

Spatial occupancy models for data collected on stream networks

Olivier Gimenez^{1*}

2024-07-25

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

18 Introduction

19 Streams and rivers provide essential habitats for numerous species of animals and plants, many
20 of which are endemic ([Dudgeon et al. 2006](#), [Reid et al. 2019](#)). The ecological health of these
21 freshwater ecosystems is paramount not only for the biodiversity they harbor but also for the
22 ecosystem services they provide, which are indispensable to both wildlife and human
23 populations ([Vári et al. 2021](#)). However, human activities are altering the natural conditions of
24 streams, rivers and their associated riparian habitats, jeopardizing the persistence of these
25 ecosystems ([Albert et al. 2020](#)).

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

32 SDMs are known to be affected by two main issues, namely imperfect detection and spatial
33 autocorrelation ([Guélat and Kéry 2018](#)). First, a species present in a given area may go
34 undetected during surveys due to various factors such as observer experience, species behavior,
35 and environmental conditions. Ignoring *imperfect detection* can lead to biased estimates of
36 species distribution, and flawed inference about the relationship between species presence and
37 environmental factors ([Kéry and Schmidt 2008](#), [Lahoz-Monfort et al. 2014](#)), potentially
38 misinforming conservation strategies and habitat management decisions. To address this issue,
39 occupancy models are SDMs that rely on repeated visits of spatial sampling units for inferring
40 distribution ([MacKenzie et al. 2017](#)). These models have been widely used in freshwater
41 ecosystems for various taxa ([Hamer et al. 2021](#), [Preece et al. 2021](#), [Charbonnel et al. 2022](#),
42 [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. 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 the spatial statistics literature, e.g. conditional autoregressive models (Johnson et al. 2013, Broms et al. 2014) and geoaddivitive models (Rushing et al. 2019). However, these models rely on the Euclidean distance between the spatial sampling units, which does not acknowledge the spatial structure in networks of streams and rivers.

Here I propose spatial occupancy models that allow spatial autocorrelation structured according to stream flow and flow connectivity (Peterson et al. 2013). 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

Occupancy models

To account for imperfect detection, I use occupancy models that allow inferring the actual species distribution (MacKenzie et al. 2017). We assume the monitoring of a species occurs over S spatial sampling units, or *sites*. If detection was perfect, the state z_i of a site i would be a Bernoulli random variable taking value 1 with *occupancy probability* ψ_i if the site was occupied, and 0 otherwise with probability $1 - \psi$. However, the ecological process of state occupancy z_i is

only partially known, because the species may go undetected whereas actually present of that site i . Therefore we need to consider the observation process, also a Bernoulli random variable, on top of the ecological process. When the species is detected on site i , say $y_i = 1$ with *detection probability* p , then that site is occupied, whereas if the species goes undetected with probability $1 - p$, i.e. $y_i = 0$, we simply do not know whether the site is occupied or not. Both parameters, ψ and p , can be modeled as functions of explanatory spatial variables, in the spirit of generalized linear models and logistic regressions. The only requirement to separately estimate the occupancy and detection probabilities is to collect data in at least two independent visits in time for a number of sites, and this temporal replication should be over a short period so that sites remain in the same state.

Spatial autocorrelation for stream networks

How is spatial autocorrelation accounted for in occupancy models? The usual way is to write the probabilities of occupancy $\psi = (\psi_1, \dots, \psi_S)$ on some scale, say the logit scale, as a function of explanatory variables gathered in a matrix \mathbf{X} with corresponding regression parameters β to be estimated, and add a random effect ϵ to capture spatial autocorrelation (Guélat and Kéry 2018):

$$\text{logit}(\psi) = \mathbf{X}\beta + \epsilon.$$

The random effect ϵ can be structured using a conditional autoregressive models and its extensions (Johnson et al. 2013) or geoaddivitive models (Rushing et al. 2019). Whatever the method, the proximity among sites is assessed using the Euclidean distance, which fails to adequately capture complex spatial dependencies in streams and rivers. Specifically, we are interested in flow connectivity and stream and river topology (Peterson et al. 2013). Following (Peterson and Hoef 2010, Ver Hoef and Peterson 2010), I consider two sites as being

flow-connected when water flows from an upstream site to a downstream site, and
flow-unconnected when they share a common confluence downstream but do not share flow.
Then I parameterize occupancy by rewriting the random effect as a mixture of four components
as follows:

$$\text{logit}(\psi) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\tau}_{tu} + \boldsymbol{\tau}_{td} + \boldsymbol{\tau}_{eu} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\tau}_{tu}$ is a random effect with spatial covariance between flow-connected sites that can
occur in the same direction of the river flow (*tail-up*, think of organisms that move passively
like, e.g., mussels), $\boldsymbol{\tau}_{td}$ is a random effect with spatial covariance between flow-connected and
flow-unconnected sites that can occur with or against the direction of the flow (*tail-down*, think
of organisms that move actively like, e.g., semi-aquatic mammals), $\boldsymbol{\tau}_{eu}$ is a random effect with a
spatial covariance independent of the network topology (generated by, e.g., air temperature or
precipitation), and $\boldsymbol{\epsilon}$ is a random effect which variance, often called the nugget, can absorb
extra-variability. Writing these covariance components is described in details elsewhere:
Appendix B in (Peterson and Hoef 2010), (Isaak et al. 2014). I provide an example below.

Implementation

For all analyses, I used the statistical language R (R Core Team 2023). I fitted models in the
Bayesian framework by specifying weakly informative priors (Northrup and Gerber 2018),
implementing a marginalized likelihood (Clark and Altwegg 2019) and using the `rstan` (Stan
Development Team 2023) package. I also used the `openSTARS` (Kattwinkel et al. 2020) and `SSN`
(Hoef et al. 2014) packages to build and characterize the network and calculate hydrological
distances.

Case study

(Couturier et al. 2023)

(Kervellec et al. 2023)

Sites in counties Lot (46), Aveyron (12) and Cantal (15). First two are in Pyrénées, third is in Massif Central.

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.

I use tail-down only. Exponential structure, with decreasing correlation with increasing distance. Ecrire le modèle ici avec l'exponentielle:

$$\text{Cov}(\epsilon_i, \epsilon_j) = \begin{cases} \sigma^2 \exp(-d_{ij}/\theta), & \text{if sites } i \text{ and } j \text{ are connected} \\ 0, & \text{otherwise} \end{cases}$$

Expliquer les paramètres d, s et theta.

Results

See Fig. 1.

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

148 l'occupancy.

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

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

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

160 From paper by Johnson. By exploiting autocorrelation when present, one can potentially
161 decrease the number of visits to specific survey units within a study area, instead relying on
162 spatial dependence to take the place of some temporal replication. ALSO. The issue of spatial
163 confounding is relatively new in the statistics literature. As Hodges and Reich (2010) state, it is
164 a "rich area of research." Experienced spatial modelers might say that confounding is desirable
165 as it allows an adjustment in fixed effects inference due to unmeasured (unknown), spatially
166 correlated covariates. However, Hodges, Hodges and Reich (2010) and Paciorek (2010)
167 illustrate bias and variance inflation depend on the structure of these unmeasured variables.
168 Because the structure of these latent spatial variables is unknown, a spatial model may not
169 account for them the way the researcher intends. Hodges and Reich (2010) note that without
170 knowledge of the missing structure, purposefully adding spatial confounding is a haphazard
171 adjustment that may bias the known fixed effects in unknown ways. There may be some
172 middle ground, however, that is worth exploring (Paciorek 2010) and could be the topic of

future research in the context of occupancy modeling and spatial modeling in general.

From Rushing paper. The distributions of most species are characterized by complex and dynamic variation in occurrence. Species distribution modeling seeks to relate this variation to environmental covariates and extrapolate these relationships to unsampled sites and times. Because habitats and species-habitat relationships change across both space and time, conventional GLM-based models rarely capture the inherent complexity of species distributions, especially when inferences are made across large spatial or long temporal scales. Here, we demonstrate a novel occupancy-based SDM that combines environmental predictors with a spatial GAM to model covariate relationships and complex, non-linear spatial variation in occupancy probability while accounting for imperfect detection.

Acknowledge similar approach for count data by (Lu et al. 2024).

Acknowledgments

Jay Ver Hoef, Edgar Santos Fernández and Maëlis Kervellec.

Ethics and Integrity statements

Data availability statement

Data and code are available at <https://github.com/oliviergimenez/spatial-stream-network-occupancy-model>.

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.
- Clark, A. E., and R. Altwegg. 2019. Efficient Bayesian analysis of occupancy models with logit link functions. *Ecology and Evolution* 9:756–768.
- Couturier, T., J. Steinmetz, P. Defos du Rau, D. Marc, E. Trichet, R. Gomes, and A. Besnard. 2023. Intensive agriculture as the main limiting factor of the otter's return in southwest France. *Biological Conservation* 279:109927.
- 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* 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.

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.

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.

Hoef, J. M. V., E. E. Peterson, D. Clifford, and R. Shah. 2014. SSN: An R package for spatial statistical modeling on stream networks. *Journal of Statistical Software* 56:1–45.

Isaak, D. J., E. E. Peterson, J. M. Ver Hoef, S. J. Wenger, J. A. Falke, C. E. Torgersen, C. Sowder, E. A. Steel, M.-J. Fortin, C. E. Jordan, and others. 2014. Applications of spatial statistical network models to stream data. *Wiley Interdisciplinary Reviews: Water* 1:277–294.

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

Kattwinkel, M., E. Szöcs, E. Peterson, and R. Schäfer. 2020. Preparing GIS data for analysis of stream monitoring data: The R package openSTARS. *Plos One* 15:e0239237.

Kervellec, M., T. Couturier, S. Bauduin, D. Chenesseau, P. D. du Rau, N. Drouet-Hoguet, C. Duchamp, J. Steinmetz, J.-M. Vandiel, and O. Gimenez. 2023. Bringing circuit theory into spatial occupancy models to assess landscape connectivity. *bioRxiv*.

- Kéry, M., and B. Schmidt. 2008. Imperfect detection and its consequences for monitoring in conservation. *Community Ecology* 9:207–216.
- Lahoz-Monfort, J. J., G. Guillera-Arroita, and B. A. Wintle. 2014. Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography* 23:504–515.
- Lu, X., Y. Kanno, G. P. Valentine, J. M. Rash, and M. B. Hooten. 2024. Using multi-scale spatial models of dendritic ecosystems to infer abundance of a stream salmonid. *Journal of Applied Ecology* 61:1703–1715.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2017. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier.
- Northrup, J., and B. Gerber. 2018. A comment on priors for Bayesian occupancy models. *Plos One* 13:e0192819.
- Peterson, E. E., and J. M. V. Hoef. 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology* 91:644–651.
- Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. A. Falke, M.-J. Fortin, C. E. Jordan, K. McNyset, P. Monestiez, A. S. Ruesch, A. Sengupta, N. Som, E. A. Steel, D. M. Theobald, C. E. Torgersen, and S. J. Wenger. 2013. Modelling dendritic ecological networks in space: An integrated network perspective. *Ecology Letters* 16:707–719.
- Preece, E. P., M. Bryan, S. M. Mapes, C. Wademan, and R. Dorazio. 2021. Monitoring for freshwater mussel presence in rivers using environmental DNA. *Environmental DNA* 3:591–604.
- R Core Team. 2023. [R: A language and environment for statistical computing](#). R Foundation for Statistical Computing, Vienna, Austria.
- Reid, A. J., A. K. Carlson, I. F. Creed, E. J. Eliason, P. A. Gell, P. T. J. Johnson, K. A. Kidd, T. J. MacCormack, J. D. Olden, S. J. Ormerod, J. P. Smol, W. W. Taylor, K. Tockner, J. C. Vermaire,

269 D. Dudgeon, and S. J. Cooke. 2019. Emerging threats and persistent conservation challenges
270 for freshwater biodiversity. *Biological Reviews* 94:849–873.

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

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

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

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

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

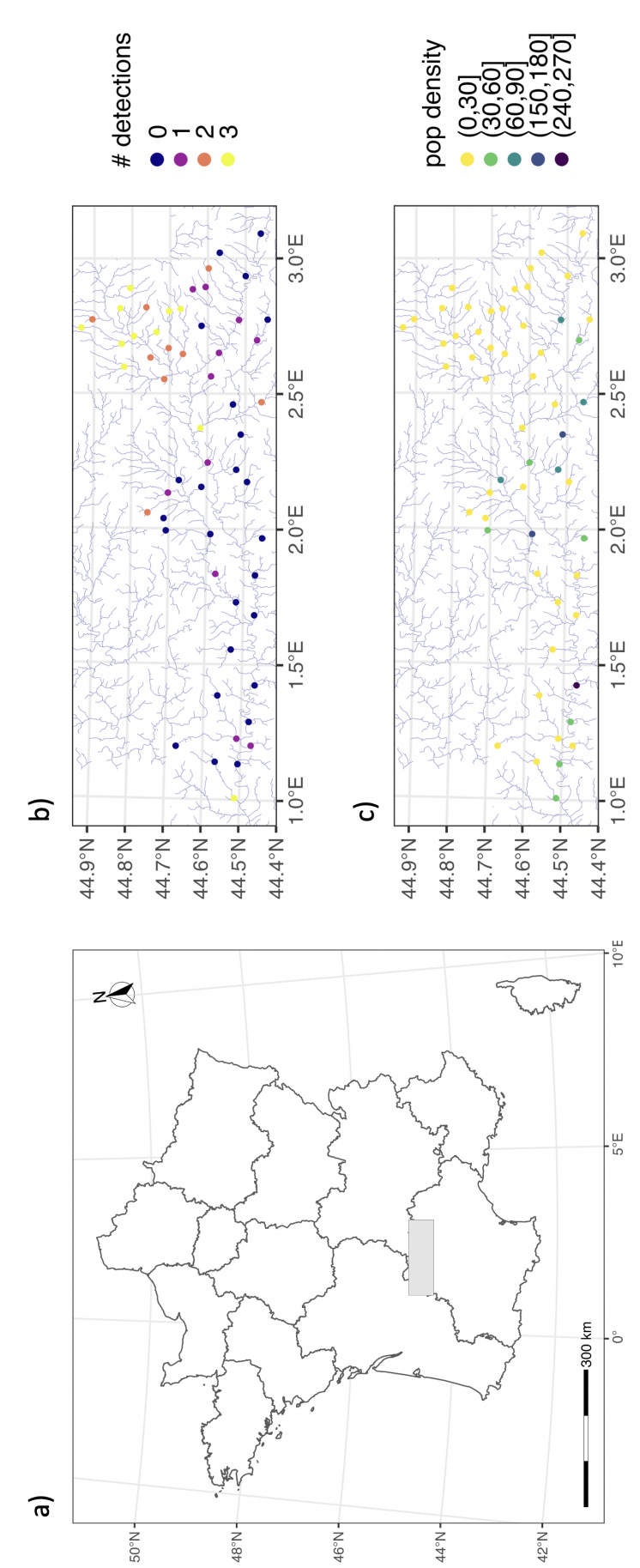


Figure 1: Second figure in landscape format.