

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Correction to: A cross-validation-based approach for delimiting reliable home range estimates

### Permalink

<https://escholarship.org/uc/item/6d08b1rk>

### Journal

Movement Ecology, 5(1)

### ISSN

2051-3933

### Authors

Dougherty, Eric R  
Carlson, Colin J  
Blackburn, Jason K  
et al.

### Publication Date

2017-12-01

### DOI

10.1186/s40462-017-0116-y

Peer reviewed

CORRECTION

Open Access



# Correction to: A cross-validation-based approach for delimiting reliable home range estimates

Eric R. Dougherty<sup>1\*</sup>, Colin J. Carlson<sup>1</sup>, Jason K. Blackburn<sup>2,3</sup> and Wayne M. Getz<sup>1,4</sup>

**Correction to: *Movement Ecology* (2017) 5:19**

**DOI:10.1186/s40462-017-0110-4**

## Original text

A grid-based exploration of parameter space was then conducted (Figure 2), whereby each of the 100 training/testing datasets was analyzed at every combination of  $k$  and  $s$  values on the grid. This analysis entailed the creation of local convex hulls with  $k$  nearest neighbors and a scaling factor of  $s$ . In all subsequent analyses, we assume that the scaling of time follows a linear formulation; however, when movement patterns more closely exemplify diffusion dynamics, an alternative equation for the TSD may be more accurate [1]. The test points were then laid upon the resulting hulls, and the probability of each was calculated as the proportion of the total number of hulls (equivalent to the total number of points in the training dataset) that contained the test point (Figure 1). Test points that were not contained within any hulls were assigned a probability equal to the inverse of the total number of points in the full movement path divided by 100, effectively penalizing any hull sets that did not include each of the test points. Though an arbitrary selection, the choice of a consistent penalty term across individuals will serve to standardize the procedure. A larger penalty will likely result in a higher optimal  $k$  value and bear a closer resemblance to the MCP. The natural log of the probability was calculated and information criterion values analogous to Akaike's Information Criterion (AIC) were derived using the equation:

$$IC = -2 * \ln \left( \sum_{i=1}^n P(\text{test points} | \text{training hullsets}) \right) + 2 * k$$

The choice of  $2k$  as the penalty term was made to maintain a structure analogous to the AIC equation. Given

\*Correspondence: [dougherty.eric@berkeley.edu](mailto:dougherty.eric@berkeley.edu)

<sup>1</sup>Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, USA

Full list of author information is available at the end of the article

the expansive literature concerning the performance and behavior of AIC under various scenarios, maintaining this structure may offer insight into similar strengths and weaknesses of the proposed approach. Ultimately, without such a penalty, all movement paths would tend towards a  $k$  equal to the number of points in the training set, such that each individual point was assigned a probability of one. It should be noted that this penalty term is specific to the  $k$  (nearest neighbors) method, but the underlying cross-validation procedure could very easily be extended for the optimization of the  $a$  (adaptive parameter) method if an appropriate penalty term is selected. An ideal penalty term would likely result in a increase of the information criterion value by a similar magnitude as in the  $k$ -based formulation above (i.e., ranging from approximately  $10^0$  to  $10^2$ ).

## Revised text

A grid-based exploration of parameter space was then conducted (Figure 2), whereby each of the 100 training/testing datasets was analyzed at every combination of  $k$  and  $s$  values on the grid. This analysis entailed the creation of local convex hulls with  $k$  nearest neighbors and a scaling factor of  $s$ . In all subsequent analyses, we assume that the scaling of time follows a linear formulation; however, when movement patterns more closely exemplify diffusion dynamics, an alternative equation for the TSD may be more accurate [1]. The test points were then laid upon the resulting hulls, and the probability of each was calculated as the proportion of the total number of hulls (equivalent to the total number of points in the training dataset) that contained the test point (Figure 1). Test points that were not contained within any hulls were assigned a probability equal to the inverse of the total number of points in the full movement path divided by 100, effectively penalizing any hull sets that did not include each of the test points. Though an arbitrary selection, the choice of a consistent penalty term across individuals will serve to standardize the procedure. A larger

penalty will likely result in a higher optimal  $k$  value and bear a closer resemblance to the MCP. The natural log of the probability was calculated and information criterion values analogous to the **Bayesian Information Criterion (BIC)** were derived using the equation:

$$IC = -2 * \ln \left( \sum_{i=1}^n P(\text{test points} \mid \text{training hullsets}) \right) + k * \ln(P)$$

$$\text{where } P = \sum_{i=1}^n (\text{test points})$$

The choice of  $k * \ln(P)$  as the overall penalty term was made to maintain a structure analogous to the **BIC** equation. Given the expansive literature concerning the performance and behavior of **BIC** under various scenarios, maintaining this structure may offer insight into similar strengths and weaknesses of the proposed approach. Ultimately, without such a penalty, all movement paths would tend towards a  $k$  equal to the number of points in the training set, such that each individual point was assigned a probability of one. **An alternative method akin to Akaike's Information Criterion can also be applied, but the penalty term ( $2 * k$ ) does not scale with the total number of test points (in turn, a function of the total length of the movement path) and will likely result in higher optimal  $k$  values than the BIC analogue.** It should also be noted that this penalty term is specific to the  $k$  (nearest neighbors) method, but the underlying cross-validation procedure could very easily be extended for the optimization of the  $\alpha$  (adaptive parameter) method if an appropriate penalty term is selected. An ideal penalty term would likely result in an increase of the information criterion value by a similar magnitude as in the  $k$ -based formulation above (i.e., ranging from approximately  $10^0$  to  $10^3$ ).

#### Explanation of correction

After the publication of this article [2], it came to our attention that the results presented throughout were based on an alternative Information Criterion (IC) equation that did not appear in the original article. The alternate formulation (akin to the Bayesian Information Criterion, rather than Akaike's Information Criterion) should be calculated as:

$$IC = -2 * \ln \left( \sum_{i=1}^n P(\text{test points} \mid \text{training hullsets}) \right) + k * \ln(P)$$

$$\text{where } P = \sum_{i=1}^n (\text{test points})$$

The only difference between the equation here and the one in the original article is the penalty term. In the equation above, increases in the  $k$  value are penalized more heavily than the simpler  $2 * k$  term. The additional benefit of this equation, and the primary reason for its use in the analysis in [2], is that the penalty term scales (in a non-linear fashion) with the total number of test points, offering more flexibility when considering trajectories of varying lengths.

Despite this issue, the fundamental principles underlying the cross-validation method remain sound, and both the original IC equation and the one presented here can be used with confidence. The logic for utilizing a BIC analogue is the same as that for formulating an AIC analogue; the correction outlined here simply enables the replication of the results in the article. The equation and the text in bold above have been altered from the original version of the paper.

#### Author details

<sup>1</sup>Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, USA. <sup>2</sup>Spatial Epidemiology and Ecology Research Laboratory, Department of Geography, University of Florida, Gainesville, FL, USA. <sup>3</sup>Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA. <sup>4</sup>School of Mathematical Sciences, University of KwaZulu-Natal, Durban, South Africa.

Received: 7 November 2017 Accepted: 7 November 2017

Published online: 05 December 2017

#### References

1. Lyons AJ, Turner WC, Getz WM. Home range plus: a space-time characterization of movement over real landscapes. *Mov Ecol.* 2013;1(1):2.
2. Dougherty ER, Carlson CJ, Blackburn JK, Getz WM. A cross-validation-based approach for delimiting reliable home range estimates. *Mov Ecol.* 2017;5(1):19.