ESTIMATING THE WINTERING LOCATION OF THE WOOD THRUSH

HANNA JANKOWSKI, VERONICA SABELNYKOVA, JOHN SHERIFF

Abstract. We consider the estimation of the wintering location of the wood thrush based on geolocator data. Some challenges associated with the method-specific errors observed in the data are discussed.

1. Introduction

An ongoing study, led by Professor Bridget Stutchbury of York University, follows the migration journey of songbirds. Songbirds are too small for satellite tracking, and previously scientists had relied on the use of radio transmissions "the best previous data ... came from a researcher who followed the birds in a plane for about four days – in 1973" (The Globe and Mail, 2009). The Stutchbury lab has employed a new method of bird tracking using geolocators. The geolocator is mounted on the bird (not unlike a small backpack) and collects sunrise and sunset data daily. Once the bird is captured, the data from the geolocator is downloaded, and the sunrise and sunset times may be translated into latitudes (length of day) and longitudes (sunrise time) giving information as to the birds' location on a daily basis.

The breakthrough of this method is that it allows biologists to generate the location of individual birds, a task previously impossible. However, the method is also highly prone to measurement error. The readings are based on light levels, which can be affected by clouds or even time spent in shaded locations - a particular problem for the wood thrush. Furthermore, during the equinox, the length of day gives no information as to the latitude.

This report studies different methods of estimating the wintering location of the wood thrush based on the collected data.

2. Data Collection and Description of Methodology

Wintering locations were observed for a total of 14 birds. For each bird, daily location data was collected. In addition, the Stutchbury lab determined which observations were taken during the wintering period for each of the birds. During the equinox, only longitudes were observed, and some data were deemed as outliers and removed from the data set. The original approach of Stutchbury et al. (2009) was to consider a mean and standard deviation approach to find the centre and range of the

---

Date: September 13, 2011.

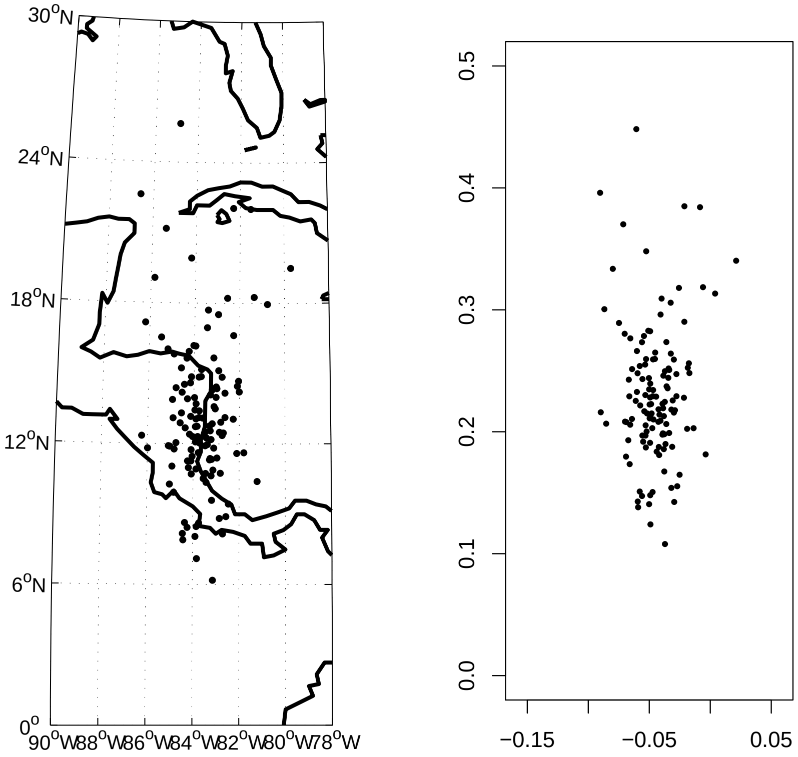

FIGURE 1. Original location data for wood thrush 10 (left) and after translation into UTM coordinates (right).

wintering location of the birds. However, the measurement error is expected to be skewed, especially in the north-south axis. We therefore consider methodology which allows for such behaviour in the errors. As in Stutchbury et al. (2009), we consider only those data points for which both a longitude and latitude were measured.

First, we transform all longitude and latitude values into universal transverse mercator (UTM) coordinates Y_{ij}. As the overall range of the observations was sufficiently small, this allows us to work on a flat surface, as opposed to a spherical one. Figure 1 shows the data points for wood thrush 10 in the original and UTM coordinate systems. Both plots show that the data is highly skewed and is not elliptical in shape. Most points, however, are highly clustered around the east coast of Nicaragua. The plot and UTM conversions were done in MATLAB using the `m_map` software (Pawlowich, 2005). The UTM plots were done using the shareware R program (R Development Core Team, 2011).

Let Y_{ij} denote the ith observation for the jth bird. For each bird, we are interested in recovering an appropriate "measure of centre". In the winter, the birds are largely sedentary, with a reported range of 0.12-1.03 ha (Roth et al., 2011; Stutchberry, 2009). Thus, we may assume that each bird is essentially stationary at some location, and the variability in observations is due entirely to error. This allows us to further assume that the observations Y_{ij} are independent within and between birds. As our measure of centre, we consider both the mean and the median either (a) assuming nothing about the underlying distribution, or (b) assuming that the underlying distribution is log-concave.

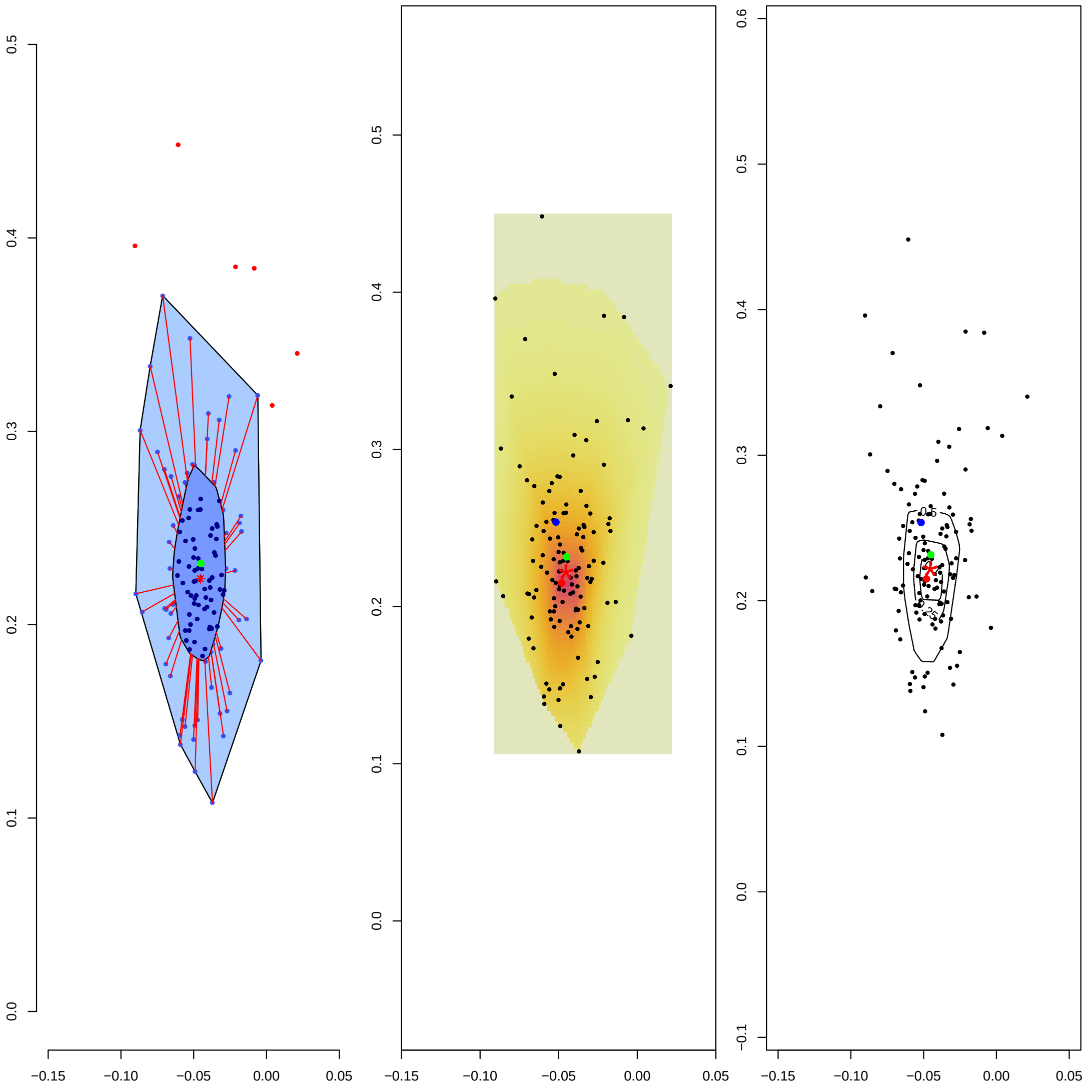

FIGURE 2. Bagplot (left) and fitted logconcave density (centre and right) for wood thrush 10.

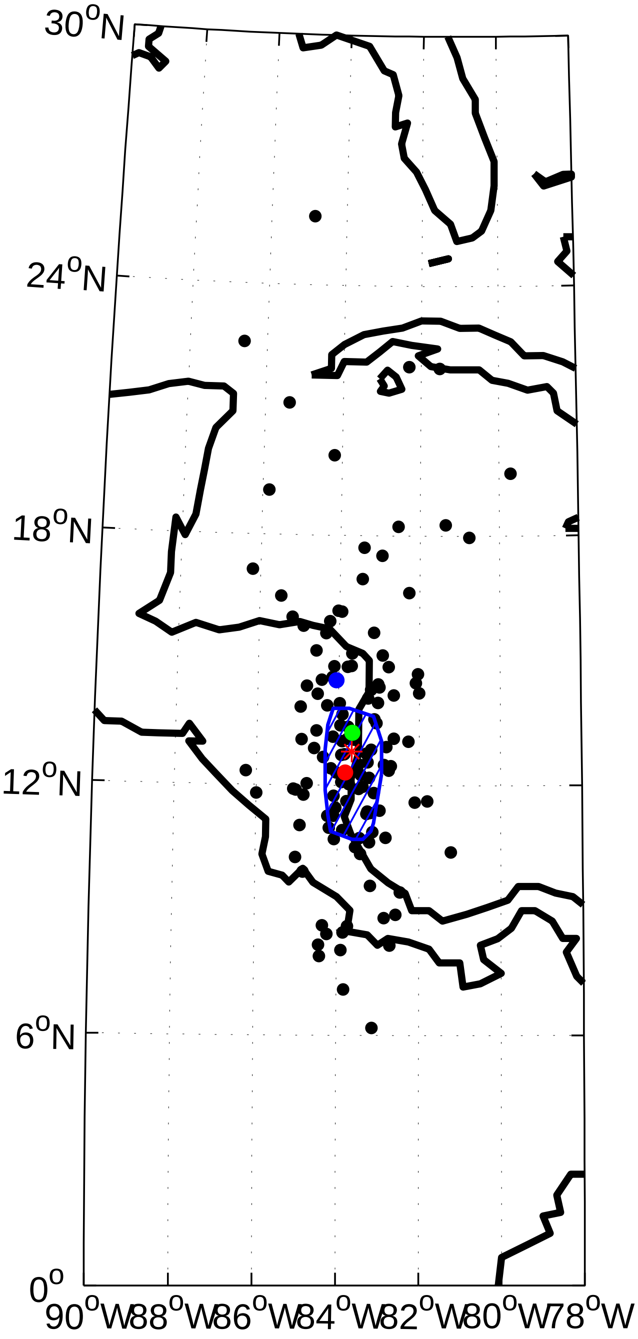

Figure 3. The depth median of data (red star) along with the mode/mean of the fitted logconcave density (red/blue) as well as the empirical mean (green) shown in latitude and longitude coordinates.

The empirical mean of the distribution is calculated in the standard way. To find the *multivariate* median, we use the so-called "depth median" introduced in Donoho and Gasko (1992). This is easily calculated, along with the bivariate boxplot called the bagplot, using the package `aplpack` (Wolf, 2010) available through the R project (R Development Core Team, 2011).

A log-concave distribution is one where the density has the form $f(x) = \exp(\varphi(x))$, for some concave function $\varphi(x)$. The appeal of the log-concave shape-constraint is that this class of densities includes many of the popular models, such as the bivariate normal, but allows for skewness. Log-concave estimation has received much attention recently in the statistics literature, notably Cule et al. (2010); Cule and Samworth (2010). To fit a log-concave distribution we use the R package `LogConcDEAD` (Cule et al., 2009). We note that the estimation procedure is that of maximum likelihood.

Figure 2 shows the results of both fitting procedures. The bagplot along with the empirical mean (green dot) and depth median (red star) are shown in the leftmost plot of the figure. The bagplot, developed in Rousseeuw et al. (1999), is the bivariate version of the popular boxplot. Here, the dark blue region denotes the convex hull of 50% of the data which is closest to the median. The points outside of the light blue region are considered to be outliers.

The other two plots in Figure 2 show a colourmap of the fitted bivariate log-concave density (centre) as well as the 10%, 25%, and 50% highest density regions of this density (right). Both plots (centre and right) show the mean (blue dot) and mode (blue dot) of the fitted density. The empirical mean and median are also shown for comparison. We can see that the fitted log-concave density is highly skewed to the north, which matches well our understanding of the errors of data collection: cloudy days decrease the amount of sunlight resulting in observations which appear further north than in reality.

The four estimators of centre are shown in latitude and longitude coordinates in Figure 2. Of the four centres of measure, the mode of the fitted density appears to be most consistent with our apriori belief of the location of the bird based on our understanding of the error behaviour. We therefore choose this as our measure of centre in what follows.

3. RESULTS

The methods discussed above were implemented on the data for each of the 14 songbirds, see Figure 4 (more detailed plots are shown in the Appendix). Based on these results, we can see that wood thrush 6 is a highly atypical bird, and is hence removed from further analysis. Songbirds 3, 4, 7, 8, and 11 all have centre estimates falling off land. One possible reason for these findings is that the missing data (outliers and equinox values) is skewing the observations resulting in the observed bias. However, for birds 3 and 7 the 10% highest density region intersects land, and for birds 4, 8, and 11 the 25% highest density region intersects land. Our intuition is

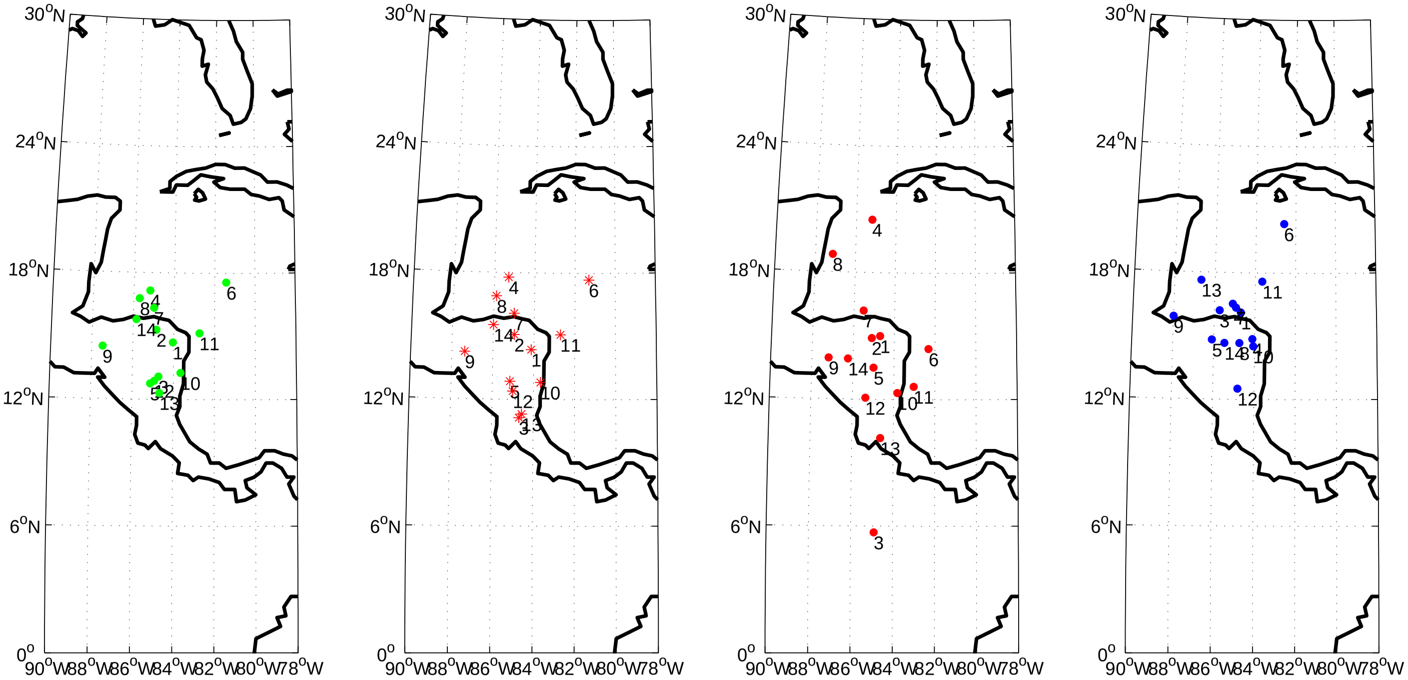

FIGURE 4. All four estimators for the fourteen songbirds. From left to right the plot shows the empirical mean (green dots) and median (red stars), and the log-concave mode (red dots) and mean (blue dots)

that, due to the slow convergence rate of the mode estimate, there is error in these estimates. However, the highest density regions, when intersected with land, should give a reasonable indication of where the bird is likely to be.

The final estimates of the wintering locations are shown in Figure 5 (25% highest density region) and 6 (10% highest density region).

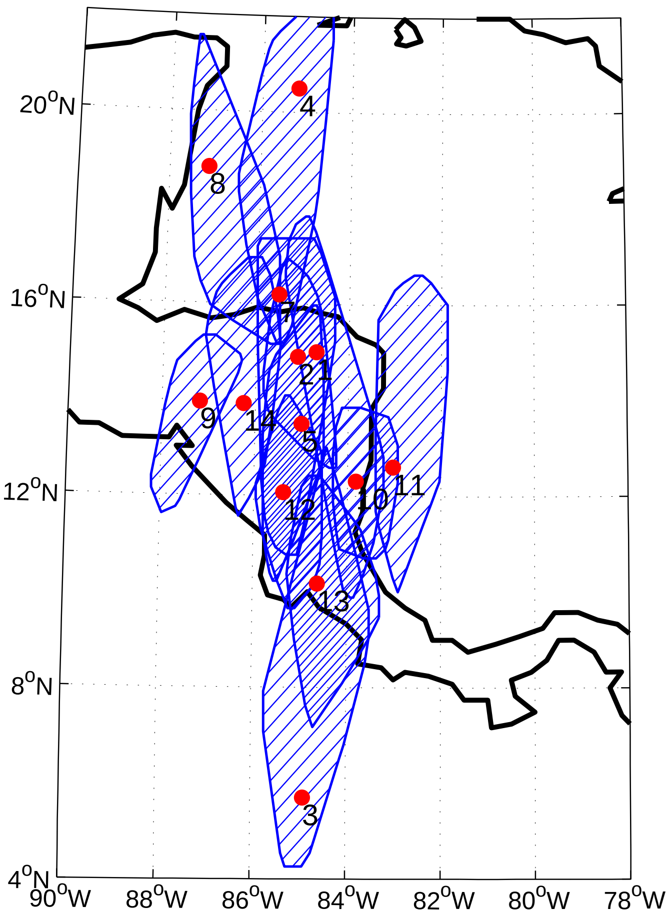

Figure 5. Final point estimates along with a shading of the 25% highest density region for each bird.

4. Future Directions

Ideally, the above mode estimates would be accompanied by appropriate confidence regions describing the variability of the estimates. Unfortunately, such calculations are still the subject of ongoing research, and at present, no confidence regions can be calculated in the bivariate case. The bootstrap is a popular statistical method which can overcome the lack of theoretical quantiles in such cases. However, it is well-known

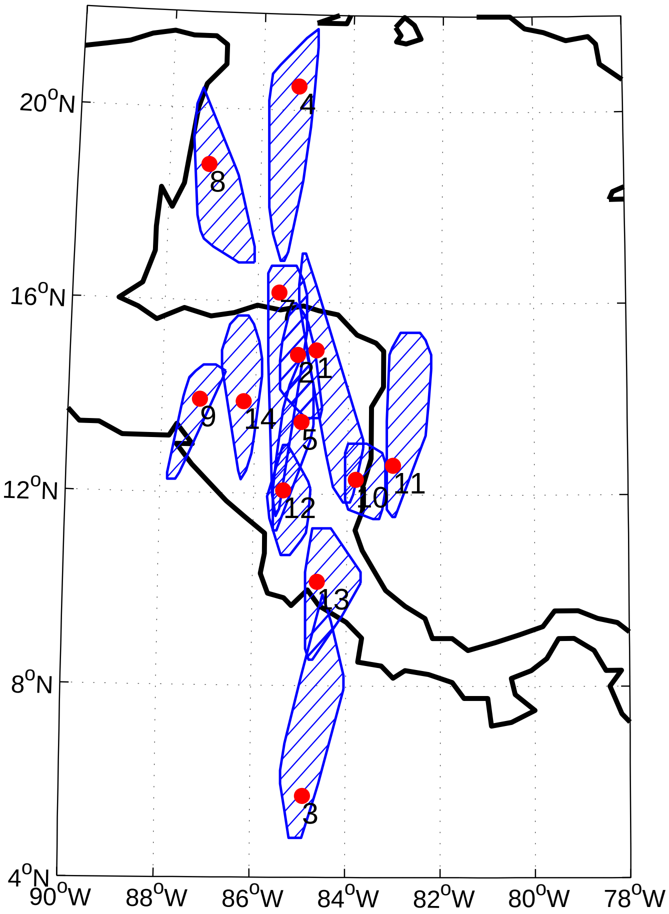

Figure 6. Final point estimates along with a shading of the 10% highest density region for each bird.

that the bootstrap can be troublesome in shape-constrained estimation (Sen et al., 2010), and we therefore abstain from providing any such estimates.

5. Appendix

Figures 7 through 20 show detailed plots for the individual birds. In each figure, the leftmost plot shows the data, along with all four centre estimates and the 25% highest density region from the log-concave density estimation. The three rightmost plots show the bagplot and density estimates as described above for wood thrush 10.

References

Cule, M., Gramacy, R. and Samworth, R. (2009). LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density.

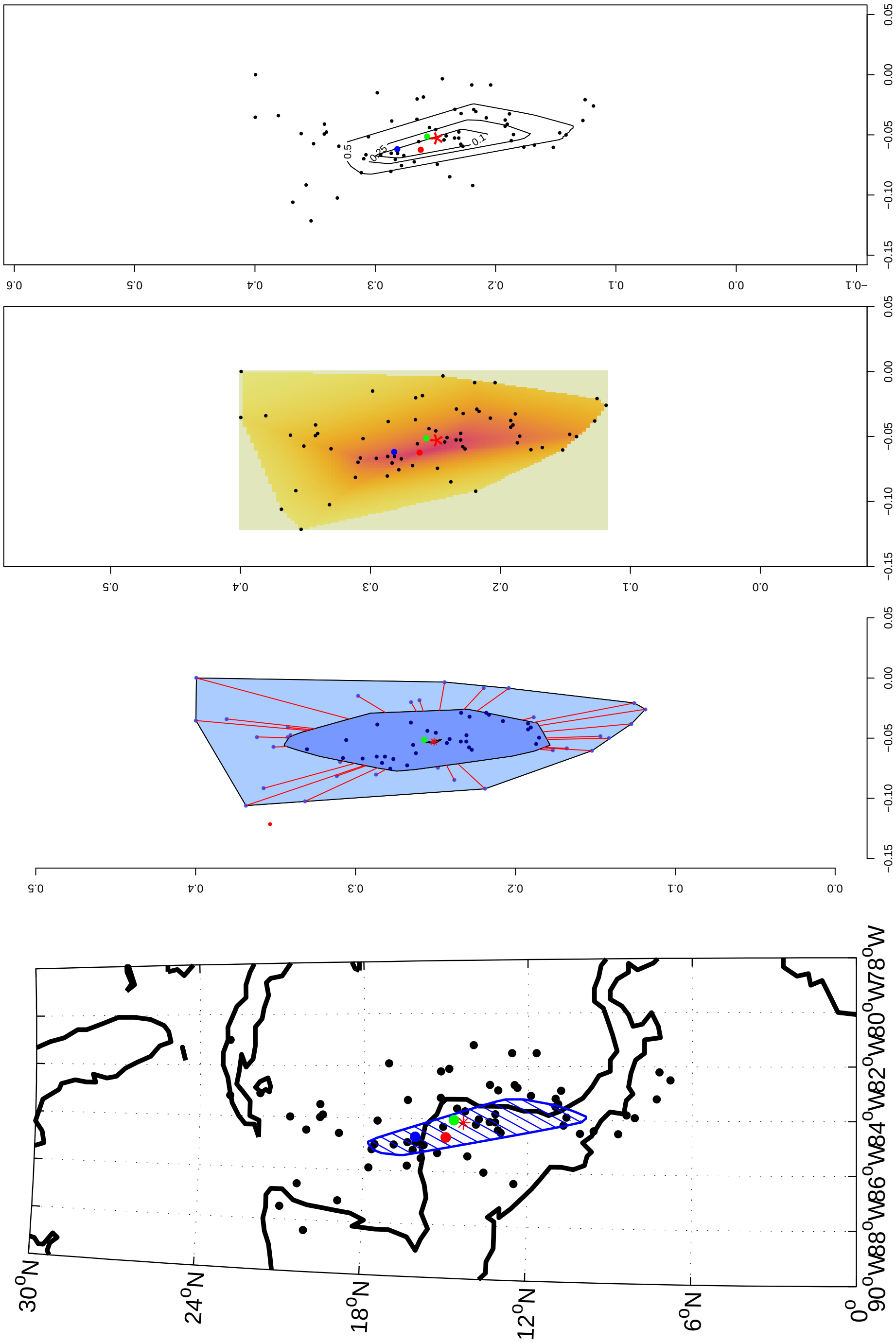

FIGURE 7. Wood thrush 1 (n=66).

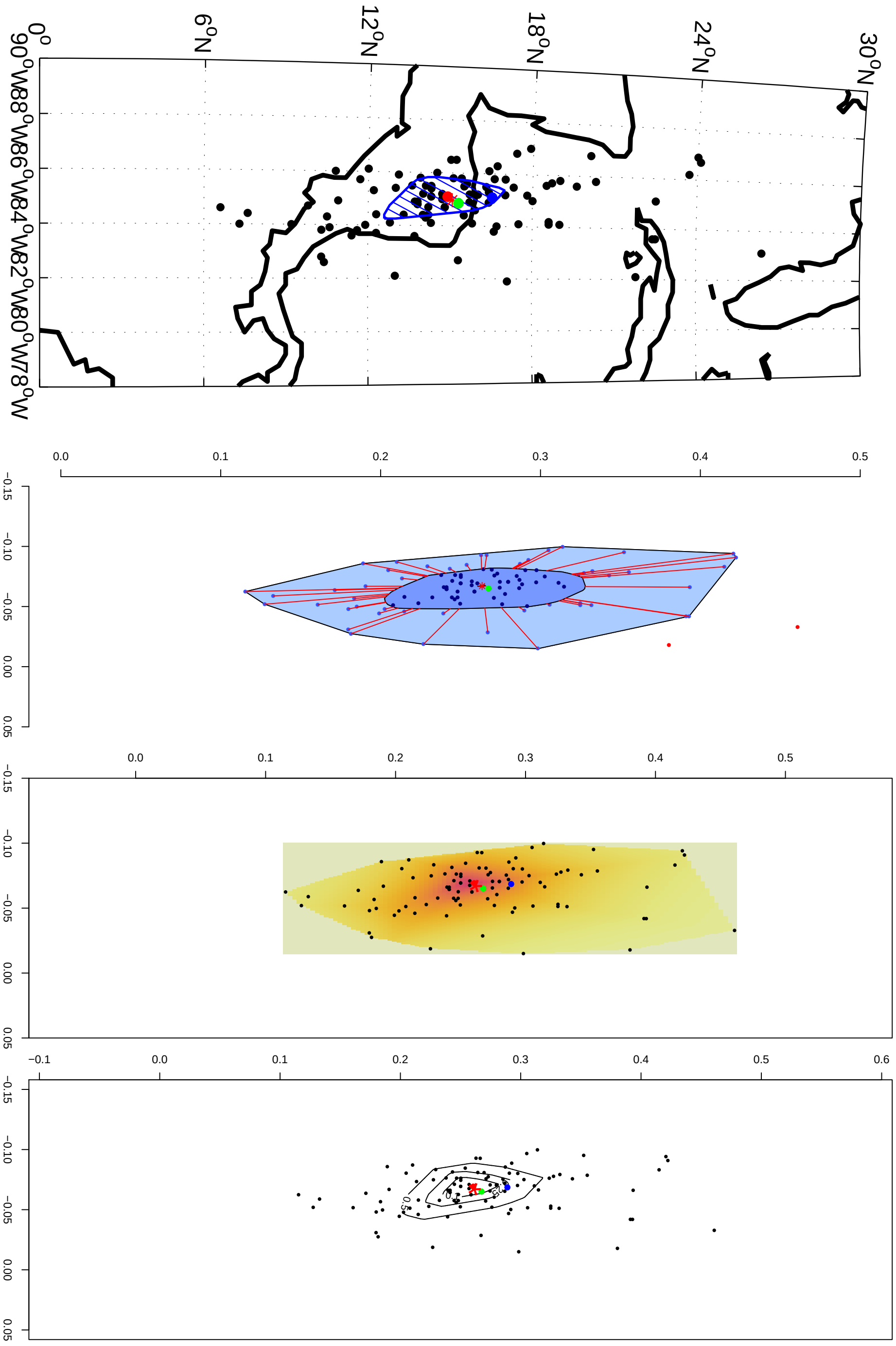

Figure 8. Wood thrush 2 (n=99).

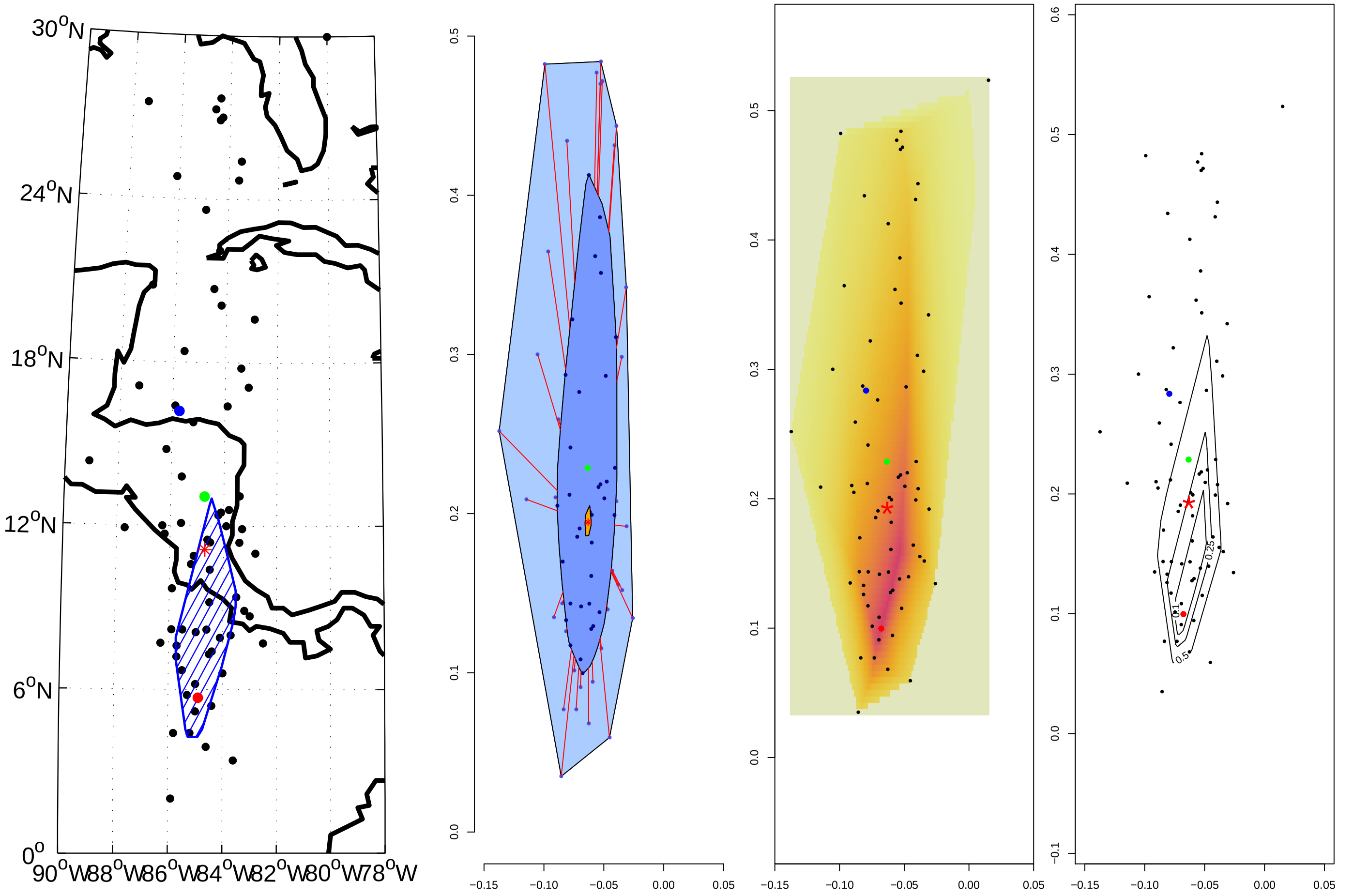

FIGURE 9. Wood thrush 3 (n=71).

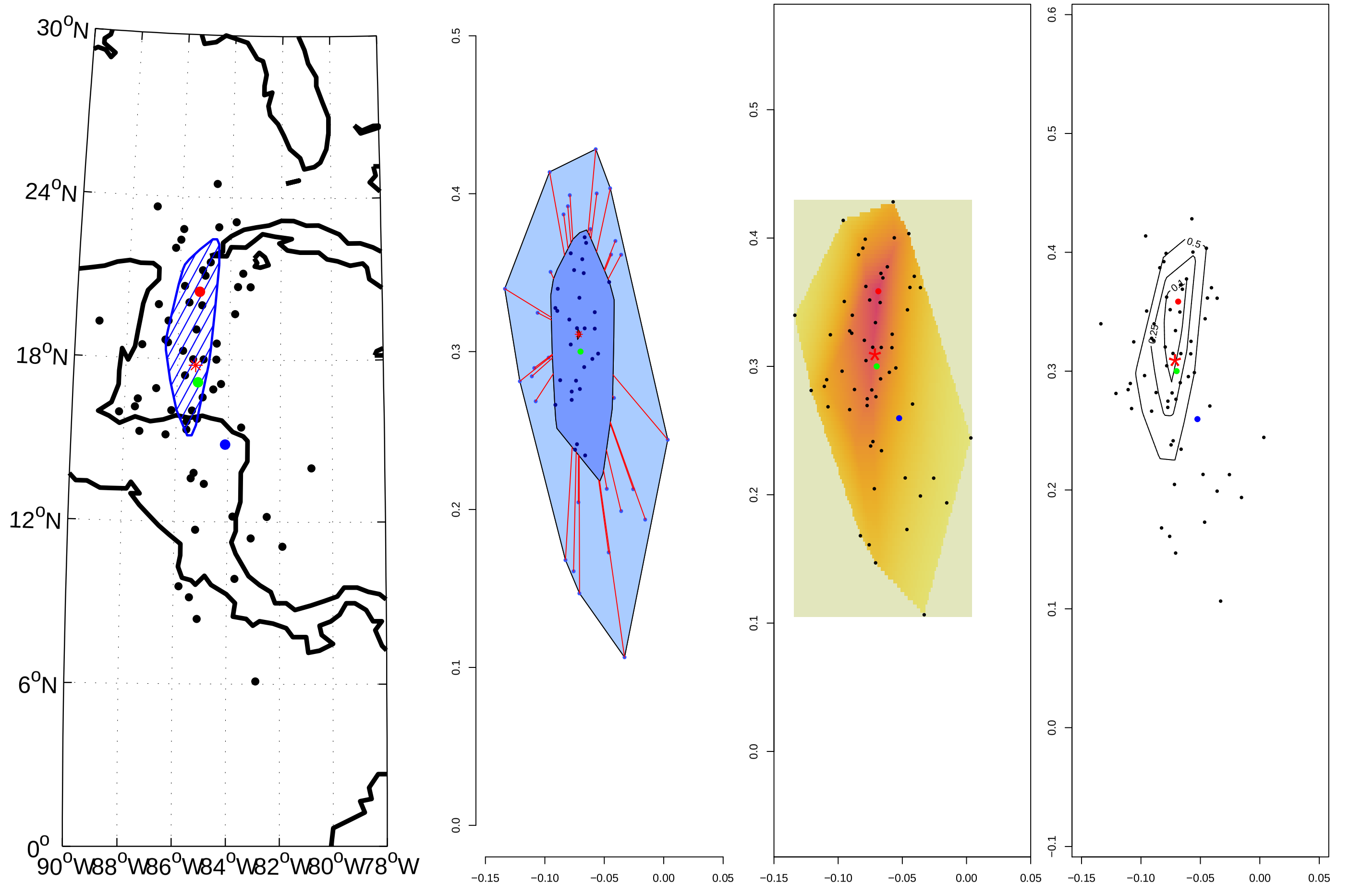

FIGURE 10. Wood thrush 4 (n=60).

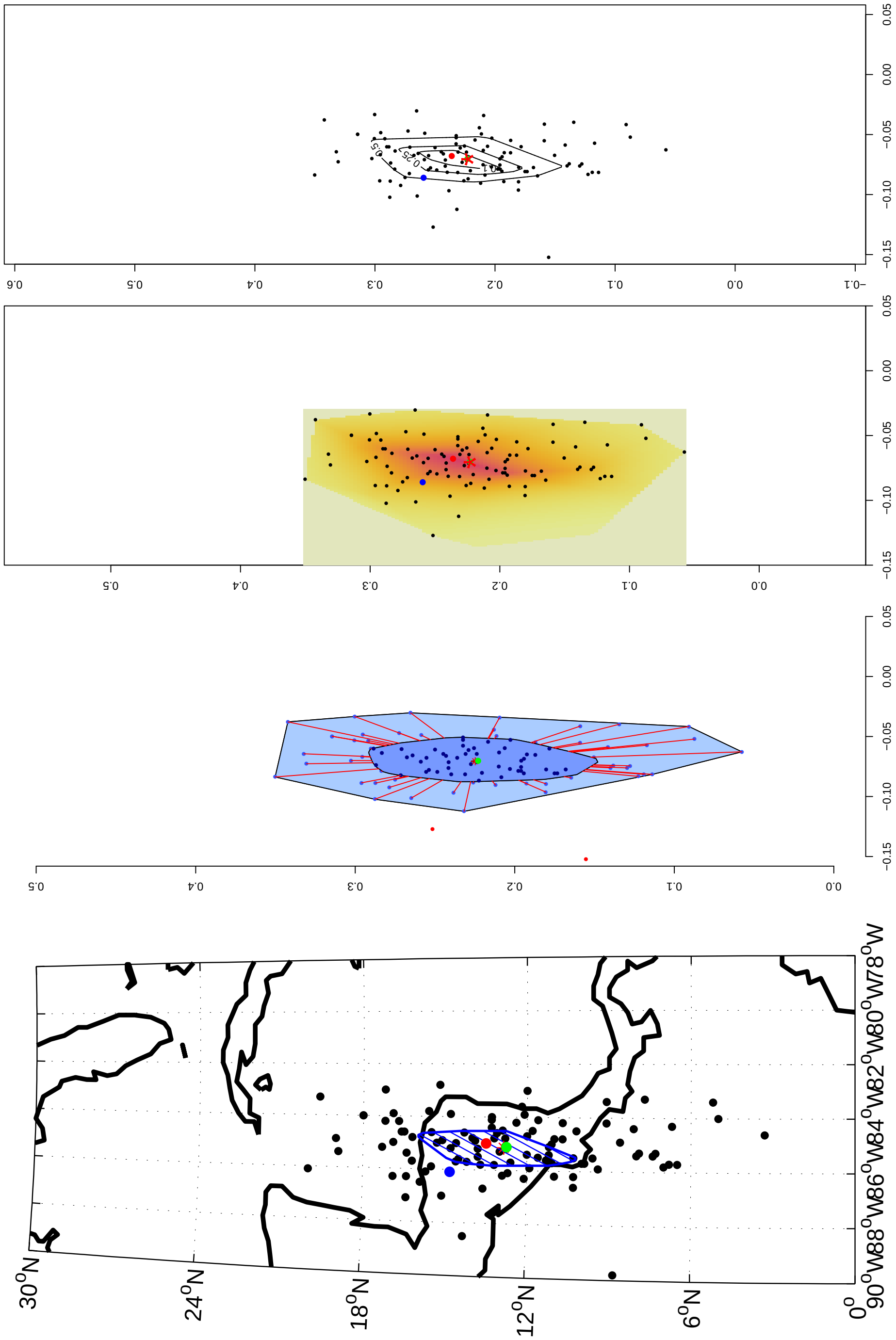

FIGURE 11. Wood thrush 5 (n=105).

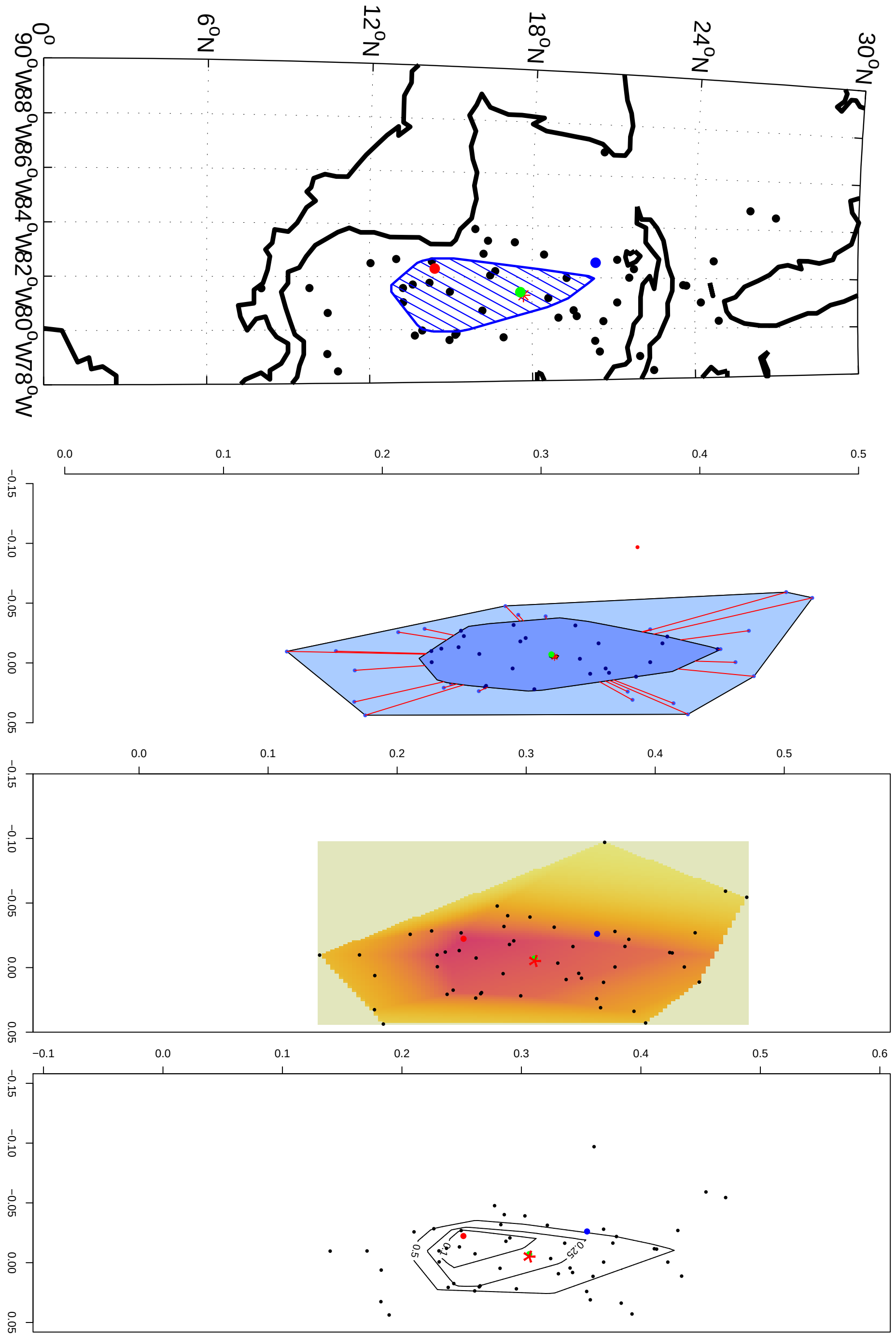

FIGURE 12. Wood thrush 6 (n=50).

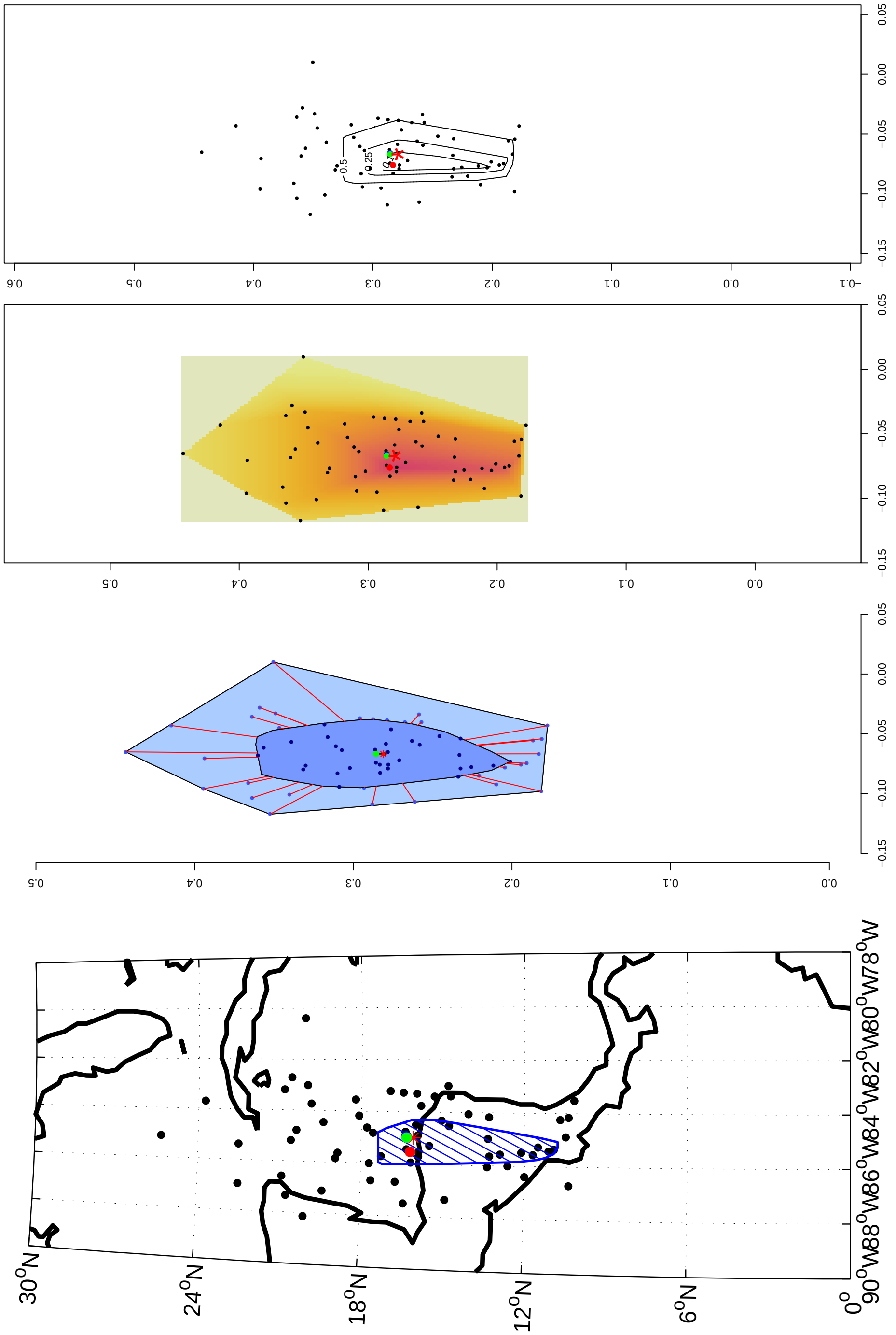

FIGURE 13. Wood thrush 7 (n=64).

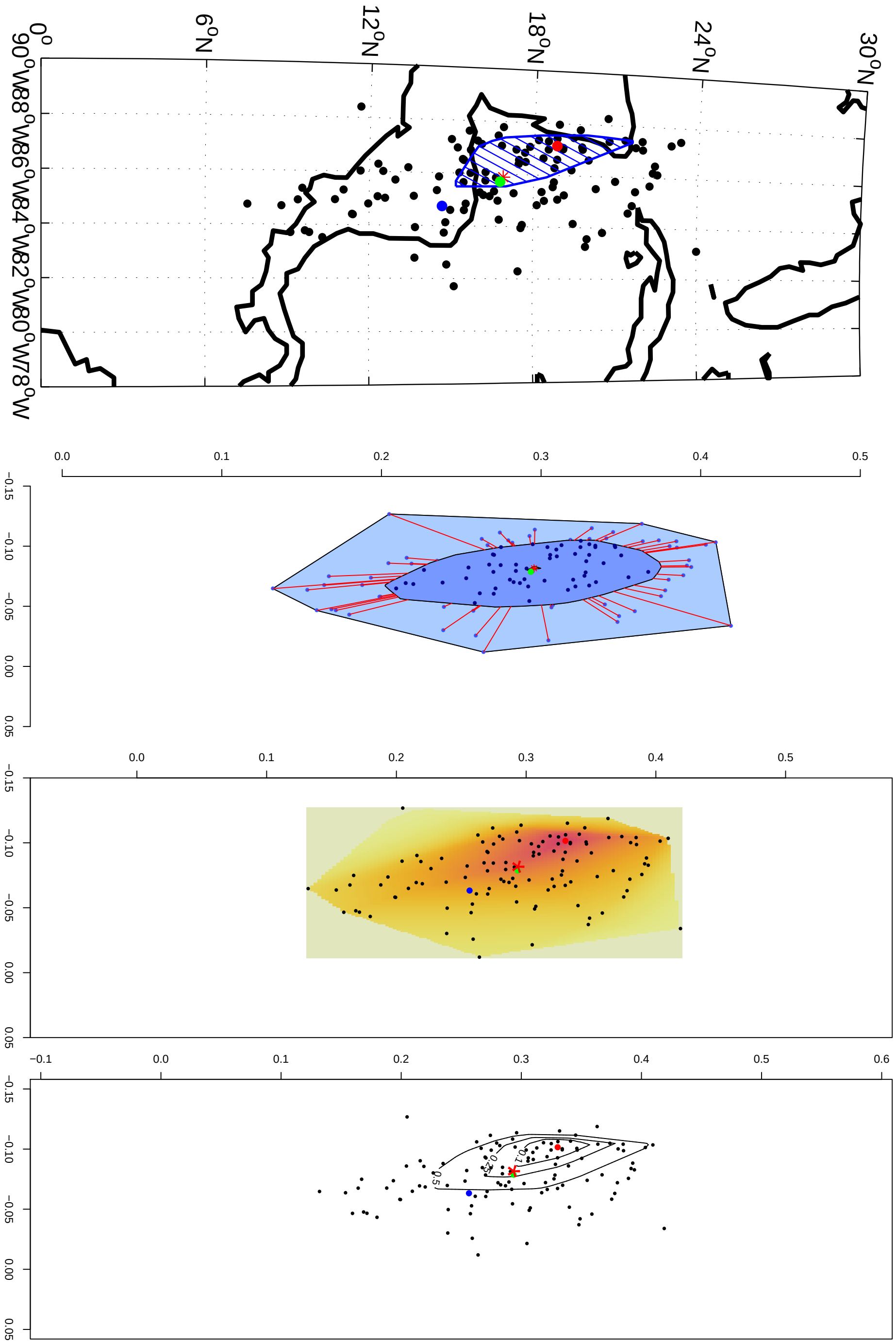

FIGURE 14. Wood thrush 8 (n=113).

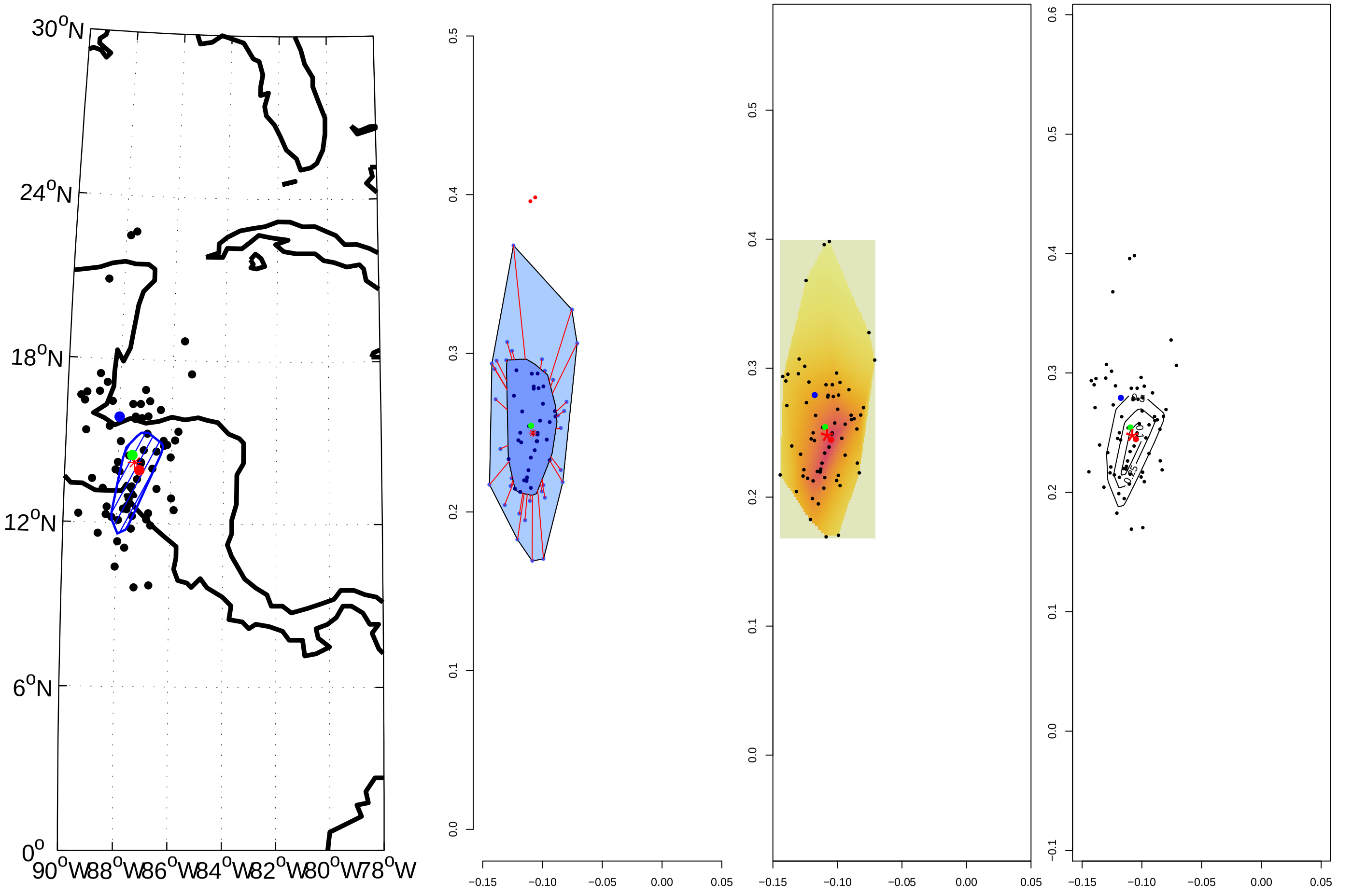

FIGURE 15. Wood thrush 9 (n=70).

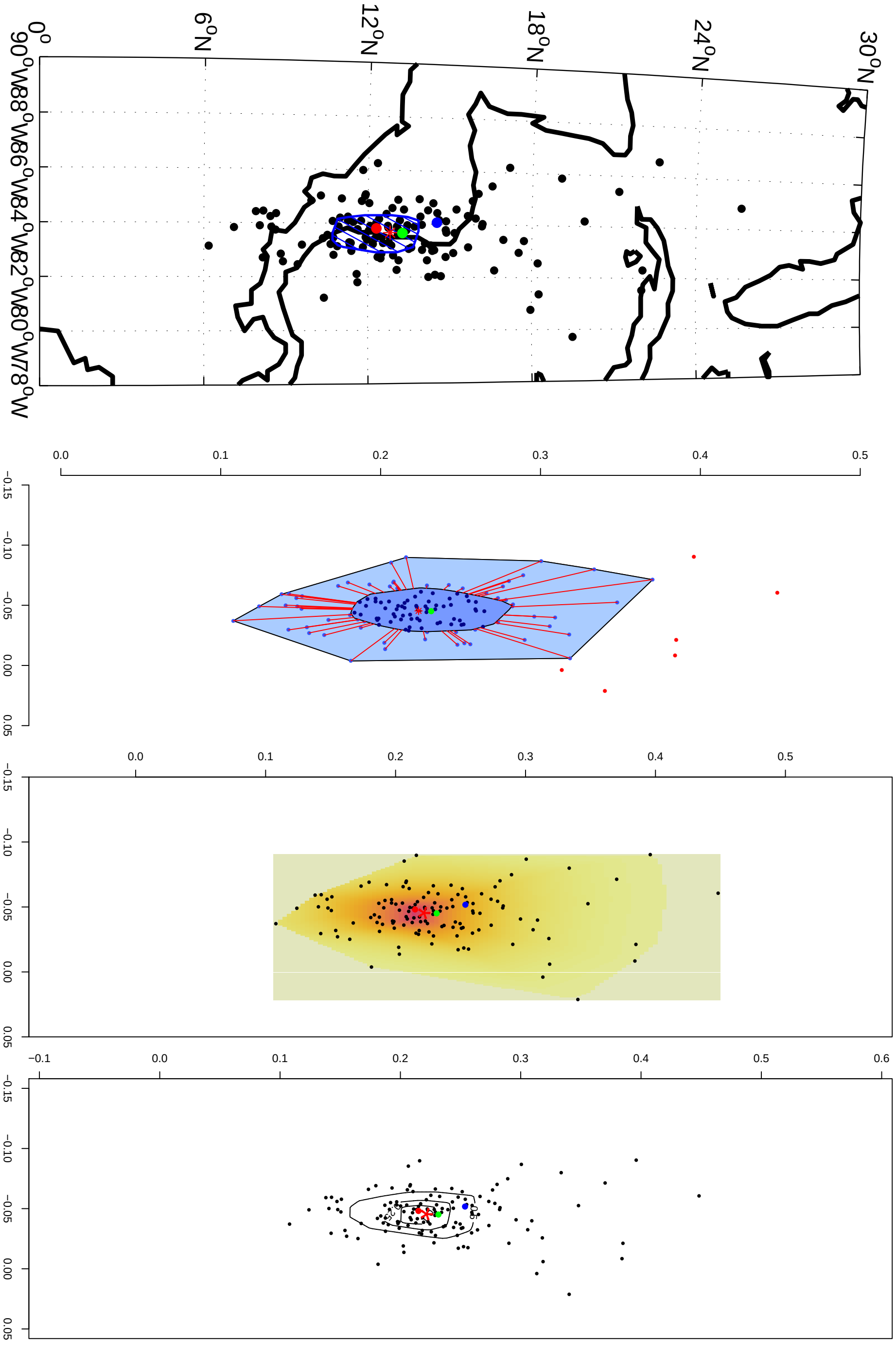

FIGURE 16. Wood thrush 10 (n=127).

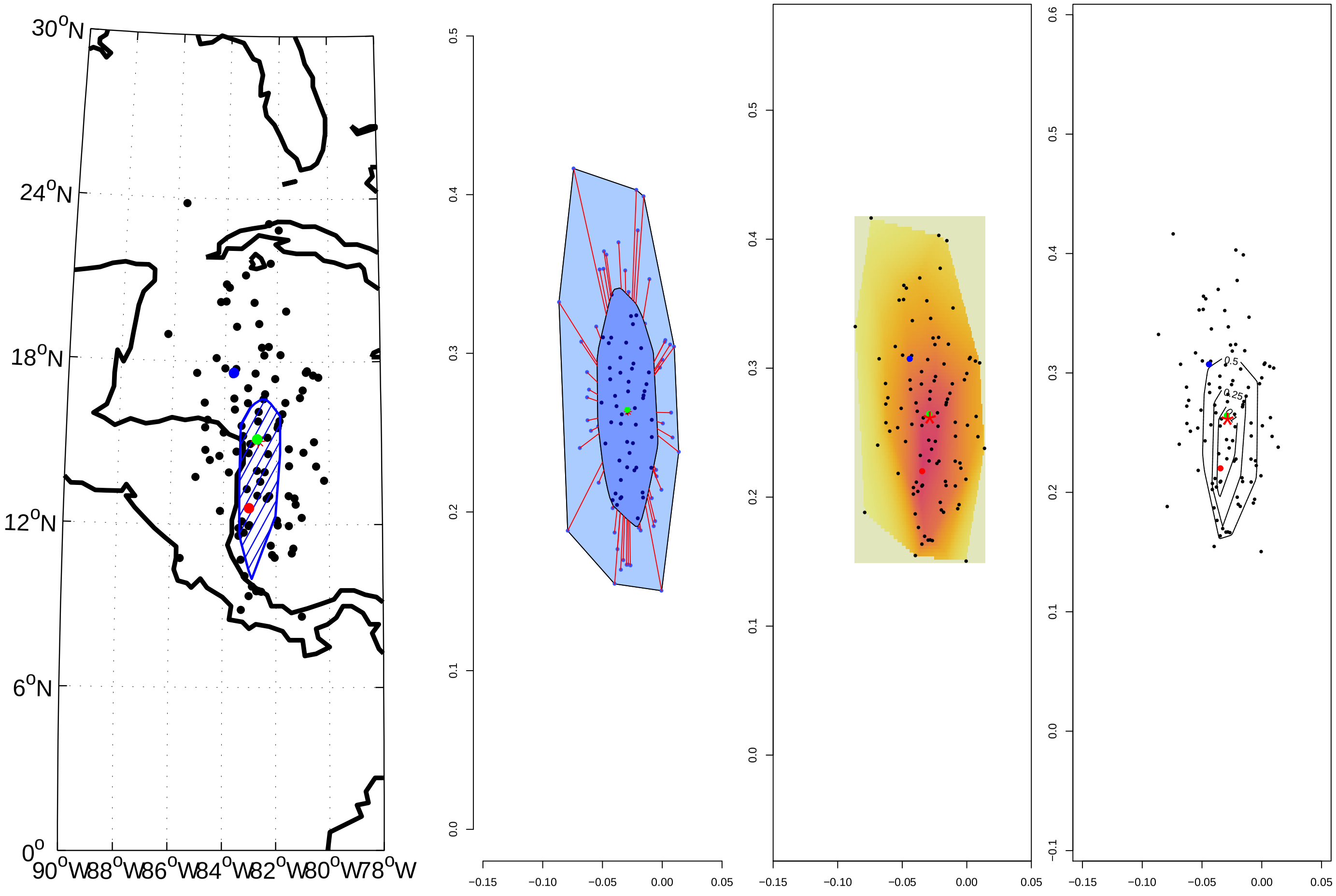

FIGURE 17. Wood thrush 11 (n=103).

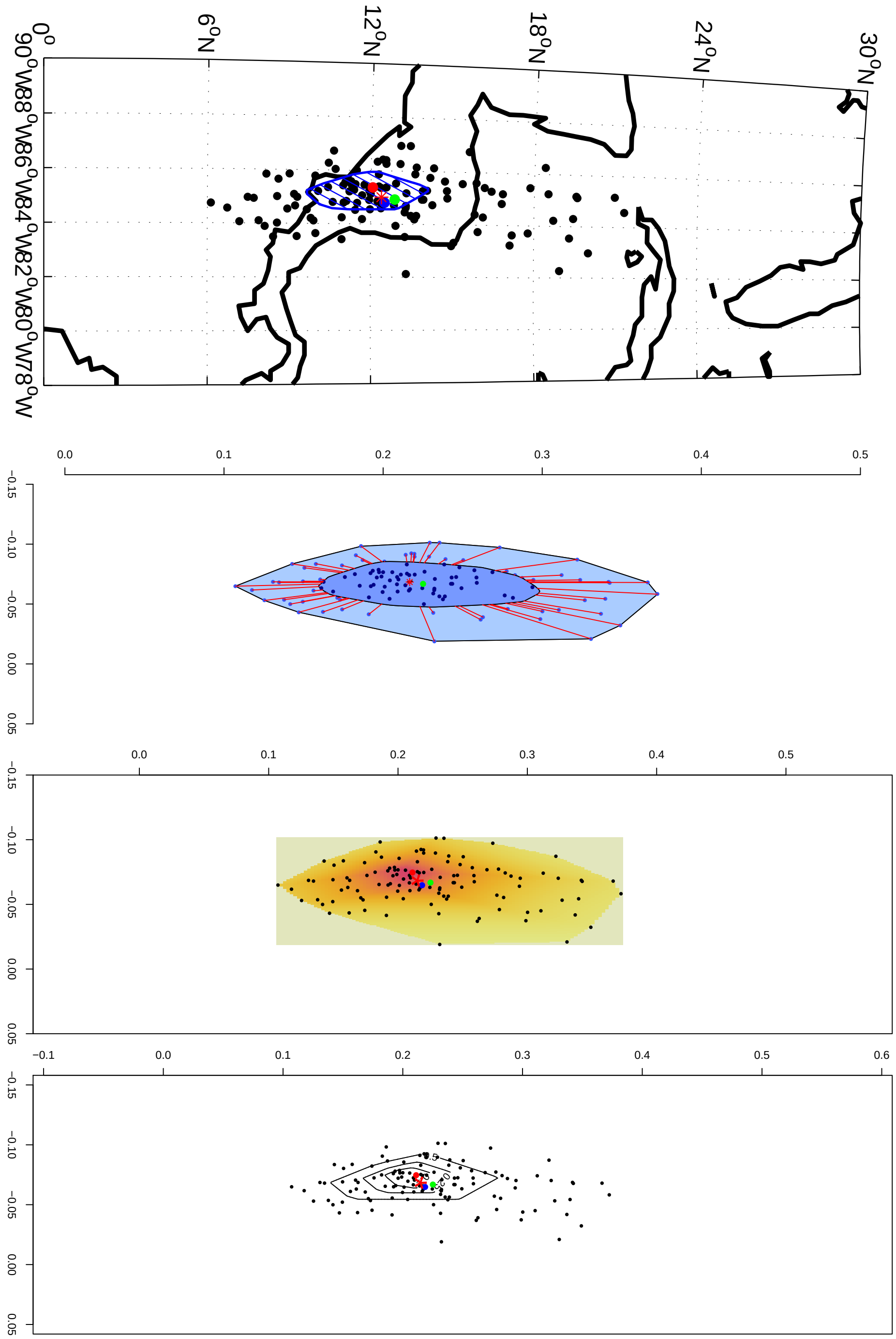

FIGURE 18. Wood thrush 12 (n=116).

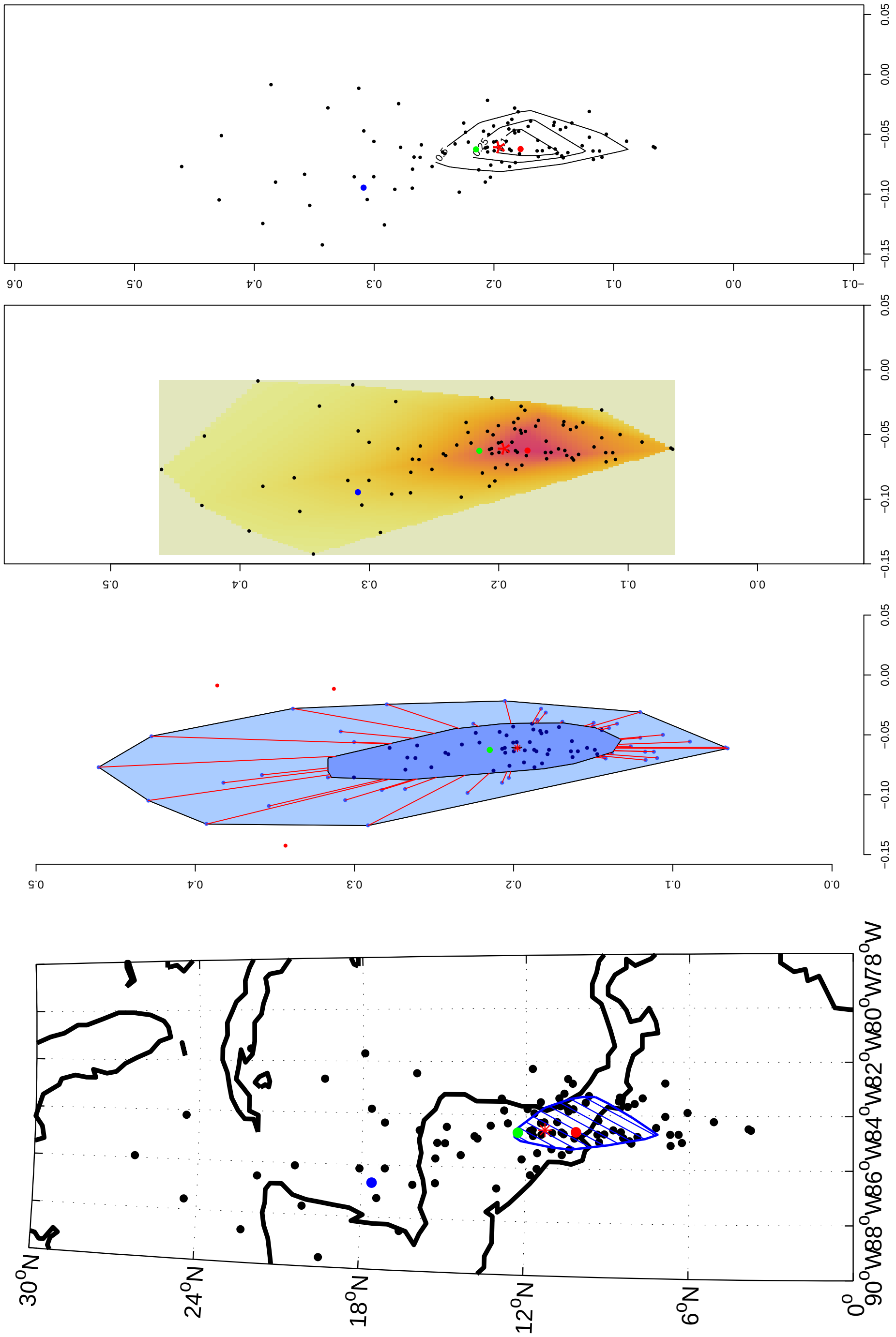

FIGURE 19. Wood thrush 13 (n=92).

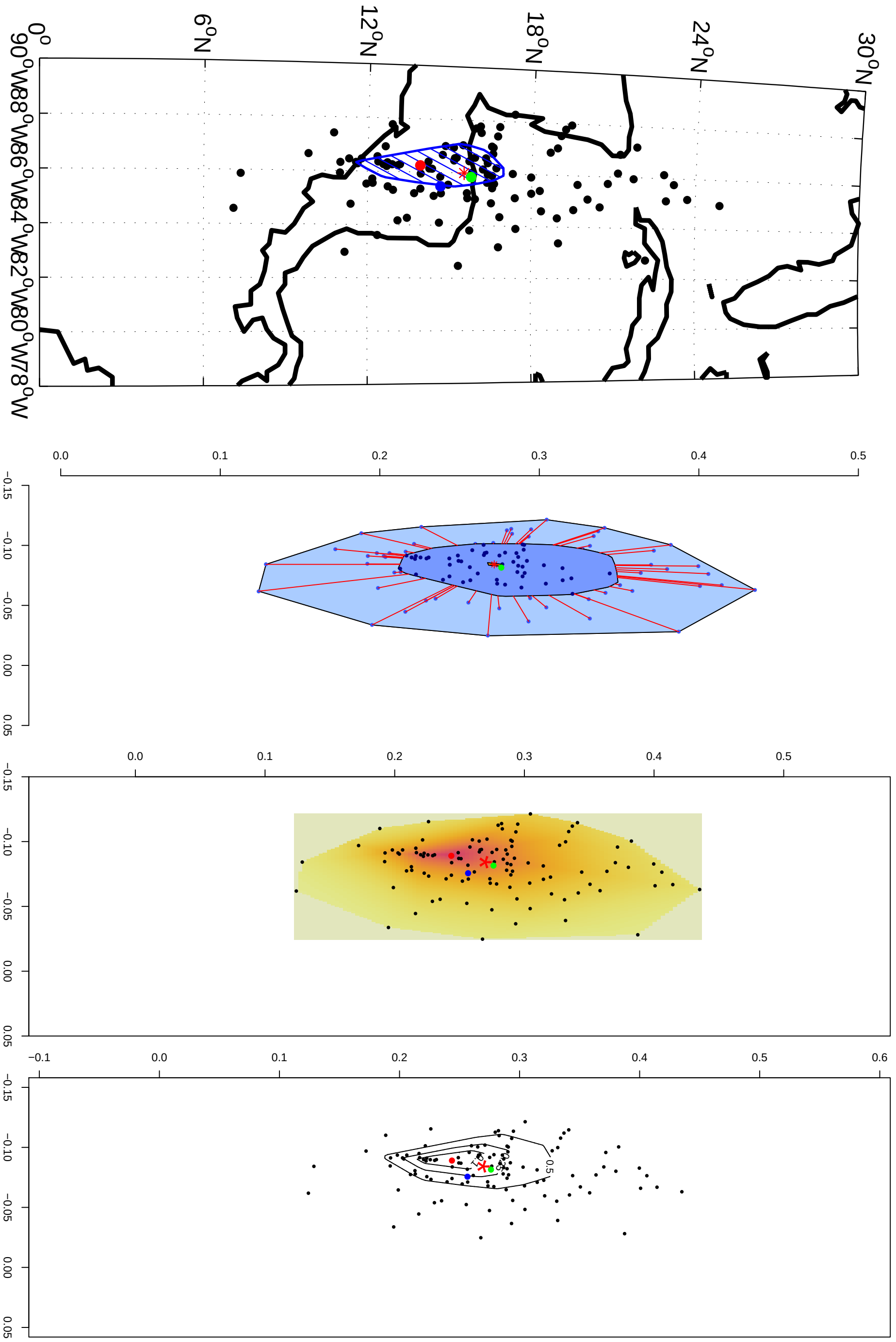

FIGURE 20. Wood thrush 14 (n=107).

Journal of Statistical Software **29**.

Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic J. Stat.* **4** 254–270.

Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607.

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. URL `http://dx.doi.org/10.1214/aos/1176348890`

Pawlowich, R. (2005). M_map - a mapping package for Matlab. Available online: www.eos.ubc.ca/ rich/map.html.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL `http://www.R-project.org`

Roth, R. R., Johnson, M. and Underwood, T. (2011). Wood thrush (hylocichla mustelina). *The Birds of North America (online)* **246**. Originally published 1996, revised in 2011 by Evans, Melissa, and Elizabeth Gow.

Rousseeuw, P., Ruts, I. and Tukey, J. (1999). The bagplot: a bivariate boxplot. *The American Statistician* **53**.

Sen, B., Banerjee, M. and Woodroofe, M. (2010). Inconsistency of bootstrap: the Grenander estimator. *The Annals of Statistics* 1953–1977.

Stutchberry, B. (2009). Personal communication.

Stutchbury, B. J. M., Tarof, S. A., Done, T., Gow, E., Kramer, P. M., Tautin, J., Fox, J. W. and Afanasyev, V. (2009). Tracking long-distance songbird migration by using geolocators. *Science* **323** 896.

The Globe and Mail (2009). Tiny, fast - and they sing too. February 13, 2009.

Wolf, P. (2010). *aplpack: Another PLot PACKage*.
URL `http://www.R-project.org`

Department of Mathematics and Statistics, York University
e-mail: `hkj@mathstat.yorku.ca`