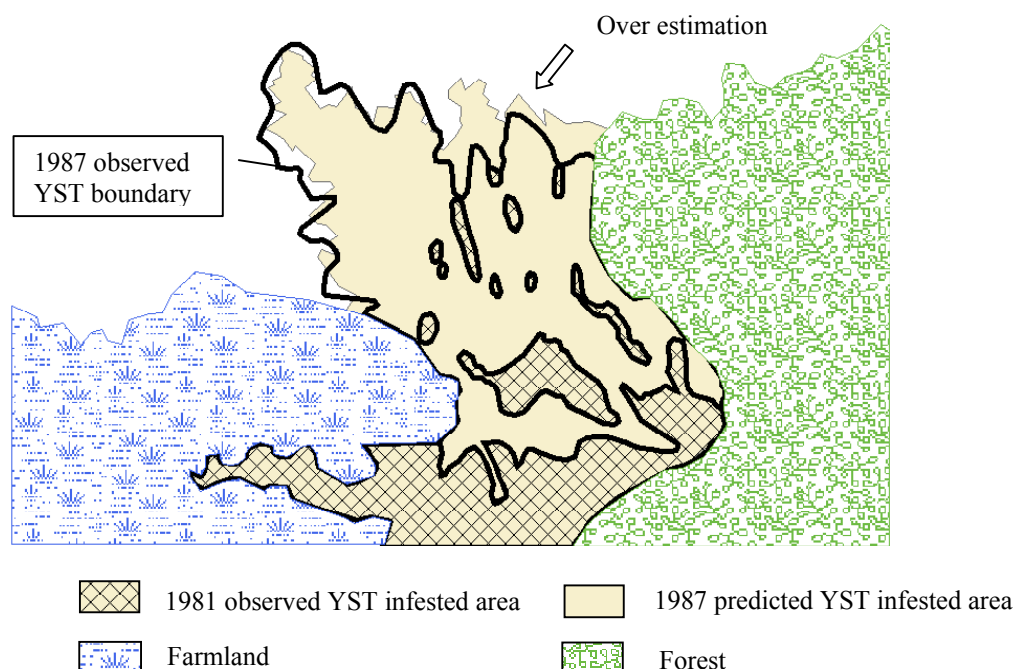


Figure 21. Validation results from Model 4.

Because of data limitations, only spatial validation is considered here. The area chosen for testing was an independent area more than 45 times larger than the training site. The validation network had 497887 links. Areas classified as farmland or forest were clipped or removed from the prediction. Using the parameters developed from the training area, Model 4 was applied to this new data. Based on 6 levels of TTC, a calibrated optimum model was developed with an OTC of 0.87. The optimum prediction boundary fit the observed boundary well with the exception of a small amount of over prediction in the upper right region (Figure 21). The mean residual was 55 and the median was -27. All tests for $\mu_0=0$ were significant. The residuals for the validation are shown in Figure 22.

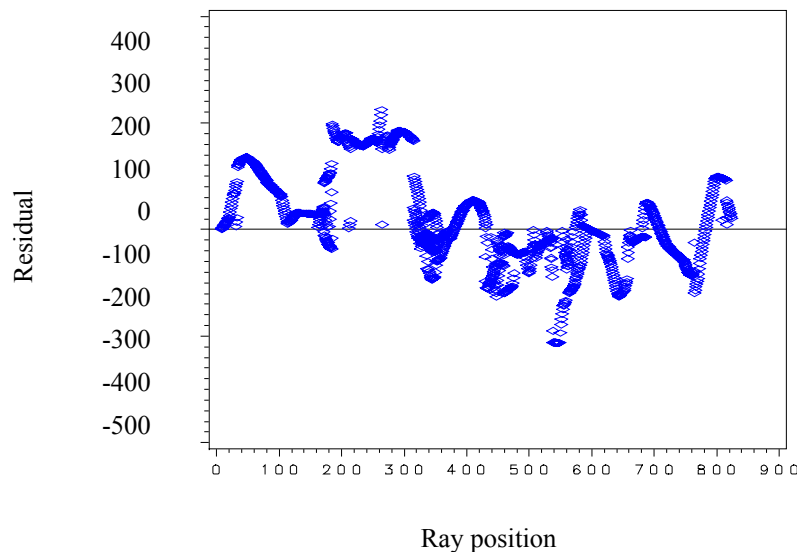


Figure 22. The residuals plot for the validation data.

Relative to the small prediction area, the mean residual value was still quite small. As a whole, the prediction was slightly over estimated while the prediction pattern was good with the exception of the upper right area.

Theoretically, Model 4 considering both survival and movement components would give better prediction results than Model 3. With sufficient training data Model 4 can yield the smallest mean and median of residuals. At the same time, the predicted boundary from Model 4 indicates the ability to produce curved and dynamic trends which fit the observed areas well. Furthermore, validation data also results in very good prediction, especially where YST moves fast (Figure 21). Thus, Model 4 is the best model for prediction of YST infestations.

IV. CONCLUDING REMARKS

This study investigated four prediction models which differed in cost functions based on three regression models and network structure. Model 1 used TIN and its cost function only considered YST survival. Prediction of YST infestation of this model was inadequate with YST moving at the same rate in all directions. Model 2 used SRN and its cost function was related to both plant survival and seed movement. However, its prediction pattern was also inadequate. YST invaded faster in some areas where it should have infested at a slower rate. Model 3 was based on the same network as Model 2, but only considered seed movement developed from a velocity model. The predicted shape of this model was better, but the trend or pattern in the boundary did not fit well. Model 4 used the same network structure and YST movement factor as Model 3, but was based on YST survival factor from a previously developed occurrence model. The associated prediction boundaries of this model fit the observed data very well. After residual analysis and validation, Model 4 had better biological relevance and may be preferred over Model 3. Therefore, the best prediction model for YST dispersal over time and space should consider the mechanisms of YST plant survival and seed movement set out in a SRN arrangement. Land managers could potentially use this model for future predictions to protect canyon rangeland from YST invasion, or limit entry into infested areas.

One advantage of this model is that, even with only two years YST presence data available in small area, the YST dispersal boundary of larger areas of interest can be predicted with high precision. However, there are also some limitations for the model. First, the model is only suitable for the local region where it is developed. Using the model to predict in a completely new region would require new data selection, regression modeling, network construction and prediction model development. A second problem is the requirement of two years YST presence data to fit the velocity model and compute residuals for the calibration model. Multiple year data, however may not always be available. The third concern is that the prediction model only considers natural conditions. The training data for the velocity model, for example, must not include any obstructing human factors, such as forest, farmland, roads etc. Only those radii which are not influenced by human activity should be used to develop the velocity model. Hence, more training areas are required to make up for any lost information if no satisfied data available.

For future research or application the following items should be explored:

- Temporal validation should be considered if more than two years YST presence data is available.
- Parametric or non-parametric simulation could be carried out to assess the behavior (variability) of the prediction model.
- More human activities, such as road and building, should be considered into a new prediction model as dummy variables.
- The network modeling approach may be applied to other plants with different mechanisms of dispersal.

V. REFERENCES

- [ArcView] software,2002. version 3.3. NY:Environment Systems Research Institute.
- [ArcGIS] software,2002. version 8.2. four parts: ArcINFO work station, ArcMap, ArcCatalog, and ArcToolbox. NY:Environment Systems Research Institute.
- Callihan, R.H., Sheley, R.L., and Thill, D.C. 1982. Yellow starthistle: identification and control. Current Info. Series No. 634, 4 pp. University of Idaho, College of Agriculture.
- Chang, Kang-tsung. 2002. *Introduction to geographic information systems*. CH. 16. Pages 305-318. McGraw-Hill Co., New York, NY.
- Cheeke, P.R. and L.R. Shull. 1985. Other plant toxins and poisonous plants. Ch. 11. Pages 358-392. In, *Natural Toxicants in Feeds and Poisonous Plants*. The Avi Publ. Co., Westport, CT.
- Cordy, D.R. 1954. Nigropallidal encephalomalacia (chewing disease) in horses on rations high in yellow star thistle. *Proc., Am. Vet. Med. Assoc.* 91:149-154.
- Fisher, D. H. & Langley, P. 1990. The structure and formation of natural categories. In Bower, G. H., *The Psychology of Learning and Motivation*, 25. San Diego, CA: Academic Press.
- Fitzgerald, R.W., Lees, B.G. 1992. The application of neural networks to the floristic classification of remote sensing and GIS data in complex terrain. In: *Proceedings of the 17th Congress of the International Society for Photogrammetry and Remote Sensing*, Washington, USA.
- Guisan, A. and Zimmermann, N.Z. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* **135** (2000), pp. 147–186.
- Lass,L.W.,J.P.McCaffrey,D.C. Thill, and R.H. Callihan. 1999. Yellow starthistle biology and management in pasture and rangeland. *Univ. Idaho Bulletin No.805*,18 pp.

- Lehmann, A., Overton J.M., and Leathwick, J.R. 2002. GRASP: generalized regression analysis and spatial prediction. *Ecological Modeling* 157(2002)189-207.
- Maddox, D.M. 1981. Introduction, phenology, and density of yellow starthistle in coastal, intercoastal, and central valley situations in California. ARR-W-20, pp. 1-33. USDA-ARS.
- Maddox, D.M., A. Mayfield, and N.H. Poritz. 1985. Distribution of yellow starthistle (*Centaurea solstitialis*) and Russian knapweed (*Centaurea repens*). *Weed Science* 33(3):315-327.
- Miller, J., and J. Franklin. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modeling* 157:227-247.
- Prentice, C. I., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud and A. M. Solomon. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography* 19:117-134.
- Roché, B.F., Jr. 1992. Achene dispersal in yellow starthistle (*Centaurea solstitialis* L.). *Northwest Science* 66:62-65.
- Roché, C.T. and B.F. Roché, Jr. 1988. Distribution and amount of four knapweed (*Centaurea* L.) species in eastern Washington. *Northwest Science* 62:242-253.
- Schwartz, M.W. 1993. Modeling the effect of habitat loss on potential rates of range change for trees in response to global warming. *Biodiversity and Conservation* 2:51-61.
- Shafii, B., W. J. Price, T. S. Prather, L. W. Lass, and D.C. Thill. 2003. Predicting the likelihood of yellow starthistle occurrence using landscape characteristics. *Weed Science*, 51:784-751.