

# A Nonparametric Analysis of the Distribution of *Convolvulus arvensis* in Wheat-Sunflower Rotations

Mario Francisco-Fernández  
Universidad de A Coruña \*

Montserrat Jurado-Expósito                      J. D. Opsomer  
Institute for Sustainable Agriculture <sup>†</sup>      Iowa State University <sup>‡</sup>

Francisca López-Granados  
Institute for Sustainable Agriculture <sup>†</sup>

June 4, 2004

## Abstract

This article describes an application of nonparametric regression to study the spatial structure and identify persistent spatial patterns of the perennial weed *Convolvulus arvensis* L. in four years of wheat-sunflower crop rotation in Southern Spain. The annual spatial distributions of weed patches over the study field are estimated using local linear regression. These are then used to delimit areas whose infestation is above an economic threshold. In order to identify the areas at the highest risk of weed infestation across years, a multi-year index is developed and mapped. A parametric bootstrap is used to quantify the variability of the multi-year map. In a precision agriculture environment, such maps can be a useful component of a long-term weed management strategy.

*Key Words:* local polynomial regression; parametric bootstrap; weed management.

---

\*Departamento de Matemáticas, Facultad de Informática, La Coruña, 15071, Spain.

<sup>†</sup>Institute for Sustainable Agriculture, Spanish Council for Scientific Research (CSIC), Apdo 4084, 14080, Córdoba, Spain

<sup>‡</sup>Department of Statistics, Iowa State University, Ames, IA 50011, USA.

# 1 Introduction

This article describes an application of local linear regression, a popular nonparametric regression method, to the problem of mapping the spatial distribution of the perennial agricultural weed *Convolvulus arvensis* L. in a field over the course of four growing seasons. Local linear regression has been broadly studied in the context of univariate regression, and we refer to Wand and Jones (1995) for an overview. For bivariate local linear regression, Ruppert and Wand (1994) provide the relevant statistical theory for the case in which the errors are independently distributed. In the spatial context, this assumption of independence is often not appropriate, and accounting for possible correlation is required for both inference and smoothing parameter selection. For a review of issues related to nonparametric regression with correlated errors, see Hart (1996) and Opsomer *et al.* (2001). Recently, Francisco-Fernández and Opsomer (2003) discussed spatial smoothing and proposed a bandwidth selection method that allows for the presence of correlated errors.

*C. arvensis*, the plant species of interest in the current article, is a very important perennial weed that infests wheat (*Triticum aestivum* L.) and sunflower (*Helianthus annuus* L.), the main crop rotation in Andalusia (southern Spain). Reduced and no-tillage production has increased in Spain in the last 10 years and now accounts for 2 million ha of annual crops (Anonymous, 1998). An important characteristic of *C. arvensis* is that it produces few viable seeds when growing in competition with

agricultural crops, and reproduces primarily vegetatively by underground rootstock. The adventitious shoots arising from a network of rootstocks reduce crop yields and interfere with the harvest. As many fields of wheat–sunflower rotations have been converted into no-tillage or reduced tillage, perennial weeds like *C. arvensis* have become more troublesome since they cannot be reduced in abundance by repeated tillage or cultivation (Liebman *et al.* 2001).

It is well known that many weed populations have a patchy distribution (Johnson *et al.* 1996), with aggregated weed patches of varying size and density interspersed with areas with few or no weed seedlings. A weed patch is considered stable if it is consistent in density and location over time (Wilson and Brain, 1991). Stability is important from the perspective of patch management, since knowledge of the location of patches with high weed density can be used to direct weed control in subsequent years. This is especially true for perennial weeds in reduced tillage systems, where ploughing and cultivation are no longer considered acceptable management options and where farmers want to make informed decisions on the precise use of herbicides (Webster *et al.* 2000).

Currently, herbicides are most often applied to the entire field even though spraying might be unnecessary in some places. An important goal of site-specific weed control is to apply herbicide only in areas where weed density exceeds an economic threshold (ET) (Dammer *et al.* 1999). Such an approach has the potential for significantly reducing herbicide use, especially if the location of the weed patches could be determined before the weed plants are fully established. The average reduction from

site-specific weed control in cereals ranged from 47% to 80% (Heisel *et al.* 1996a). In maize, Tian *et al.* (1999) realised savings of 42%; and Timmermann *et al.* (2001) reported that with a site-specific weed control an average of 54% of the herbicides could be saved in sugar beet. In sunflower, Jurado-Expósito *et al.* (2003) achieved an average reduction in herbicide cost around 61% if a given herbicide were applied just to the areas exceeding the ET. In order to implement a site-specific herbicide application strategy, a weed patch distribution map is required.

Previous work on the mapping of weed patches includes linear triangulation (Gerhards *et al.* 1997), polynomial interpolation (Zanin *et al.* 1998), or kriging. Kriging weighs the average of observed weed densities and is the only current estimator that estimates the variance (Cressie 1993, pp.183-194; Isaaks and Srivastava 1989). The adoption of this approach in weed research has been very recent, but has shown to be useful in quantifying the spatial structure of weed populations (González-Andújar *et al.* 2001, Heisel *et al.* 1996b, Jurado-Expósito *et al.* 2003). However, kriged estimates used in this field assume a constant and known trend. When this assumption is violated, model misspecification bias can result.

In this article, we develop maps for the spatial distribution of *C. arvensis* over the course of the 1999-2002 growing seasons, for a sample of 261 locations at which the number of plants of *C. arvensis* were counted. Using the concept of economic thresholding discussed above, we identify the areas in each annual map where herbicide applications are warranted under a site-specific weed control strategy. In addition to the individual year analyses, we will also use a bootstrap-based method

to generate a map that shows the likelihood of being at high risk of infestation across years. Such a map can be used in subsequent years to predict the places that are most likely to be affected by *C. arvensis*. Such a multi-year map can be useful to formulate and implement site-specific weed control strategies that take the persistent nature of *C. arvensis* infestations into account.

The organization of the remainder of this article is as follows. Section 2 describes the statistical model and reviews the nonparametric estimator. Section 3 provides information on the study area and the sampling design used to obtain the data. In Section 4, we describe the results of the data analysis.

## 2 Local linear regression for spatial data

We briefly describe the spatial nonparametric regression model to be used for the *C. arvensis* data. Assume that a set of  $\mathbb{R}^3$ -valued random vectors,  $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ , are observed, where the  $Y_i$  are scalar responses variables and the  $\mathbf{X}_i$  are predictor variables with a common density  $f$  and compact support  $\Omega \subseteq \mathbb{R}^2$ . In this article, we will refer to the  $\mathbf{X}_i$  as the *locations* corresponding to the  $Y_i$ . The relationship between the locations and the responses variable is assumed to be of the form

$$Y_i = m(\mathbf{X}_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where

$$\mathbb{E}(\varepsilon_i | \mathbf{X}_i) = 0, \quad \text{Var}(\varepsilon_i | \mathbf{X}_i) = \sigma^2,$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j | \mathbf{X}_i, \mathbf{X}_j) = \sigma^2 \rho(\mathbf{X}_i - \mathbf{X}_j),$$

with  $\rho(\mathbf{x})$  continuous, satisfying  $\rho(\mathbf{0}) = 1$ ,  $\rho(\mathbf{x}) = \rho(-\mathbf{x})$ , and  $|\rho_n(\mathbf{x})| \leq 1$ ,  $\forall \mathbf{x}$ . The presence of the function  $\rho$  implies that the observations are spatially correlated. Francisco-Fernández and Opsomer (2003) discuss the asymptotic framework under which a local linear regression estimator for this model is consistent.

The estimator for  $m(\cdot)$  at the location  $\mathbf{x}$  is the solution for  $\alpha$  to the least squares minimization problem

$$\min_{\alpha, \boldsymbol{\beta}} \sum_{i=1}^n \{Y_i - \alpha - \boldsymbol{\beta}^T(\mathbf{X}_i - \mathbf{x})\}^2 K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}),$$

where  $\mathbf{H}$  is a  $2 \times 2$  symmetric positive definite matrix;  $K$  is a bivariate kernel and  $K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-1} K(\mathbf{H}^{-1}\mathbf{u})$ . The Epaneshnikov kernel function  $K(\mathbf{x}) = \frac{2}{\pi} \max \{ (1 - \|\mathbf{x}\|^2), 0 \}$  will be used throughout this article. The bandwidth matrix  $\mathbf{H}$  controls the shape and the size of the local neighborhood used for estimating  $m(\mathbf{x})$ . The local linear regression estimator can be written explicitly as

$$\hat{m}(\mathbf{x}; \mathbf{H}) = e_1^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{Y} \equiv \mathbf{s}_x^T \mathbf{Y} \quad (2)$$

where  $e_1$  is a vector with 1 in the first entry and all other entries 0,  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ ,  $\mathbf{W}_x = \text{diag} \{ K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}), \dots, K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x}) \}$ , and

$$\mathbf{X}_x = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \end{pmatrix}.$$

This estimator depends on the choice of the values used in the matrix bandwidth  $\mathbf{H}$ . We will base our choice of bandwidth values on the “correlation-corrected” generalized cross-validation method of Francisco-Fernández and Opsomer (2003), who

showed both theoretically and through simulations that this method works well for spatially correlated data, as long as the correlation can be reasonably approximated by a smoothly decaying function of distance between locations.

### 3 Description of Data

The data were collected during the course of four growing seasons (1999–2002). The four surveys were conducted in a field of about 1.6 ha located at Monclova (La Luisiana, Seville), within one of the most important and technologically advanced farming areas in Andalusia, southern Spain. The field site was farmer-managed using no-tillage production methods. Wheat (*Triticum aestivum* L.) was sown in 1999 and 2001, and sunflower (*Helianthus annuus* L.) in 2000 and 2002. Conventional herbicides practices for weed control were used. Glyphosate was applied preemergence at a rate of 2 L/ha for the control of annual weed seedlings in wheat and sunflower. At these rates the herbicides had no significant activity on perennial shoots of *C. arvensis*.

Weed density was sampled in early May before crop harvesting. An area measuring 65 m wide by 250 m long was selected for the intensive survey in 1999, and the same area was sampled again in subsequent years. The survey area was located in a larger field of approximately 40 ha, and its borders were at least 50 m from the main borders of the field. Crop rows were always oriented south-north across the study area during the course of the study. *C. arvensis* density assessments were performed following an approximate 7- by 7-m grid pattern, resulting in a total of 261 sampling

units. The position of each grid point was georeferenced using a Differential Global Positioning System (DGPS) and recorded in UTM coordinates. At each node, the number of plants of *C. arvensis* were counted in a 2 by 2 m square. Figure 1 shows the 261 locations where the number of plants were counted each year. A portion of the data used in this study were previously studied by Jurado-Expósito *et al.* (2003), where a more detailed description of the study design and measurements is provided.

[Figure 1 about here.]

Based on previous research, the economic threshold (ET), i.e., the *C. arvensis* density causing a reduction in net wheat or sunflower yield equal to the control treatment cost, was estimated at 14 plants/m<sup>2</sup> (Castro-Tendero and García-Torres, 1995). Hence, if site-specific herbicide application is to be used in this field every year, the location of persistent patches of *C. arvensis* with density exceeding ET should be located and treated. This cut-off value allows us to produce target herbicide application maps for each year. Moreover, the percentage of saved herbicide compared to wall-to-wall application is readily estimated from these maps.

In the case of perennial weeds like *C. arvensis*, there is a clear interest not only in locating the high density patches in any given year, but also to determine the location of areas most at risk of multi-year infestations. If those areas in particular can be targeted for treatment, it might be possible to further reduce the long-term treatment needs for the overall field. Because the crop planted varies across years,



the size and exact location of *C. arvensis* “clumps” change in character, so that a multi-year map will need to be able to incorporate such heterogeneity.

## 4 Results

We begin by fitting the nonparametric regression model (1) to the data from each year separately. In order to produce a map for the survey area of interest, the local linear estimates were computed on a dense regular  $200 \times 200$  grid overlaying the field. Use of the “corrected GCV” method for bandwidth selection of Francisco-Fernández and Opsomer (2003) requires the selection of a pilot model for the correlation. We used the exponential model

$$\rho(\mathbf{x}) = \exp(-\alpha \|\mathbf{x}\|) \quad (3)$$

for this purpose, where  $\alpha$  is an unknown parameter, and fitted that model to the residuals of a pilot local linear regression fit. Visual inspection of the plots (not shown) comparing observed and model predicted correlations at a range of distances indicated that the exponential model fitted the data reasonably well. Note that this model specification is used only in the selection of bandwidth values, and does not determine the actual shape of the spatial distribution function  $m(\cdot)$  in (1).

The bandwidth selection method was applied to each of the four years individually, and resulted in bandwidth matrices that had similar but not identical characteristics. In order to avoid introducing differences between the years due to the fitting method, it was decided to use a single bandwidth matrix for all the years,

by taking the average of the individual bandwidth matrices. Hence, the bandwidth matrix used for all years was equal to

$$\mathbf{H} = \begin{bmatrix} 44.8113 & 10.3409 \\ 10.3409 & 39.4077 \end{bmatrix},$$

corresponding to a moderate-to-large amount of smoothing.

[Figure 2 about here.]

Figure 2 shows the estimated weed densities for the four years obtained using this bandwidth matrix. A visual assessment reveals distinct aggregation of infested areas for all years. The *C. arvensis* populations are more highly aggregated in sunflower years (Fig. 2, 2000 and 2002) compared to the wheat years, and a higher amount of surface area was free of *C. arvensis* plants in sunflower years. This implies that when growing in competition with sunflower, *C. arvensis* patches were less and smaller than in competition with wheat. However, crop rotation did not explain all of the observed year-to-year variability, with significantly different patterns emerging for the two pairs of years with the same crop.

If we want to apply annual location-specific herbicide treatments to this field using the ET as a guide, Figure 3 displays which areas would need to be treated at the threshold value of 14 plants/m<sup>2</sup> mentioned in the previous section. As these maps show, a large fraction of the field would have to be treated every year except for the year 2000. Specifically, the fractions of the surface area that would have to be treated annually according to this rule are 76.0% in 1999, 52.6% in 2000, 88.3% in 2001 and 78.0% in 2002.

[Figure 3 about here.]

[Figure 4 about here.]

While some parts of the field are included in the area to be treated each year, others appear to need treatment during some years but not during others. Overall, spatial stability of the main high-density focal points was observed for each *C. arvensis* seedling patch map across the four years survey (Fig. 2). If the areas most prone to weed infestation across all years could be identified, this knowledge could guide weed scouting and weed management. This information could also be a key component of a precision agriculture system. In order to develop a multi-year map for the persistence of *C. arvensis* patches in the field, we implemented the following simple rule after some experimentation with alternatives:

a field point is considered significantly at risk of infestation if its weed density exceeds the ET in at least 3 of the 4 years.

Figure 4 displays the resulting spatial distribution map.

[Figure 5 about here.]

While the above analysis is useful in identifying portions of the field that are most vulnerable to infestation by *C. arvensis* in individual years as well as across years, so far there is no accompanying measure of variability. In particular, the map in Figure 4 is sensitive to locations whose estimated weed densities are highly variable, as is likely to happen on the boundaries of the study region, or to locations

whose weed densities are close to the ET in any given year. For instance, it is not clear how to interpret the irregular boundary region seen in the South-East corner of Figure 4, or the small area in the center top of the plot. In order to address these problems, we decided to supplement the above analysis by a bootstrap-based evaluation.

For each year, a bootstrap dataset is generated by taking the estimated mean spatial distribution (as shown in Fig. 2), and adding bootstrap errors generated as a spatially correlated set of errors with known distribution. The annual distribution functions for generating these bootstrap errors are obtained by taking the residuals from the original regression for that year and fitting the exponential model in (3). Once the annual bootstrap datasets are obtained, the above annual nonparametric regressions are repeated for each bootstrap sample using the same bandwidth  $\mathbf{H}$  as for the original analysis, and a bootstrap map of at-risk areas as in Figure 4 is produced. This process is repeated 1000 times. The result is the map in Figure 5, which displays the frequency, across bootstrap replicates, for each location of how often that location is included in the at-risk area. This analysis shows that the area of most concern is primarily the South-West quadrant of the field, while the areas stretching in the North-East and South-East observed in Figure 4 have become somewhat “fainter,” in the sense that their probability of being at risk are lower.

[Figure 6 about here.]

In order to evaluate the sensitivity of this procedure to the choice of ET, we

repeated the complete analysis for higher values of the threshold,  $ET=16$ ,  $18$  and  $20$ . Figure 6 displays the results. The main difference is that the overall area of concern shrinks, with the North-East and South-East components almost completely disappearing for higher  $ET$ , while the South-West quadrant continues to be identified as the main problem area.

## 5 Conclusion

In this article, we have developed a method for displaying the portion of a field that is most at risk for infestation by perennial weeds, through a combination of nonparametric regression and a parametric bootstrap. Our method used an economic threshold on weed patch density to determine what constitutes an “at risk” location as well as a heuristic rule to combine data from different years. Both of these features of our method could easily be extended or customized for different situation, for instance by having different threshold for different crops or by replacing the multi-year measure by a different type of thresholding altogether. The overall approach of spatial smoothing and bootstrap-based density mapping is readily implemented, and the resulting map should be particularly useful in a precision agriculture environment as part of a long-term weed management strategy.

## 6 Acknowledgments

The research of Mario Francisco-Fernández has been partially supported by Grant PGIDIT03PXIC10505PN and MCyT Grant BFM2002-00265 (European FEDER

support included). The research of Francisca López-Granados and Montserrat Jurado-Expósito has been partially supported by a CICYT-FEDER project AGL2002-04468-C03-02 and a Grant I3P-CSIC.

## References

- [1] Anonymous (1998). *Guía de agricultura de conservación de cultivos anuales*. Ed. Asociación Española de Agricultura de Conservación/ Suelos Vivos, pp 35.
- [2] Castro-Tendero, A. J. and García-Torres, L. (1995). SEMAGI- an expert system for weed control decision making in sunflowers. *Crop Protection* **14**, 543–548.
- [3] Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, 2nd edition.
- [4] Dammer, K-H., Schweigert, T. and Wittmann, C. H. (1999). Probability maps for risk assessment in a patchy weed control. *Precision Agriculture* **1**, 185–198.
- [5] Francisco-Fernández, M. and Opsomer, J. D. (2003). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. Preprint #03–07, Department of Statistics, Iowa State University.
- [6] Gerhards, R., Wyse-Pester, D. Y., Mortensen, D. and Johnson, G. A. (1997). Characterizing spatial stability of weed populations using interpolated maps. *Weed Science* **45**, 108–119.

- [7] González-Andújar, J. L., Martínez-Cob, A., López-Granados, F. and García-Torres, L. (2001). Spatial distribution and mapping of *Orobanche crenata* infestation in continuous *Vicia faba* cropping for six years. *Weed Science* **49**, 773-779.
- [8] Hart, J. (1996). Some automated methods of smoothing time-dependent data, *Journal of Nonparametric Statistics*, **6**, 115–142.
- [9] Heisel, T., Christensen, S. and Walter, A. M. (1996a). Weed managing model for patch spraying in cereal. Precision agriculture. Proceedings of the 3rd International conference, Minneapolis, Minnesota, USA, eds. Robert, P. C., Rust, r. H., Larson, W. E. 999-1007.
- [10] Heisel, T., Andersen, C. and Ersbøll, A. K. (1996b). Annual weed can be mapped with kriging. *Weed Research* **36**, 325–337.
- [11] Isaaks, E.H., Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*. Oxford University. New York.
- [12] Johnson, G. A., Mortensen, D. A. and Gotway, C. A. (1996). Spatial and temporal analysis of weed seedling populations using geostatistics. *Weed Sci.* 44:704-710.
- [13] Jurado-Expósito, M., López-Granados, F., García-Torres, L., García-Ferrer, A., Sanchez de la Orden, M. and Atenciano, S. (2003). Multi-species weed spatial

- variability and site-specific management maps in cultivated sunflower. *Weed Science* **51**, 319-328.
- [14] Liebman, M., Mohler, C.L., Staver, C.P. (2001). *Ecological Management of Agricultural Weeds*. Cambridge University Press, Cambridge.
- [15] Opsomer, J. D., Wang, Y. and Yang, Y. (2001). Nonparametric regression with correlated Errors. *Statistical Science* **16**, 134–153.
- [16] Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics* **22**, 1346–1370.
- [17] Tian, L., Reid, J. F. and Hummerl, J. W. (1999). Development of a precision sprayer for site-specific weed management. *Transactions of the ASAE* **42**, 893-900.
- [18] Timmermann, C., Gerhards, R., Krohmann, P., Sökefeld, M. and Kühbauch, W. (2001). The economical and ecological impact of the site-specific weed control. Third European Conference on Precision Agriculture, Montpellier, France. Eds: G. Grenier and S. Blackmore. 563-568.
- [19] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman and Hall, London.
- [20] Webster, T.M. and Cardina, J. (1998). Spatial and temporal emergence patterns of hemp dogbane (*Apocynum cannabinum*). In: Proceedings of the Weed Science Society of America. Chicago, IL. **38**, 9-12.



- [21] Wilson, B. J. and P. Brain. (1991). Long-term stability of distribution of *Alopecurus myosuroides* Huds. within cereal fields. *Weed Research* **31**, 367–373.
  
- [22] Zanin, G., Berti, A. and Riello, L. (1998). Incorporation of weed spatial variability into the weed control decision-making process. *Weed Research* **38**, 107–118.

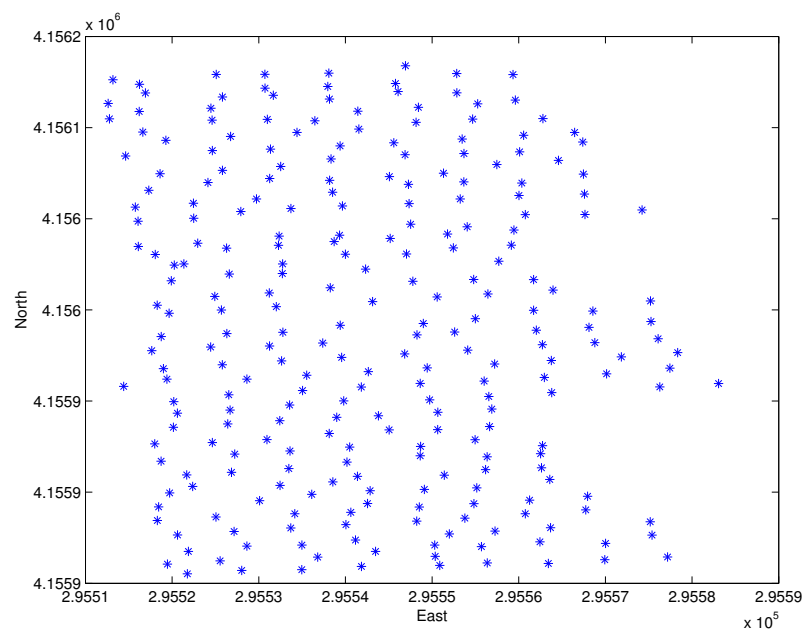


Figure 1: Locations (in UTM East/North coordinates) of *C. arvensis* measurements.

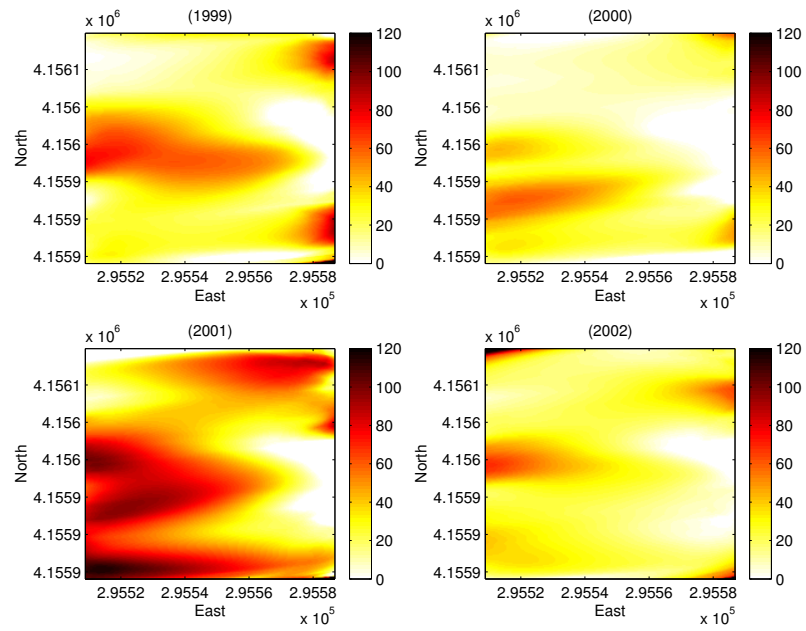


Figure 2: Local linear regression of *C. arvensis* distribution in years 1999–2002 (plants/m<sup>2</sup>).

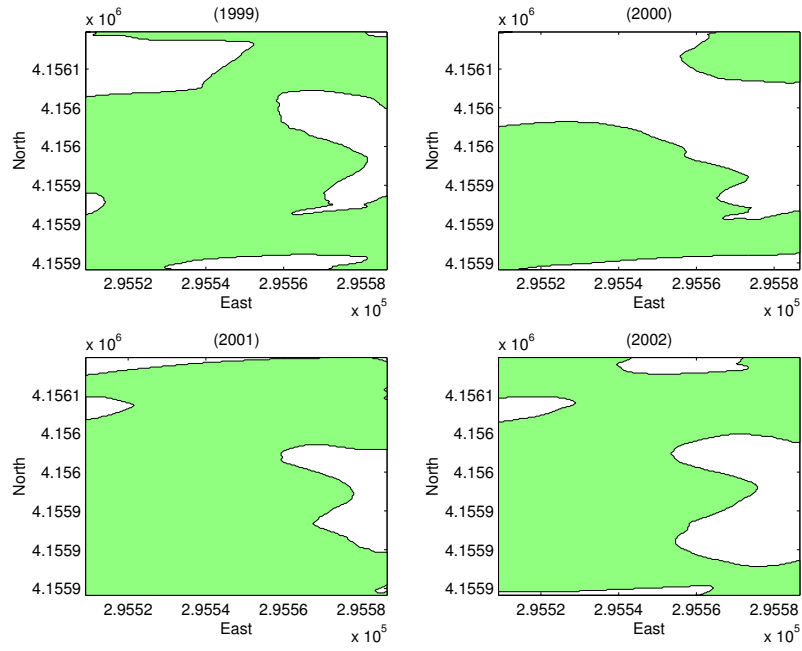


Figure 3: Site-specific herbicide application maps obtained for  $ET \geq 14$  weeds/ $m^2$ ; shaded areas are those needing herbicide treatment.

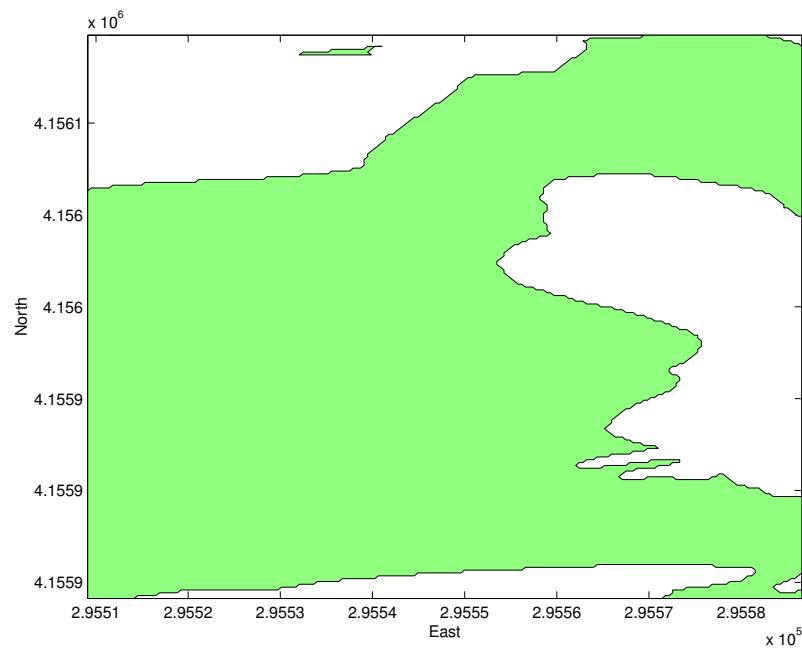


Figure 4: Multi-year map of areas at risk (green) and not at risk (white) of persistent *C. arvensis* infestation.

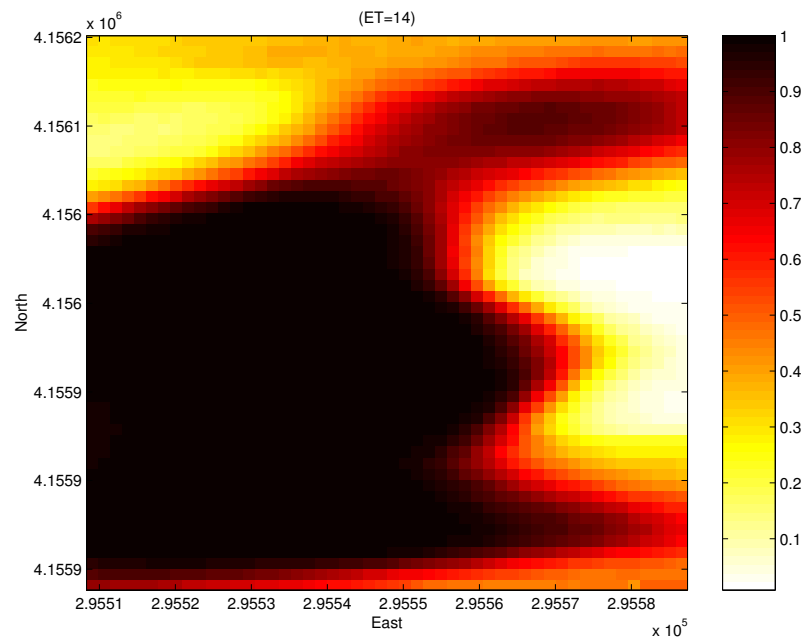


Figure 5: Map with pointwise bootstrap probabilities of being considered at risk of infestation.

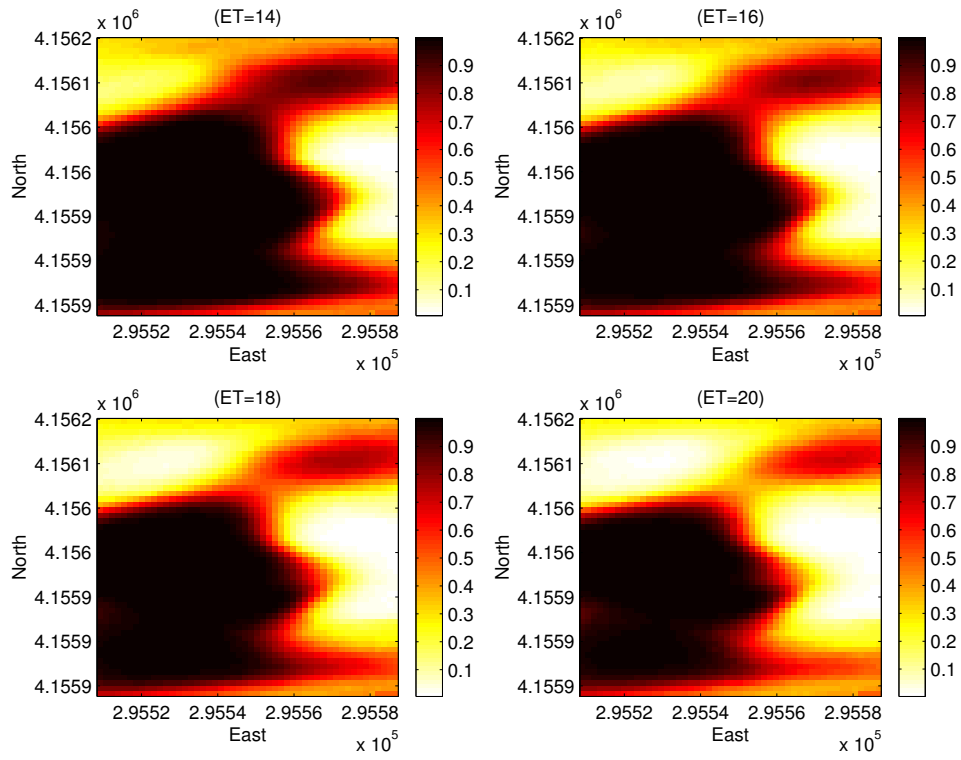


Figure 6: Map with pointwise bootstrap probabilities of being considered at risk of infestation, for different values of the economic threshold.