

Spatial occupancy models for data collected on stream networks

Olivier Gimenez^{1*}

2024-07-22

¹ CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

* Corresponding author: olivier.gimenez@cefe.cnrs.fr

To monitor streams and rivers biodiversity, we need to quantify species distribution. To do so, occupancy models allow distinguishing the non-detection of a species from its actual absence. Occupancy models can account for spatial autocorrelation, but are not suited for streams and rivers because of their spatial structure in networks. Here I propose spatial occupancy models for data collected on stream and river networks. I present the statistical developments of the model, then I illustrate the approach on a semi-aquatic mammal. Overall, spatial stream network occupancy models provide a formal approach to assess biodiversity in streams and rivers.

Keywords: Bayesian statistics, Spatial stream network models, Occupancy models, Spatial autocorrelation, Wildlife monitoring

INTRODUCTION

Streams and rivers provide essential habitats for numerous species of animals and plants, many of which are endemic (Dudgeon et al. 2006). The ecological health of these freshwater ecosystems is paramount not only for the biodiversity they harbor but also for the ecosystem services they provide, which are indispensable to both wildlife and human populations (Vári et al. 2021). However, human activities are altering the natural conditions of streams, rivers and their associated riparian habitats, jeopardizing the persistence of these ecosystems (Albert et al. 2020).

In that context, species distribution models (SDMs) are essential tools in understanding and preserving biodiversity (Elith and Leathwick 2009). SDMs predict the distribution of species, helping scientists and conservationists identify critical habitats and biodiversity hotspots. SDMs also inform strategies aimed at mitigating impacts of climate and land-use changes, manage invasive species or enhancing habitat connectivity in freshwater ecosystems (Domisch et al. 2015).

SDMs are known to be affected by two main issues, namely imperfect detection and spatial autocorrelation (Guélat and Kéry 2018). First, a species present in a given area may go undetected during surveys due to various factors such as observer experience, species behavior, and environmental conditions. Ignoring *imperfect detection* can lead to biased estimates of species distribution, and flawed inference about the relationship between species presence and environmental factors (Kéry and Schmidt 2008, Lahoz-Monfort et al. 2014), potentially misinforming conservation strategies and habitat management decisions. To address this issue, occupancy models are SDMs that rely on repeated visits of spatial sampling units for inferring distribution (MacKenzie et al. 2017). These models have been widely used in freshwater ecosystems for various taxa (Hamer et al. 2021, Preece et al. 2021, Charbonnel et al. 2022, Wedderburn et al. 2022). Second, SDMs rely on the assumption of independent residuals,

which may be violated if sampling sites that are close together tend to have similar probabilities of species presence **pas très bien dit, et je ne parle jamais de pr d occupancy, y remédier?**.

Ignoring *spatial autocorrelation* can lead to biased estimates of species distribution, and can inflate the probability of erroneously detecting the effect of environmental covariates (Dormann et al. 2007). Several extensions of occupancy models have been proposed to account for spatial autocorrelation, building on existing conditional autoregressive and geosadditive models in the spatial statistics literature (Johnson et al. 2013, Broms et al. 2014, Rushing et al. 2019).

However, these models rely on the Euclidean distance between the spatial sampling units, which does not acknowledge the spatial structure of streams and rivers in networks.

Here I propose spatial occupancy models that allow spatial autocorrelation structured according to stream flow and flow connectivity. I build on the linear mixed modelling approach proposed by (Peterson and Hoef 2010, Ver Hoef and Peterson 2010), which allows considering a mixture of distance-based spatial correlation structures (Euclidean or not) in a single model. I plug-in this variance component approach into occupancy models using a Bayesian approach. I present the statistical developments of the model, and I illustrate the approach on a semi-aquatic mammal in French streams and rivers.

METHODS

<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2699.2011.02487.x>

How is spatial autocorrelation accounted for in occupancy models?

The usual way is to write the probability of occupancy on some scale, here the logit scale, as a function of covariates the X and some random effect ϵ to capture spatial variation.

This ϵ is distributed as a multivariate normal distribution with a covariance matrix which we write as the product of a variance term σ^2 times a correlation matrix R .

66 Big R captures the correlation between each pair of sites, the lower-case r 's.

67 These correlations are assumed to be a function of the distance between sites i and j , and a scale
68 parameter θ . Here this function is the exponential.

69 The correlation will be 1 at distance 0 when i and j are the same site. The correlation will decline
70 with increasing distance between two sites.

71 In addition, the larger the scale parameter θ , the slower the decrease in correlation with
72 distance.

73 The question now is which distance should we use. Let's get back to being an otter travelling
74 the river to find fish and shelter. If the flow goes from top to bottom. Then going from site 1 to 4
75 is tricky and the euclidean distance is irrelevant. So we prefer using stream distance instead,
76 and go from site 1 to 4 by following the flow. We will call them flow-connected sites. Let's
77 forget about otters for a minute. Let's assume we're studying organisms that disperse passively
78 with the flow, like mussels or some insects. We need to consider models that do not allow
79 autocorrelation between sites if the water at one site does not flow into another. Like here
80 between sites 1 and 2. We will call them flow-unconnected sites. So two things are important:
81 stream distance and flow connectivity.

82 Jay Ver Hoef and Erin Peterson back in 2010 developed statistical models that allow accounting
83 for stream distance and flow connectivity in linear models.

84 These are called spatial stream network models.

85 They add a random term with spatial covariance between flow-connected sites.

86 Then they add spatial covariance between flow connected and flow-unconnected sites.

87 This is for correlation between sites that can occur with active movements, with or against the
88 river flow ;

89 Last, they add correlation that can occurs independently of the network topology using the

Euclidean distance.

In which we structure the random effect like in spatial stream network models we've just seen.

Occupancy models

Any spatial or ecologically relevant variable can be used to construct the weights, as long as it is available for every segment in the stream network (Fig. 1). Simple options are to use equal weights or Shreve's stream order (Shreve, 1967), but the watershed area (i.e. area of land that drains downhill to a common point on the stream network) is commonly used as a surrogate for water volume (Frieden et al., 2014).

Tail-down models differ from tail-up models because they allow spatial correlation between both flow-connected and flow-unconnected locations.

Occupancy models rely on several assumptions. First, we assume closure which means we assume a site remains in the same state, occupied or non-occupied. Second we assume independence of the surveys and sites.

Case study

We used the statistical language R (R Core Team 2023) for all our analyses. These were implemented in SSN2 (Dumelle et al. 2023), tidyverse (Wickham et al. 2019), sf (Pebesma 2018, Pebesma and Bivand 2023), rstan (Stan Development Team 2023), bayesplot (Gabry et al. 2019, Gabry and Mahr 2022), geodata (Hijmans et al. 2024), ggspatial (Dunnington 2023), rmapshaper and (Teucher and Russell 2023) packages.

The species went almost extinct in the 20th century, it was hunted for its fur. Thanks to the banning of hunting and its protection, it is now recolonizing our country. In that context, the question we're asking is what is its current distribution.

To answer this question, we collect scats the species leaves behind.

Now to the case study. I investigated the effect of human disturbance on otter occupancy. The study site is in this rectangle here, between Montpellier where I live and Swansea. There were 56 sites that were visited 3 times. Clearly we have some spatial variation in the number of detections. I used human density as a proxy for human disturbance. Human density can lead to disturbance of otters through urbanization for example. We recorded human density as the number of inhabitants per km² in a buffer around each stream.

RESULTS

Trees in forest *A* grew taller than those in forest *B* (mean height: 25 versus 13 m).

And many more cool results that get updated dynamically, e.g. see Table ?? and Fig. ?. Note Tables and Figures are **cross-linked and numbered automatically**. They could also appear in the middle of the document, not necessarily at the end.

See also Fig. ? and Table ? in the Supplementary Material.

When spatial autocorrelation is ignored, we found a negative effect of human density on occupancy probability.

However, when we accounted for spatial autocorrelation using our new model, human density had no longer an effect on occupancy.

The effect size of human density increases when spatial autocorrelation is ignored.

The most likely explanation is that of a bias due to an omitted variable. Human density is spatially correlated, and its effect size is inflated. This bias is controlled when spatial autocorrelation is included. There is probably a difference in occupancy according to another variable that would need to be accounted for.

DISCUSSION

Si besoin, merge results and discussion sections.

The focal ecological parameter of interest is the occupancy probability ψ , which is represented similarly in all the five models compared. However, the precision of the occupancy estimation is impacted by the quality of the estimation for the detection process (kery2008, kellner2014). In this study, we focused on cases in which data is collected continuously, for example with sensors or opportunistic data. We aimed to evaluate whether modelling the detection process in continuous time could enhance the precision of the estimated probability of occupancy.

Ecologically, omitted bias variable. Check in Couturier which one it could be. Citer quand même le papier de Hodges sur spatial confusion. Citer aussi les échanges avec Jay VerHoef pour la difficulté in estimating le sill parameter. Mais ok pour covariate et prediction de l'occupancy.

Application à eDNA. Alors besoin peut-être d'autres termes spatiaux comme tail-up, et euclidean. Ici on a utilisé que tail-down car on s'intéresse aux otters. eDNA important pour monitor biodiv in freshwater and marine realms. If false positives worked out (voir work de E. Matechou), resste à proprement prendre en compte spatial autocorrelation.

Extension à dynamic occupancy models. Colonization function of distance to features that hamper movement of individuals. Citer les papier de Chandler, Morin et celui à paraître de Kervellec. L'idée est qu'on pourrait regarder la question de la connectivité dans river et stream network. Pas super compliqué, le papier du portugais a paved the road.

Extending to dynamic occupancy to write the probability of colonization as a function of distance to landscape features that might hamper movements, and therefore assessing connectivity.

Ethics and Integrity statements

Data availability statement

Data and code are available at <https://github.com/oliviergimenez?tab=repositories>.

Funding statement

This research is a product of the DISCAR group funded by the French Foundation for Research on Biodiversity (FRB) through its synthesis center CESAB.

Conflict of interest disclosure

The author has no conflicts of interest to declare.

REFERENCES

- Albert, J. S., G. Destouni, S. M. Duke-Sylvester, A. E. Magurran, T. Oberdorff, R. E. Reis, K. O. Winemiller, and W. J. Ripple. 2020. [Scientists' warning to humanity on the freshwater biodiversity crisis](#). *Ambio* 50:85–94.
- Broms, K. M., D. S. Johnson, R. Altwegg, and L. L. Conquest. 2014. Spatial occupancy models applied to atlas data show southern ground hornbills strongly depend on protected areas. *Ecological Applications* 24:363–374.
- Charbonnel, A., F. Blanc, P. Laffaille, M. Némot, and L. Buisson. 2022. Combining spatial dependence occupancy models and conservation gap analyses to promote species conservation: A case study with a threatened semi-aquatic mammal. *Biological Conservation* 270:109567.

Domisch, S., S. Jähnig, J. Simaika, M. Kuemmerlen, and S. Stoll. 2015. Application of species distribution models in stream ecosystems: The challenges of spatial and temporal scale, environmental predictors and species occurrence data. *Fundamental and Applied Limnology / Archiv für Hydrobiologie* 186:45–61.

Dormann, F. C., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. Daniel Kissling, I. Kühn, R. Ohlemüller, P. R. Peres-Neto, B. Reineking, B. Schröder, F. M. Schurr, and R. Wilson. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* 30:609–628.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews* 81:163–182.

Dumelle, M., E. Peterson, J. M. Ver Hoef, A. Pearse, and D. Isaak. 2023. SSN2: Spatial modeling on stream networks in R.

Dunnington, D. 2023. [Ggsatial: Spatial data framework for ggplot2](#).

Elith, J., and J. R. Leathwick. 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40:677–697.

Gabry, J., and T. Mahr. 2022. [Bayesplot: Plotting for bayesian models](#).

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman. 2019. [Visualization in bayesian workflow](#). *J. R. Stat. Soc. A* 182:389–402.

Guélat, J., and M. Kéry. 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution* 9:1614–1625.

Hamer, A. J., D. Schmera, and M. J. Mahony. 2021. Multi-species occupancy modeling provides novel insights into amphibian metacommunity structure and wetland restoration. *Ecological Applications* 31:e2293.

202 Hijmans, R. J., M. Barbosa, A. Ghosh, and A. Mandel. 2024. [Geodata: Download geographic](#)
 203 [data](#).

204 Johnson, D. S., P. B. Conn, M. B. Hooten, J. C. Ray, and B. A. Pond. 2013. Spatial occupancy
 205 models for large data sets. *Ecology* 94:801–808.

206 Kéry, M., and B. Schmidt. 2008. Imperfect detection and its consequences for monitoring in
 207 conservation. *Community Ecology - COMMUNITY ECOL* 9:207–216.

208 Lahoz-Monfort, J. J., G. Guillera-Arroita, and B. A. Wintle. 2014. Imperfect detection impacts
 209 the performance of species distribution models. *Global Ecology and Biogeography*
 210 23:504–515.

211 MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2017.
 212 *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species*
 213 *Occurrence*. Elsevier.

214 Pebesma, E. 2018. [Simple Features for R: Standardized Support for Spatial Vector Data](#). *The R*
 215 *Journal* 10:439–446.

216 Pebesma, E., and R. Bivand. 2023. [Spatial Data Science: With applications in R](#). Chapman and
 217 Hall/CRC.

218 Peterson, E. E., and J. M. V. Hoef. 2010. A mixed-model moving-average approach to
 219 geostatistical modeling in stream networks. *Ecology* 91:644–651.

220 Preece, E. P., M. Bryan, S. M. Mapes, C. Wademan, and R. Dorazio. 2021. Monitoring for
 221 freshwater mussel presence in rivers using environmental DNA. *Environmental DNA*
 222 3:591–604.

223 R Core Team. 2023. [R: A language and environment for statistical computing](#). R Foundation for
 224 Statistical Computing, Vienna, Austria.

225 Rushing, C., J. andrew Royle, D. Ziolkowski Jr, and K. Pardieck. 2019. Modeling spatially and
 226 temporally complex range dynamics when detection is imperfect. *Scientific Reports* 9.

227 Stan Development Team. 2023. [RStan: The R interface to Stan](#).

228 Teucher, A., and K. Russell. 2023. [Rmapshaper: Client for 'mapshaper' for 'geospatial'](#)
 229 [operations](#).

230 Vári, Á., S. Podschun, T. Eros, T. Hein, B. Pataki, C. Ioja, C. Adamescu, A. Gerhardt, T. Gruber,
 231 A. Dedić, M. Ciric, B. Gavrilović, and A. Báldi. 2021. Freshwater systems and ecosystem
 232 services: Challenges and chances for cross-fertilization of disciplines. *Ambio* 51:135–151.

233 Ver Hoef, J., and E. Peterson. 2010. A moving average approach for spatial statistical models of
 234 stream networks. *Journal of the American Statistical Association* 105:6–18.

235 Wedderburn, S. D., N. S. Whiterod, and L. Vilizzi. 2022. Occupancy modelling confirms the first
 236 extirpation of a freshwater fish from one of the world's largest river systems. *Aquatic*
 237 *Conservation: Marine and Freshwater Ecosystems* 32:258–268.

238 Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemond, A.
 239 Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J.
 240 Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and
 241 H. Yutani. 2019. [Welcome to the tidyverse](#). *Journal of Open Source Software* 4:1686.

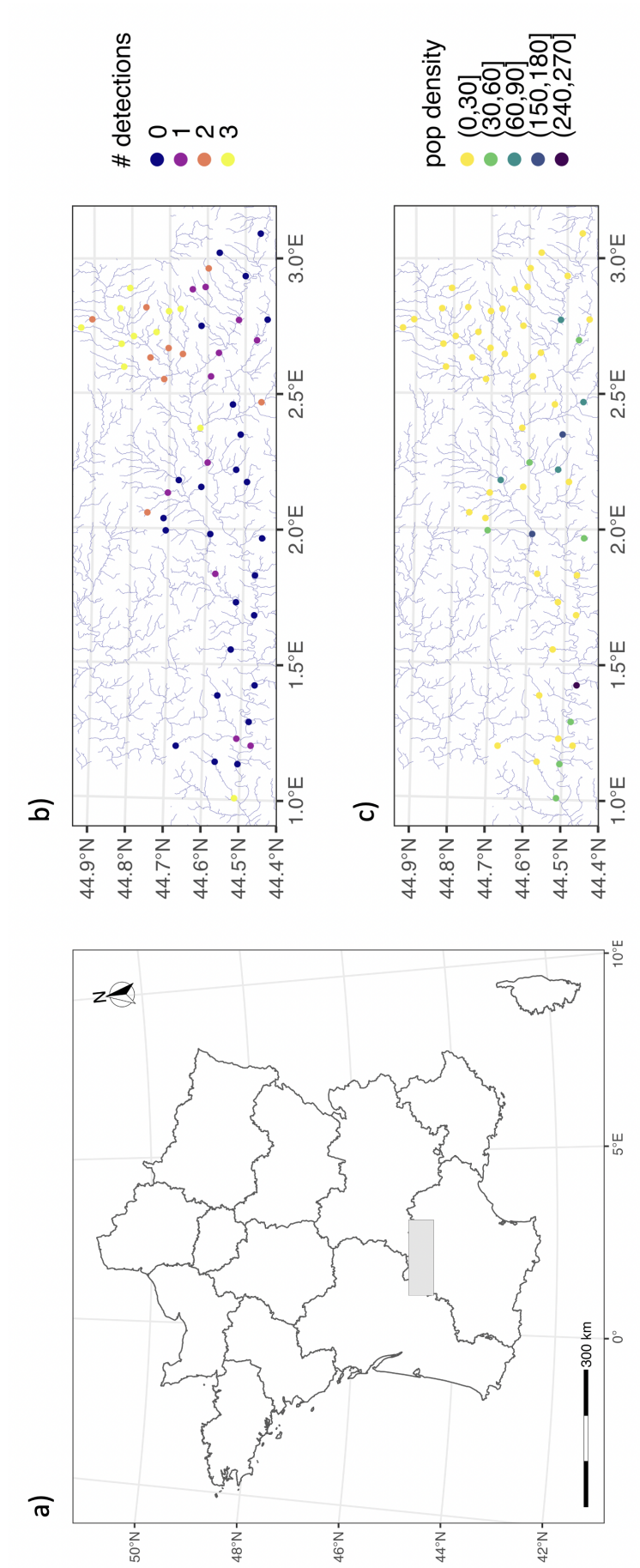


Figure 1: Second figure in landscape format.