KERNEL DENSITY ESTIMATION FOR SPATIAL ECOLOGY

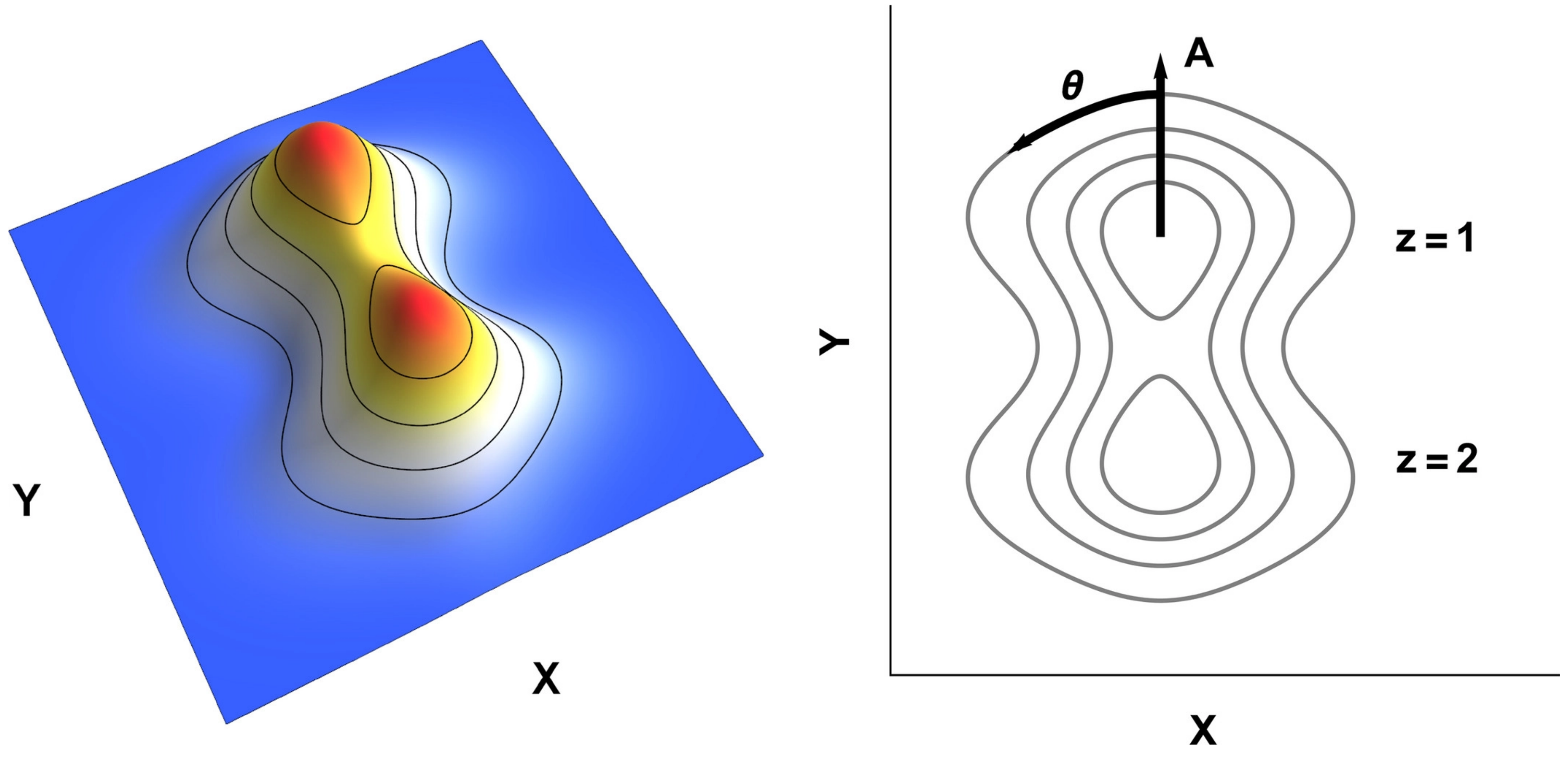
Josh Cullen

September 9, 2022

Photo by Colton Jones on Unsplash

What is kernel density estimation (KDE)?

- **Horne et al. (2020)**: "Kernel density estimation (KDE) has been used across scientific disciplines as a nonparametric approach to estimate an unknown probability distribution (Silverman 1986). KDE works by placing small kernels or bumps over each data point and then averaging the contribution from all kernels to obtain an estimate of the probability distribution at any point in space"
- **Worton (1989)**: "…kernel methods free the UD estimate from parametric assumptions and provide a means of smoothing locational data to make more efficient use of them than a histogram."
- **Seaman and Powell (1996)**: "In the context of home range analysis, the density at any location is an estimate of the amount of time spent there. This information forms a basis for ecological investigations of habitat use and preference. The density also forms a basis for measuring the overlap of individuals or species in terms of area and intensity of use (volume)."



Fleming and Calabrese 2017

Number of points

$$\hat{f}_h(\mathbf{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left[\frac{\mathbf{x} - X_i}{h}\right]$$

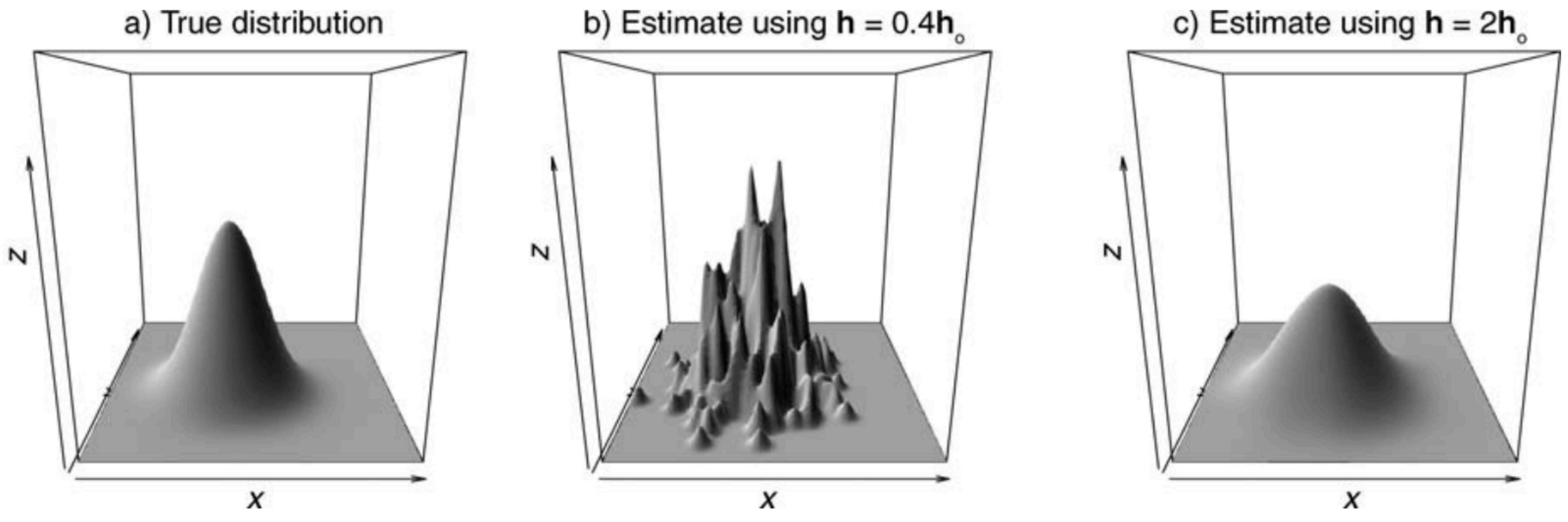
Probability density function evaluated at coordinates of $\mathbf{x}$

Smoothing bandwidth

Bivariate kernel PDF

Caveats regarding KDE

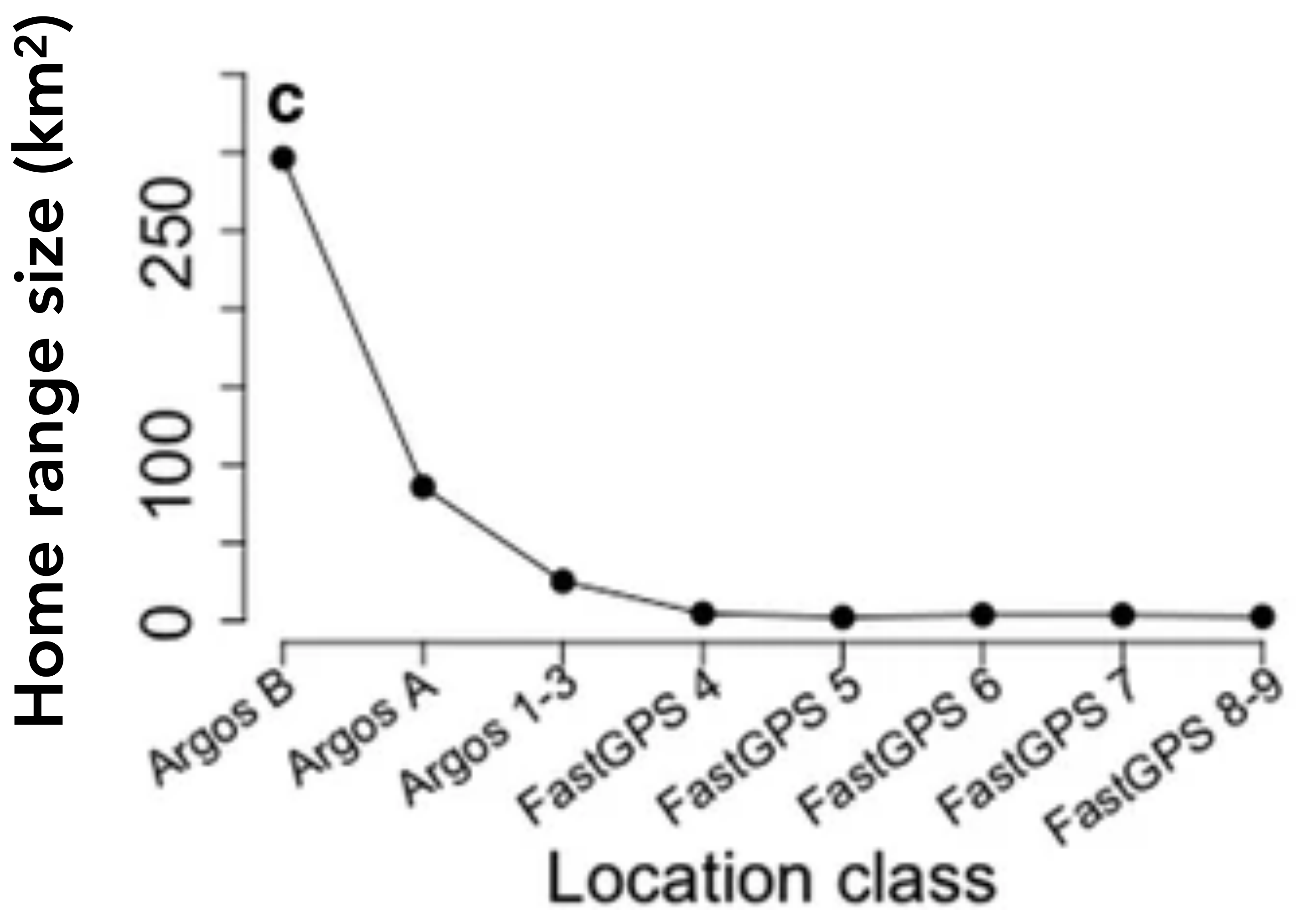
- Choice of bandwidth estimator is of critical importance (Worton 1989; Seaman and Powell, 1996)
  - Controls the smoothing (and therefore intensity) of the density surface
- Method was originally developed for UD estimation for use of VHF telemetry where autocorrelation wasn't a problem (Horne et al. 2020)
  - All points considered independent and identically distributed (IID); not the case with modern telemetry devices that transmit frequently

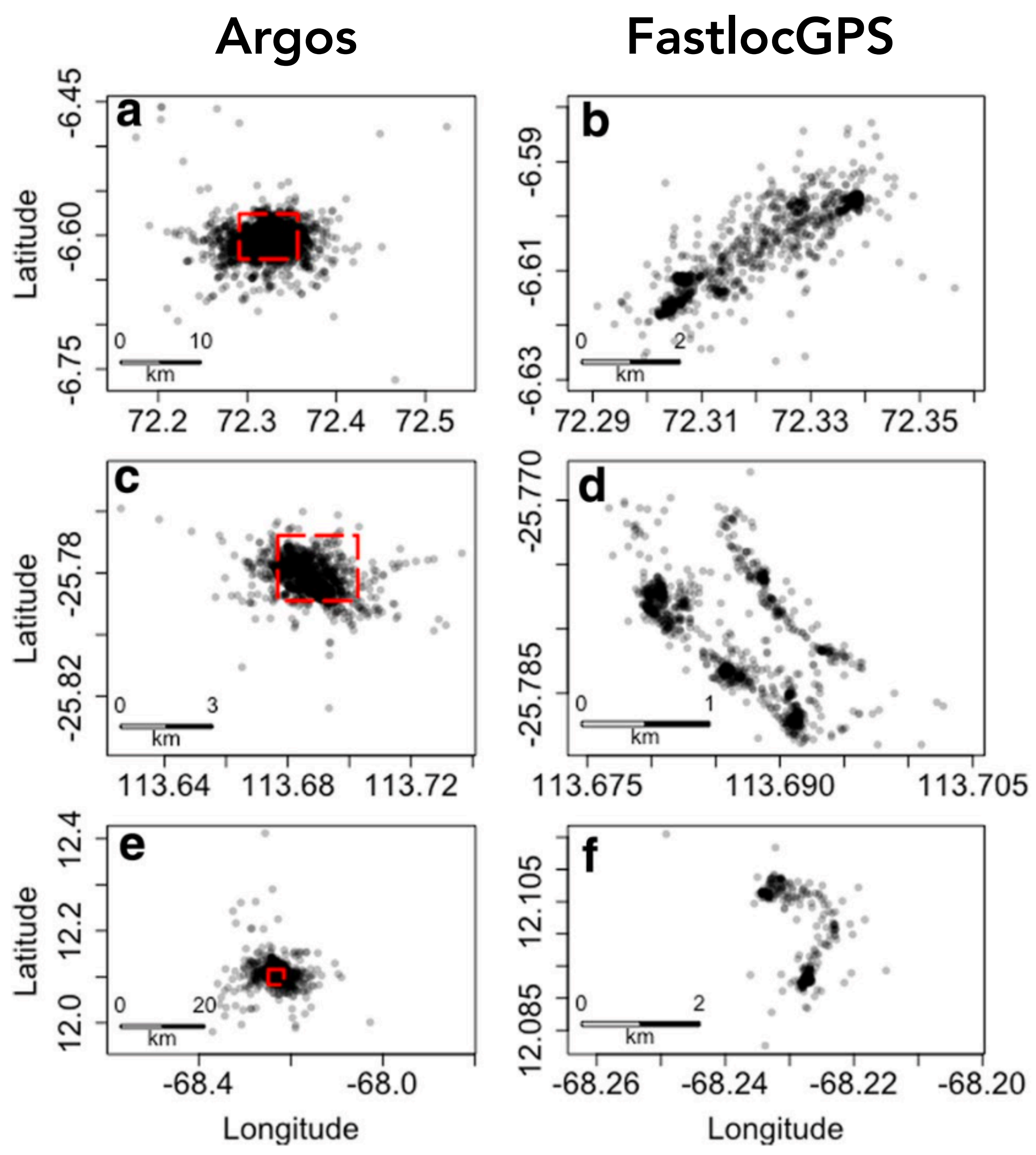


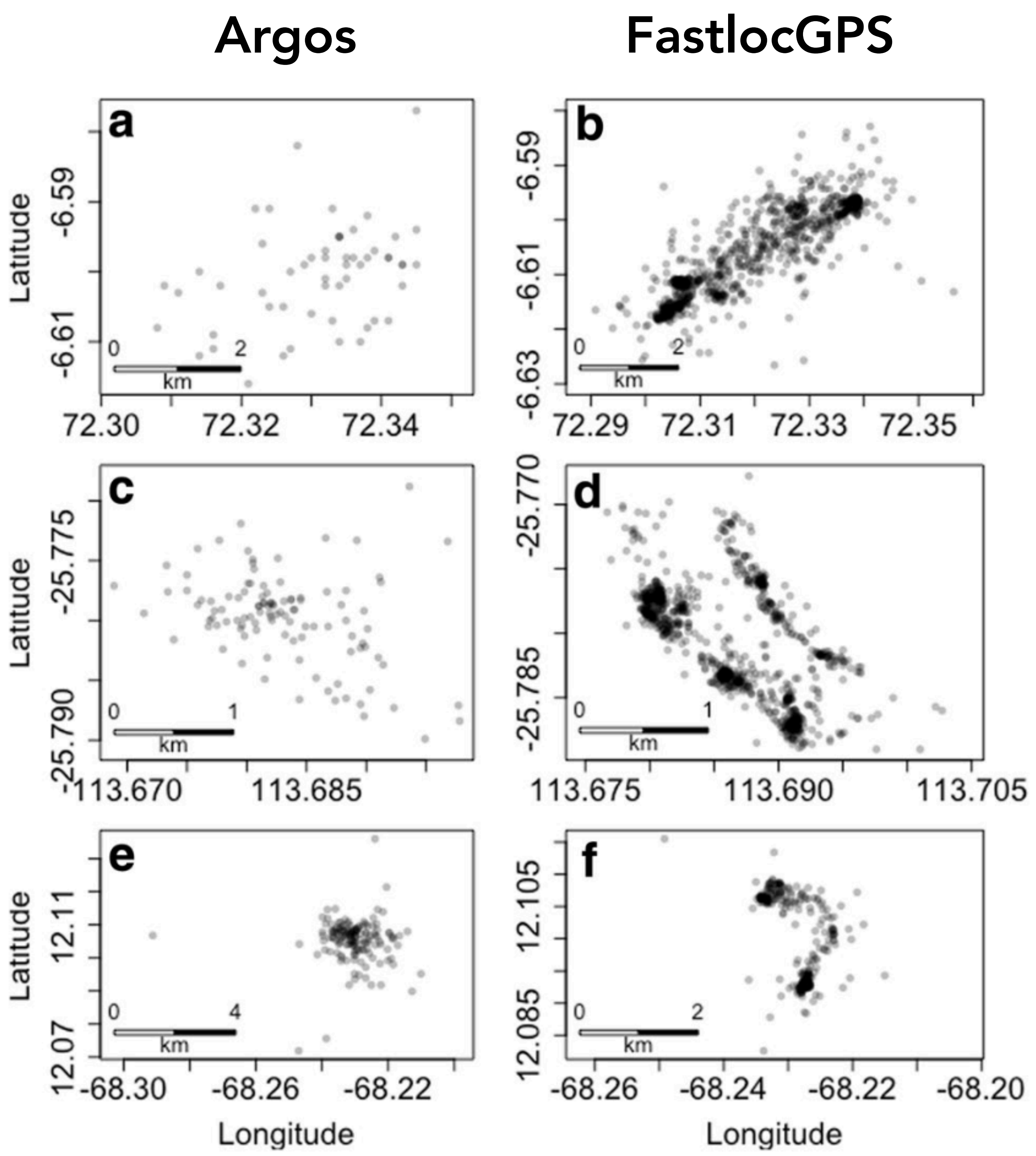
Fieberg 2007

Caveats regarding KDE

- Will be highly impacted by location error (as with most traditional methods)
  - Large Argos errors will result in large UDs, whereas more accurate FastlocGPS will result in considerably smaller UDs for the same movement path



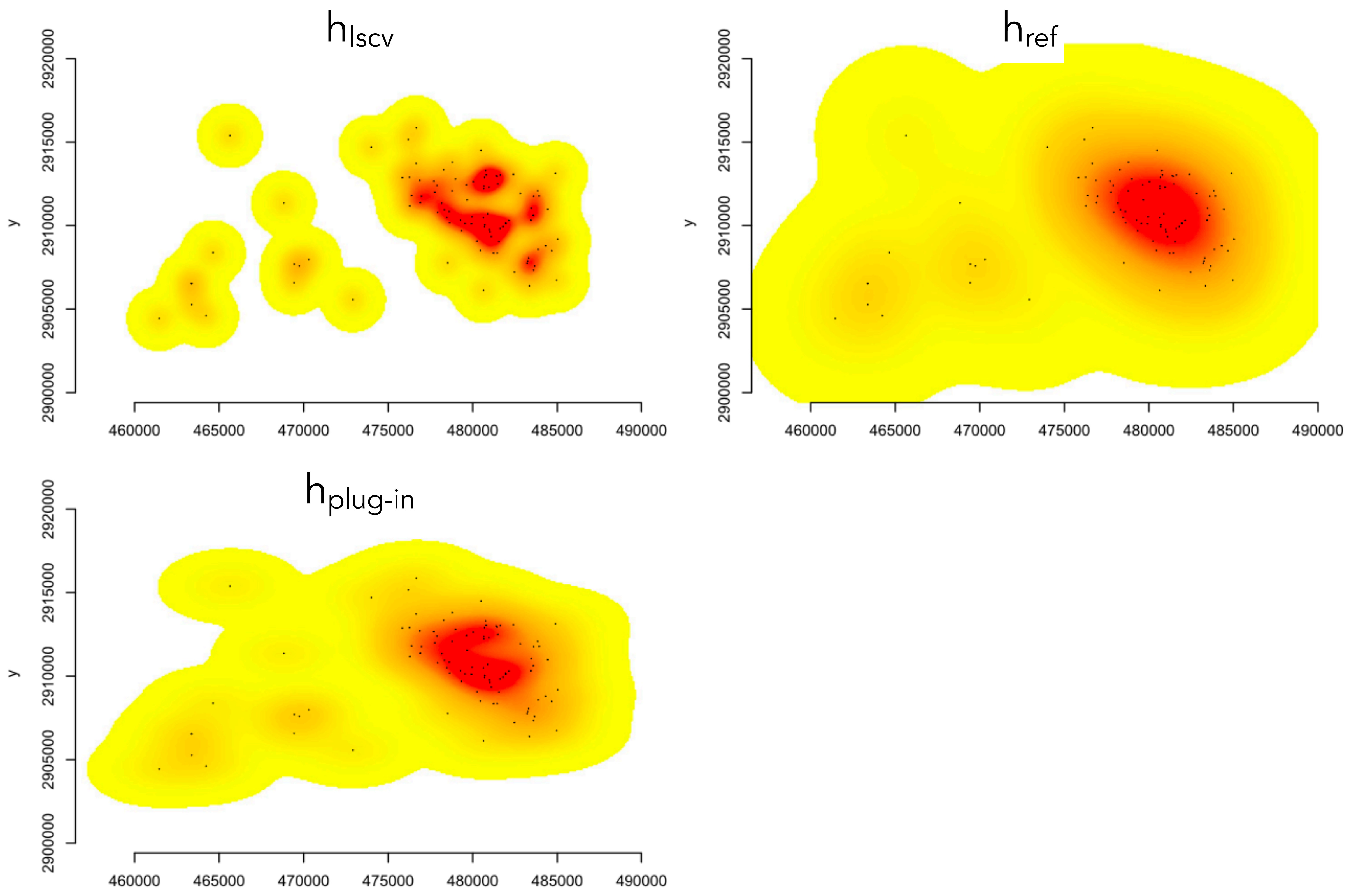




Thomson et al 2017

Types of bandwidth estimation

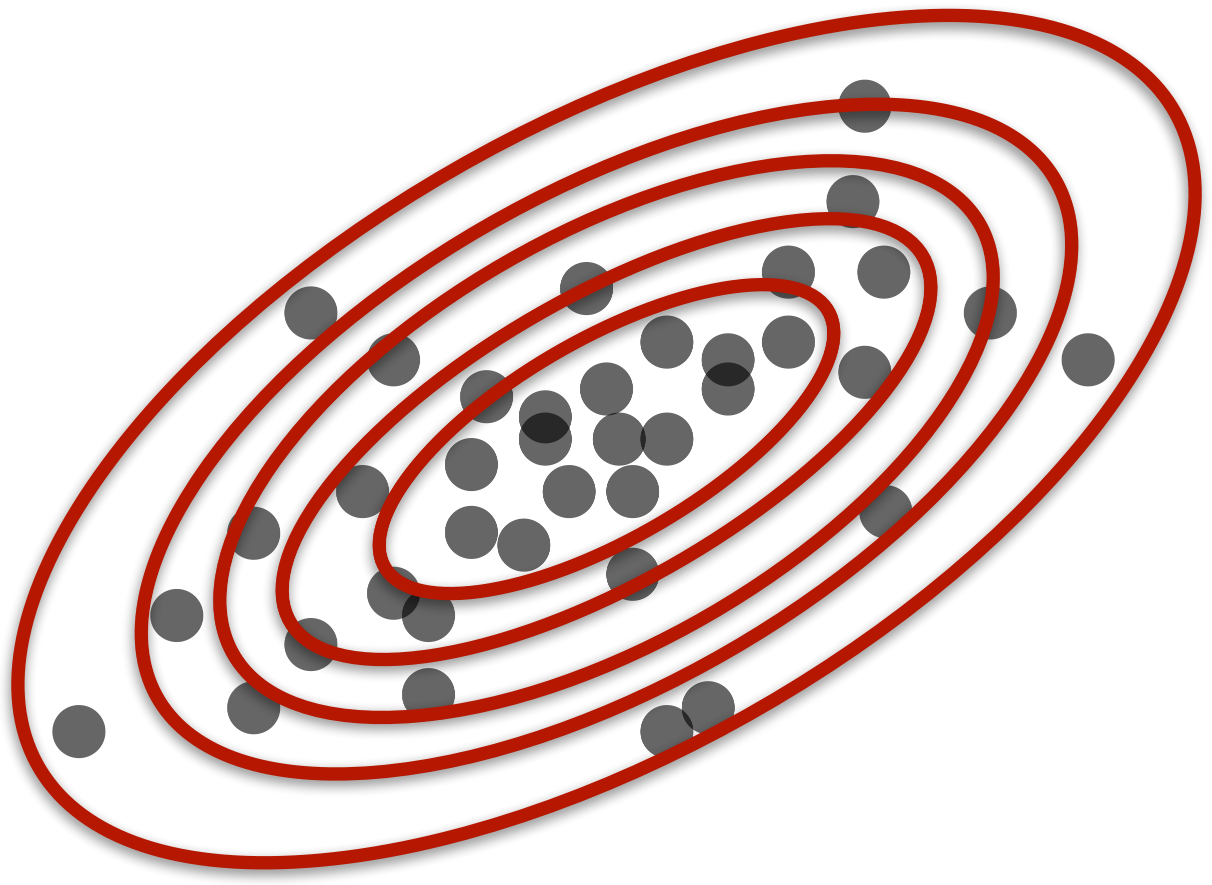
- Reference (h_{ref})
- Least-squares cross validation (h_{lscv})
- Plug-in ($h_{plug\text{-}in}$)



Walter et al. 2015

Reference method (h_{ref})

- Best-suited to a unimodal distribution
- Derived from a bivariate normal distribution by minimizing the mean integrated square error (MISE)
- Typically estimates single bandwidth in both coordinate dimensions
- Since most tracks exhibit a multimodal distribution, this violates the unimodal assumption and overestimates the UD (Seaman and Powell, 1996)



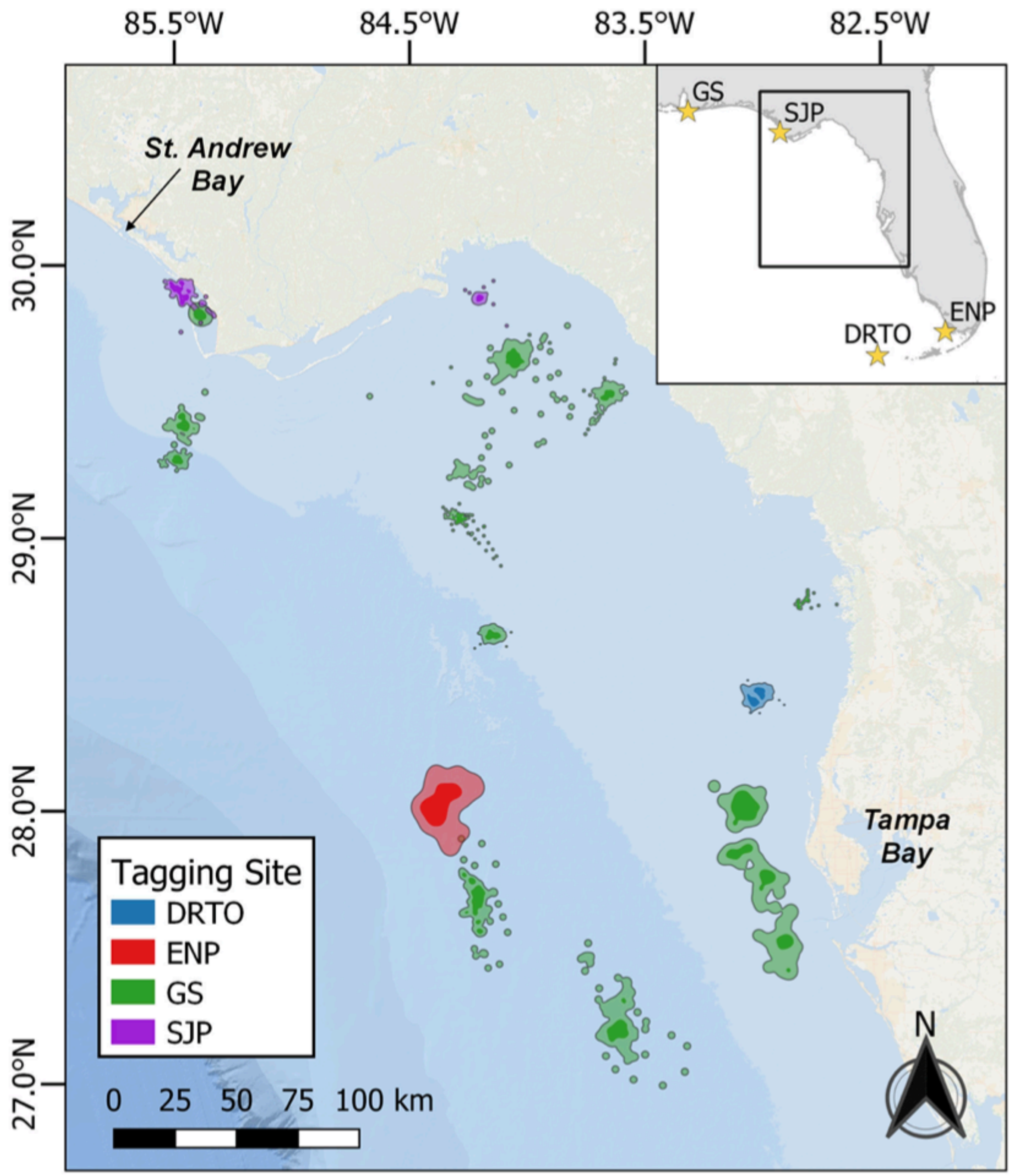
$$h_{ref} = \sqrt{\sigma^2} \times n^{-1/6}$$

Avg marginal covariance from x and y coordinates

Sample size

Least-squares cross validation method (h_{lscv})

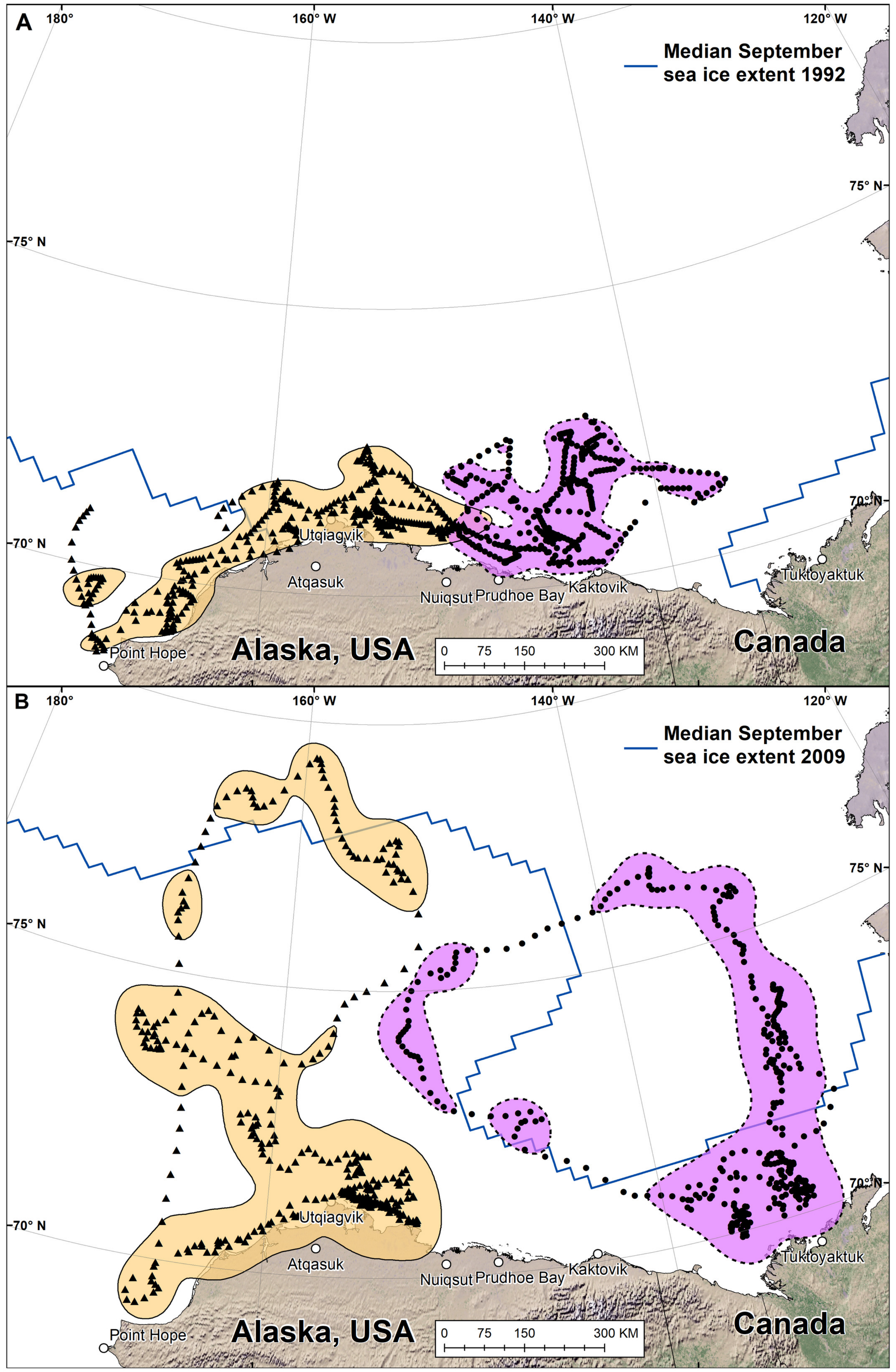
- Examines a range of potential bandwidths and finds the one that minimizes error between true and estimated distributions
- Can estimate separate bandwidths in x and y dimensions
- As time interval of tracks decreases, so will bandwidth estimates (Horne et al. 2020)
  - UD estimates will tightly cover tracks
- Provides better estimates for tracks w/ multiple tightly clustered groups of points (Gitzen et al. 2006)
  - Compared to other bandwidth estimators



Hart et al. 2020

Plug-in method ($h_{plug-in}$)

- Uses pilot bandwidth values to minimize error function and than "plugs-in" these values into the equation directly to be minimized (Gitzen et al. 2006)
- Can estimate separate bandwidths in x and y dimensions
- May oversmooth UDs if points are truly separated into distinct clusters (Gitzen et al. 2006)
  - Compared to h_{lscv}
- If points only slightly clustered or spread out, this method performs as well or better than h_{lscv}

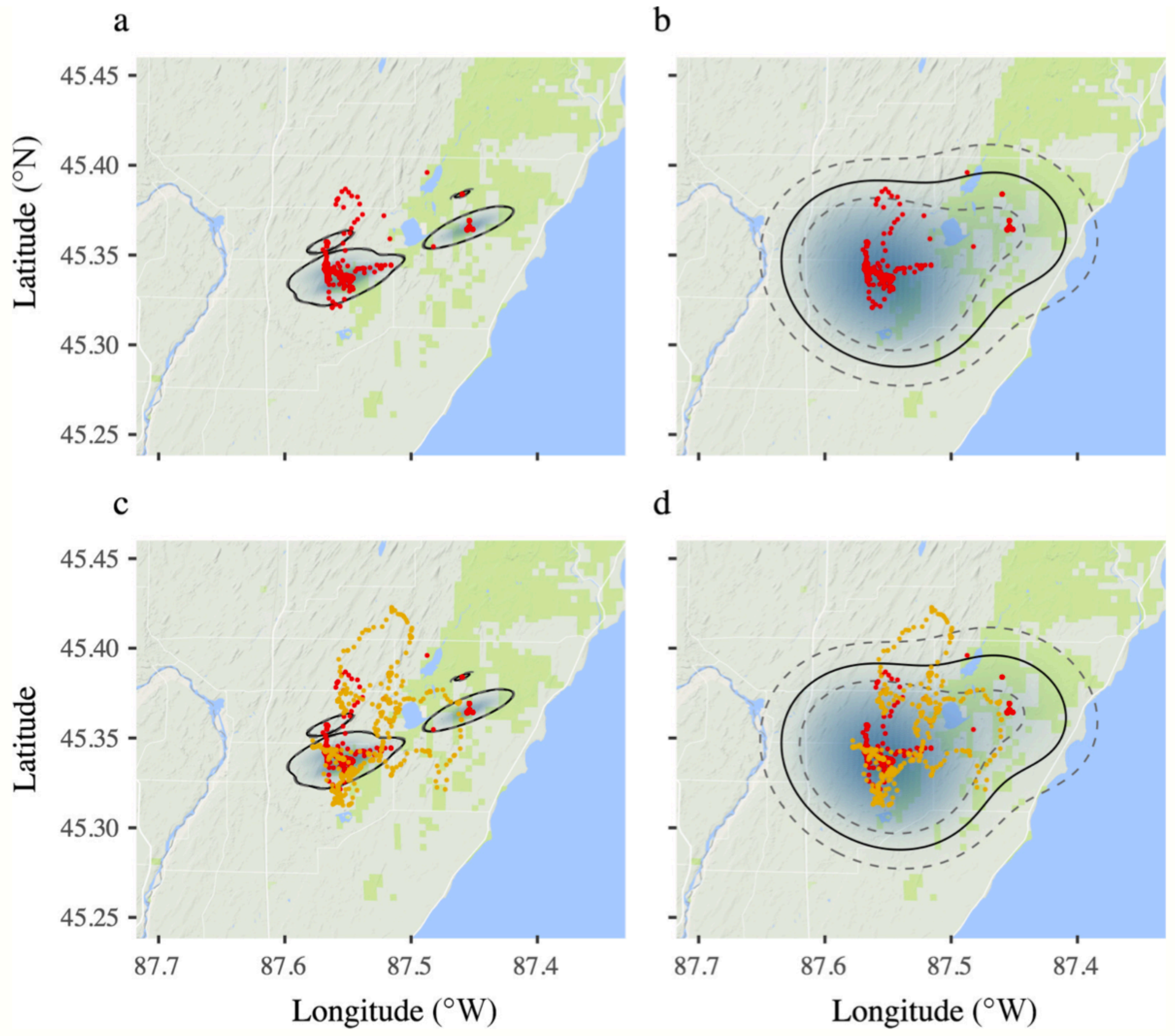


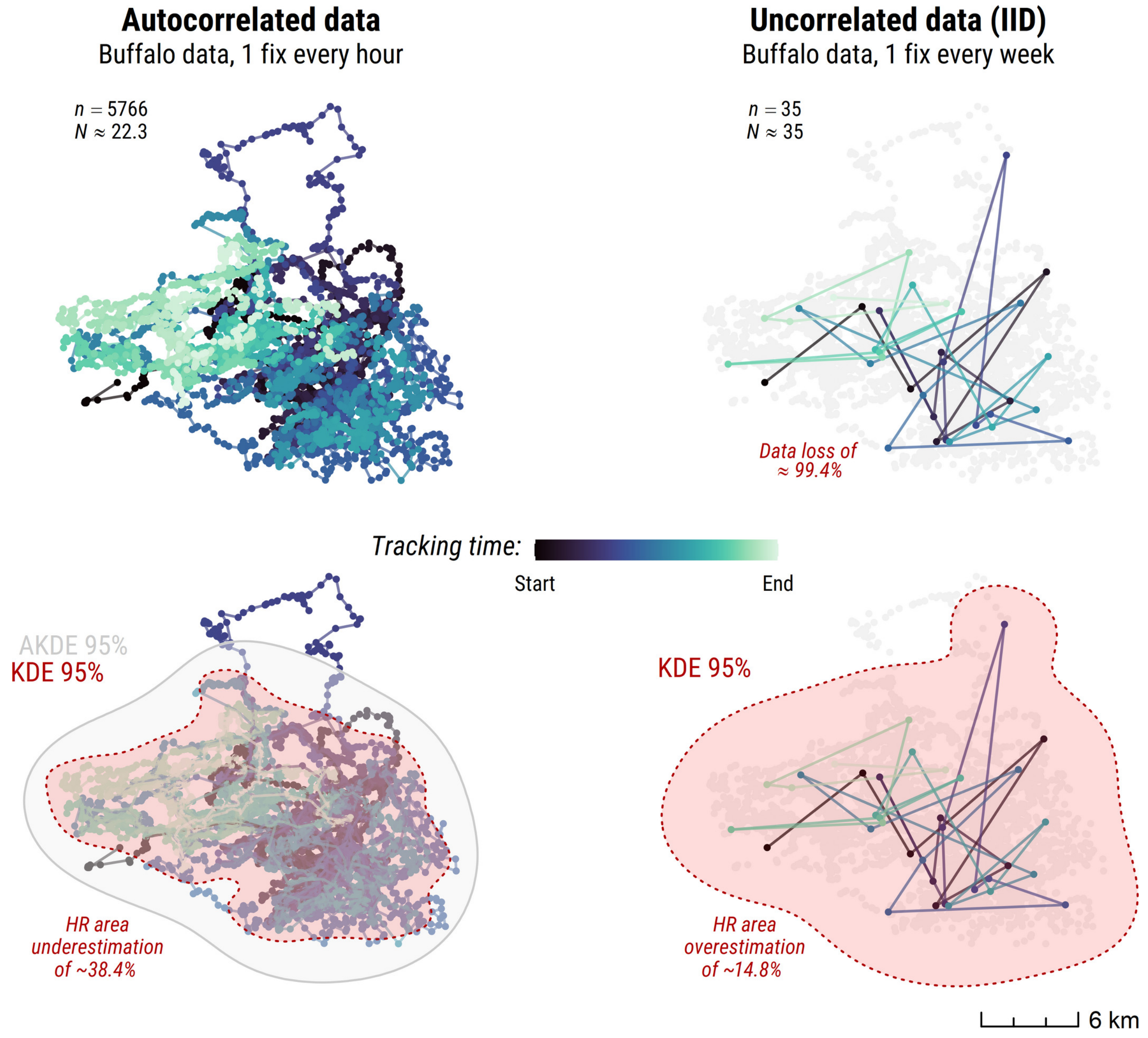
Pagano et al. 2021

Autocorrelated KDE (AKDE)

- Explicitly accounts for temporal autocorrelation
  - Properly accounts for dependence in successive observations, unlike other KDE methods
- Estimates 'range distribution' as opposed to 'occurrence distribution'
- Used for estimating home ranges
  - May provide excessively large estimates of space-use for a dispersing individual

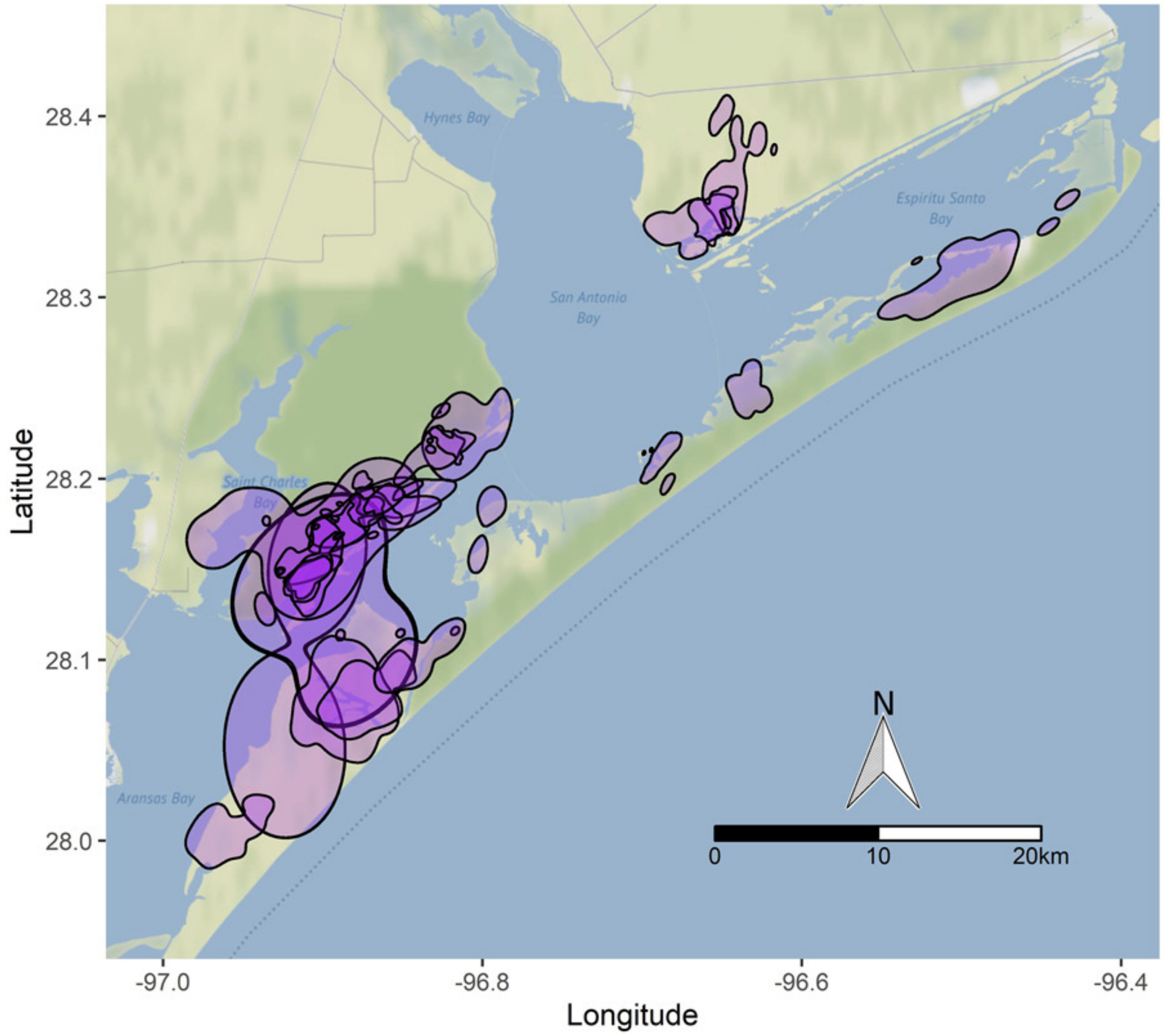
Noonan et al. 2019



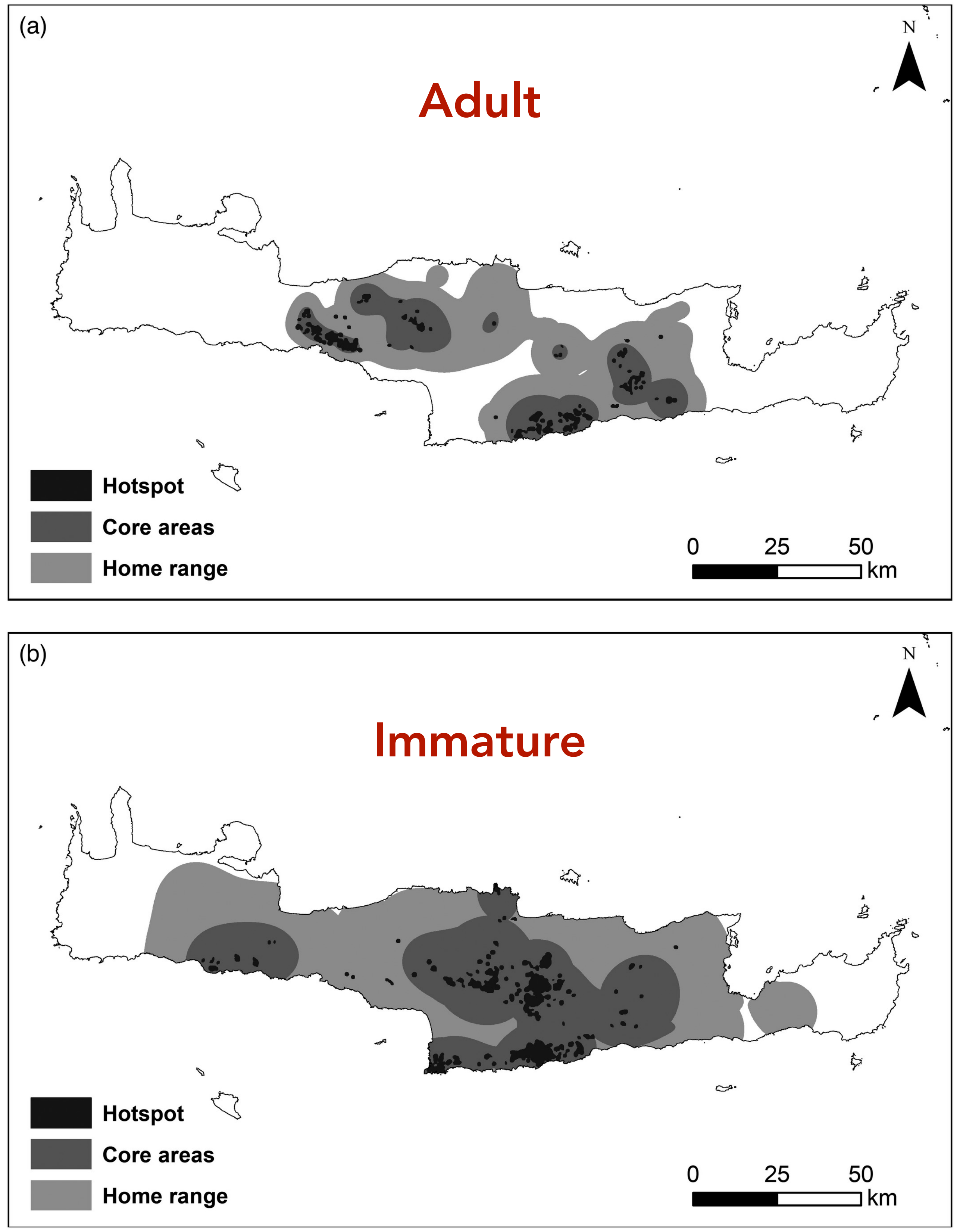


Silva et al. 2022

Motivating examples

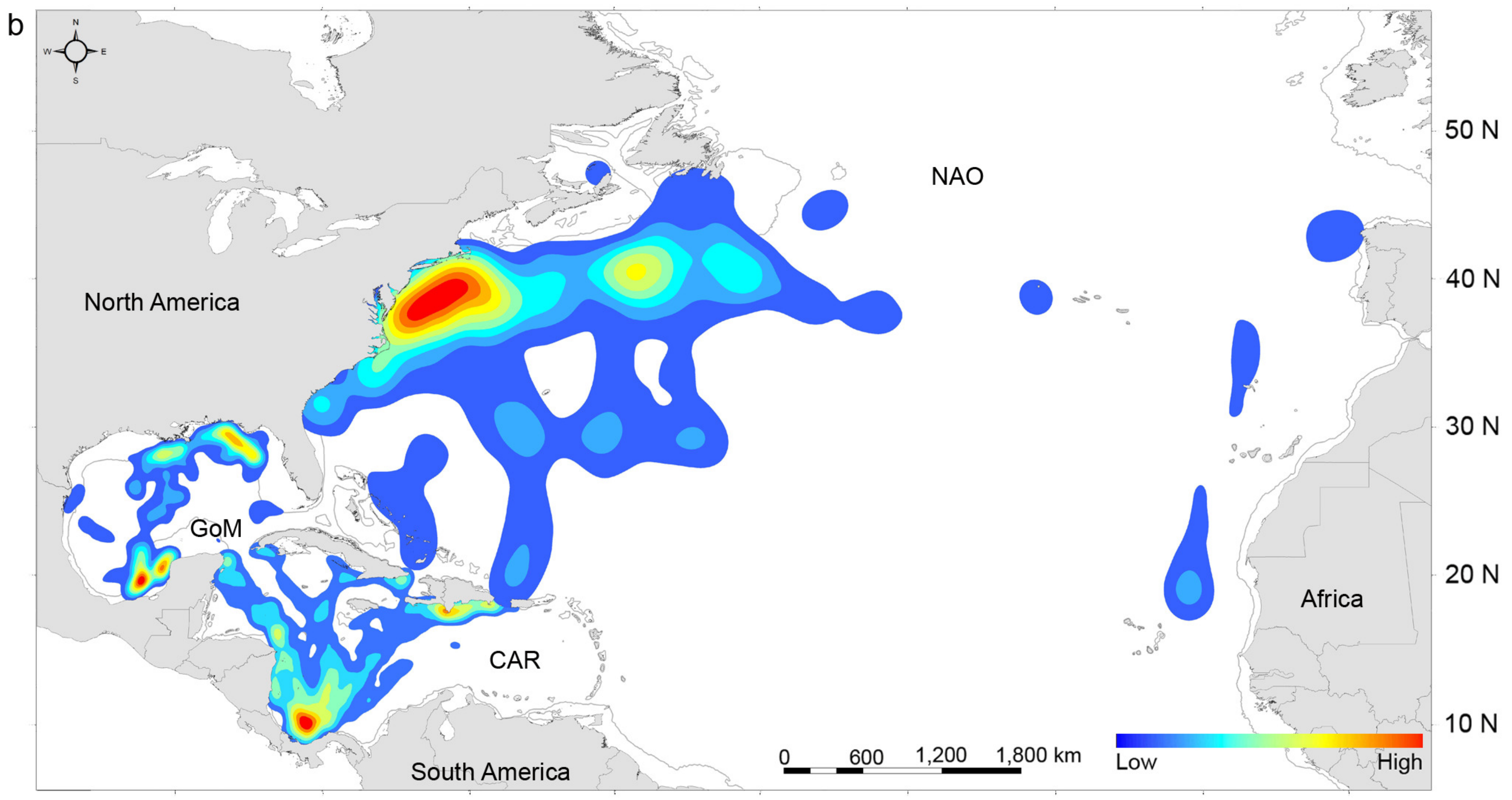


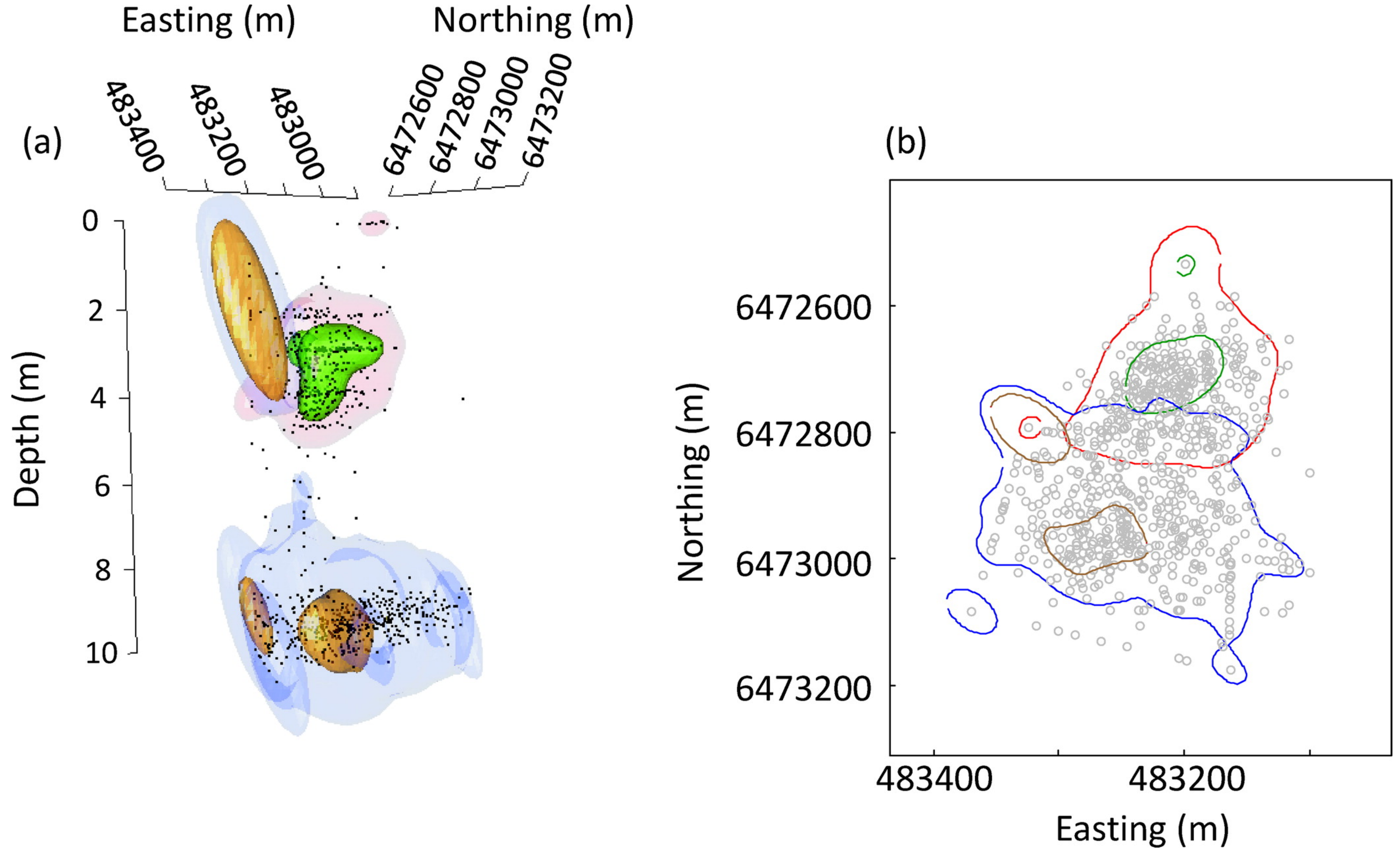
Butler et al. 2022



Xirouchakis et al. 2021

Evans et al. 2021





Simpfendorfer et al. 2012

Let’s calculate KDEs!

References

- Butler, M. J., Stewart, D. R., Harris, G. M., Bidwell, M. T., & Pearse, A. T. (2022). Space use and site fidelity of wintering whooping cranes on the Texas Gulf Coast. The Journal of Wildlife Management, e22226.
- Evans, D. R., Valverde, R. A., Ordoñez, C., & Carthy, R. R. (2021). Identification of the Gulf of Mexico as an important high-use habitat for leatherback turtles from Central America. Ecosphere, 12(8), e03722.
- Fieberg, J. (2007). Kernel density estimators of home range: smoothing and the autocorrelation red herring. Ecology, 88(4), 1059-1066.
- Fleming, C. H., & Calabrese, J. M. (2017). A new kernel density estimator for accurate home-range and species-range area estimation. Methods in Ecology and Evolution, 8(5), 571-579.
- Gitzen, R. A., Millspaugh, J. J., & Kernohan, B. J. (2006). Bandwidth selection for fixed-kernel analysis of animal utilization distributions. The Journal of Wildlife Management, 70(5), 1334-1344.
- Hart, K. M., Lamont, M. M., Iverson, A. R., & Smith, B. J. (2020). The importance of the northeastern Gulf of Mexico to foraging loggerhead sea turtles. Frontiers in Marine Science, 7, 330.
- Horne, E. J. S., Fieberg, J., Börger, L., Rachlow, J. L., Calabrese, J. M., & Fleming, C. H. (2020). Animal Home Ranges. Population Ecology in Practice.
- Noonan, M. J., Tucker, M. A., Fleming, C. H., Akre, T. S., Alberts, S. C., Ali, A. H., ... & Calabrese, J. M. (2019). A comprehensive analysis of autocorrelation and bias in home range estimation. Ecological Monographs, 89(2), e01344.
- Pagano, A. M., Durner, G. M., Atwood, T. C., & Douglas, D. C. (2021). Effects of sea ice decline and summer land use on polar bear home range size in the Beaufort Sea. Ecosphere, 12(10), e03768.
- Seaman, D. E., & Powell, R. A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. Ecology, 77(7), 2075-2085.
- Silva, I., Fleming, C. H., Noonan, M. J., Alston, J., Folta, C., Fagan, W. F., & Calabrese, J. M. (2022). Autocorrelation-informed home range estimation: A review and practical guide. Methods in Ecology and Evolution, 13(3), 534-544.
- Simpfendorfer, C. A., Olsen, E. M., Heupel, M. R., & Moland, E. (2012). Three-dimensional kernel utilization distributions improve estimates of space use in aquatic animals. Canadian Journal of Fisheries and Aquatic Sciences, 69(3), 565-572.
- Thomson, J. A., Börger, L., Christianen, M. J. A., Esteban, N., Laloë, J. O., & Hays, G. C. (2017). Implications of location accuracy and data volume for home range estimation and fine-scale movement analysis: comparing Argos and Fastloc-GPS tracking data. Marine biology, 164(10), 1-9.
- Walter, W. D., Onorato, D. P., & Fischer, J. W. (2015). Is there a single best estimator? Selection of home range estimators using area-under-the-curve. Movement Ecology, 3(1), 1-11.
- Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. Ecology, 70(1), 164-168.
- Xirouchakis, S. M., Grivas, C., Andreou, G., & Georgopoulou, E. (2021). Home range size, space use and resource selection of griffon vultures in an insular environment. Journal of Zoology, 314(2), 116-131.