The Annals of Applied Statistics
2021, Vol. 15, No. 4, 1628–1651
https://doi.org/10.1214/21-AOAS1447

MARKOV RANDOM FIELD MODELS FOR VECTOR-BASED REPRESENTATIONS OF LANDSCAPES

BY PATRIZIA ZAMBERLETTI[*], JULIEN PAPAÏX[†], EDITH GABRIEL[‡] AND THOMAS OPITZ[§]

Biostatistique et Processus Spatiaux (BioSP) , INRAE, [*]patrizia.zamberletti@inrae.fr; [†]julien.papaix@inrae.fr; [‡]edith.gabriel@inrae.fr; [§]thomas.opitz@inrae.fr

In agricultural landscapes the spatial distribution of cultivated and semi-natural elements strongly impacts habitat connectivity and species dynamics. To allow for landscape structural analysis and scenario generation, we here develop statistical tools for real landscapes composed of geometric elements, including 2D patches but also 1D linear elements (e.g., hedges). Utilizing the framework of discrete Markov random fields, we design generative stochastic models that combine a multiplex network representation, based on spatial adjacency, with Gibbs energy terms to capture the distribution of landscape descriptors for land-use categories. We implement simulation of agricultural scenarios with parameter-controlled spatial and temporal patterns (e.g., geometry, connectivity, crop rotation), and we demonstrate through simulation that pseudo-likelihood estimation of parameters works well. To study statistical relevance of model components in real landscapes, we discuss model selection and validation, including cross-validated prediction scores. Model validation with a view toward ecologically relevant landscape summaries is achieved by comparing observed and simulated summaries (network metrics but also metrics and appropriately defined variograms using a raster discretization). Models fitted to subregions of the Lower Durance Valley (France) indicate strong deviation from random allocation and realistically capture landscape patterns. In summary, our approach improves the understanding of agroecosystems and enables simulation-based theoretical analysis of how landscape patterns shape biological and ecological processes.

1. Introduction. Agroecosystems are the basis for food production and other ecosystem services, such as biodiversity, pollination and pest control (Power (2010), Foresight (2011)). Landscape heterogeneity plays an important role for many agroecological processes. It can be expressed through landscape *configuration*, referring to the size, shape and spatial-temporal arrangement of land-use patches (e.g., clustering, repulsiveness) and through landscape *composition*, referring to the number and proportion of land-use types (Martin et al. (2019)). Generative models are widely applied in landscape ecology for simulating virtual landscapes (i.e., a mosaic of fields having shapes and properties that vary in space and time and provide a support for biotic and abiotic processes) to systematically study the effects and impacts of landscape heterogeneity on ecosystem processes; see the recent reviews of Langhammer et al. (2019), Poggi et al. (2018). The purpose of such models is to generate a high number of virtual but structurally realistic maps of land-cover (Gardner (1999), Gardner and Urban (2007), Saura and Martinez-Millan (2000), Sciaini et al. (2018)), and, often, parameters related to landscape features, such as the percentage of land-cover, the habitat fragmentation or spatial autocorrelation (Langhammer et al. (2019)) can be controlled. In this paper we focus on modeling agricultural landscapes, and we consider neutral landscape models where the

Received July 2020; revised January 2021.
Key words and phrases. Graphical model, Markov chain Monte Carlo simulation, multiplex-network, pseudo-likelihood, statistical landscape modeling, stochastic geometry.

model does not directly interact with the biotic or abiotic processes (Gardner et al. (1987), With and King (1997)).

Existing models use either a vector-based or a raster-based representation, with the majority of models being of raster type. The raster approach is particularly useful for modelling gradual landscape dynamics and continuous processes (e.g., Lin et al. (2014)). However, agricultural landscapes are strongly characterized by polygon-shaped patches and piecewise linear corridors along polygon boundaries such that vector approaches seem preferable (Gaucherel et al. (2006a, 2006b), Inkoom et al. (2017), Le Ber et al. (2009), Papaïx et al. (2014), Langhammer et al. (2019)). In particular, fringe structures, such as hedgerows, roads or ditches aligned along polygon boundaries, have an important impact on many agroecological processes despite their small surface proportion. In a vector-based framework, Gaucherel et al. (2006a, 2006b) use models based on Gibbs energy terms to control certain pairwise interactions between landscape elements with the aim of simulating patches and certain fringe structures. Papaïx et al. (2014) develop a landscape generator without fringe structures that generates the landscape mosaic with two types of fields based on the Gibbsian T-tessellation model of Kiêu et al. (2013). However, existing modeling frameworks lack tools for parameter inference and model validation. Validation procedures are usually solely based on testing whether simulated landscapes are able to reproduce realistic landscape features by comparing observed and simulated landscape metrics (e.g., from the FRAGSTAT library, McGarigal and Marks (1995)). Such metrics are often directly used within simulation algorithms to enforce convergence toward target values (Langhammer et al. (2019)).

Vector-based approaches are independent of the grid resolution and give better control over small-surface elements, and they provide a sparser and more functional representation of patchy geometric structures without continuous gradients. The approach that we develop is geared toward flexible and realistic parametric stochastic modeling of fringe structures, such as hedgerow networks. For these reasons we advocate to turn away from the raster paradigm when modeling agricultural landscapes. Using a network-based representation of interactions among landscape elements, we construct Gibbs energies based on network structure (see, e.g., the recent collection of papers introduced by Fienberg (2010)) and, more specifically, models pertaining to the widely used class of discrete Markov random fields; see the seminal work of Besag (1972), Hammersley and Clifford (1971). Approaches relevant to our work are the nearest-neighbour Markov structures of Baddeley and Møller (1989) and the representations based on connected components introduced in Møller and Waagepetersen (1998). Recent developments and reviews are exposed in van Lieshout (2000, 2019), Gaetan and Guyon (2010), Green, Hjort and Richardson (2003). The use of Gibbs energies, and of Markov structures in particular, provides a natural distributional framework for controlling landscape descriptors. Likelihood-based statistical inference in such classification models, here formulated for categories of landscape elements, is notoriously difficult due to an intractable normalizing constant. We, therefore, resort to well-established pseudo-likelihood estimation, for which model selection and validation are more intricate and need to be performed carefully, especially in our setting, with only a single observed realization of the process.

We suppose that the polygon structure of patches in a bounded subset of planar space $\mathbb{R}^2$ is given, that is, a tessellation of space serves as fixed support of the model. It can be obtained by preprocessing a real landscape, or we may use simulations of a parametric tessellation model to generate realistic features (e.g., Kiêu et al. (2013)). We model the stochastic land-use allocation mechanism of patches and linear elements by assigning categories to the polygons and their edges, where dynamic structures such as crop rotation are possible.

An overarching goal is to generate visually realistic landscapes. We develop the following methodological novelties: (i) a mathematical representation of landscape composition and configuration through multilayer networks; (ii) generative stochastic parametric models coupling land-use allocation of patches and linear elements, relying strongly on Markov

interactions based on the network established in (i); (iii) simulation of such models using Markov chain Monte Carlo (MCMC) with the Gibbs sampler; (iv) statistical inference using real landscape data; (v) validation of relevant landscape characteristic based on a comparison of summaries for vector and raster representations between real and simulated landscapes. Our approach can handle relatively large landscapes by capitalizing on low computational requirements thanks to vector-based representations and to sparse-matrix structures for encoding interactions.

The paper is structured as follows. Section 2 presents real landscape data and preprocessing steps for an agricultural region in southeastern France, for which previous studies have highlighted a key role of agricultural practices and hedgerow configuration for biodiversity and pest control (Maalouly et al. (2013), Ricci et al. (2009), Lefebvre et al. (2016)). In Section 3 we propose the mathematical representation, modeling and simulation of landscapes. Tools for statistical inference, including model selection and validation, are developed in Section 4. In Section 5 we apply the developed framework to the above data, and we discuss how the goodness-of-fit and the generation of realistic landscape metrics is influenced by the choice of the descriptors in the model. A discussion in Section 6 concludes the paper. Supplementary Material contains details on the simulation algorithm and additional estimation and simulation results (Zamberletti et al. (2021)).

2. Landscape data. Real data for agricultural landscapes are based on remote sensing images, digital land registers, land-cover data bases such as CORINE (Büttner and Maucha (2006)) and field data. Often, manual annotation steps are necessary to complete and clean data. We study the Lower Durance Valley in southeastern France, depicted in Figure 1(a), stretching over 163 km^2 and mainly characterized by agricultural activity (87%) and urbanized areas, with main cultures of open area (46%) and apple/pear orchards (24%).

Data are based on manual digitalization (ArcView software) using an official French database of aerial photographs (BD ORTHO, IGN, 2004, 0.5 m resolution, updated with field monitoring in 2009).

The region has a total length of 1146 km of hedgerows, which we will represent as linear segments, whose average length amounts to 105 m. A particularity of the region is a dominance of East–West oriented hedges, whose function is to break the strong Mistral winds blowing from the North.

For the data application in this paper, we select three subdomains D1, D2 and D3 with contrasting properties and dimensions, shown in Figure 1(b) and numerically summarized in Table 1: D1 is relatively small and dominated by seminatural surfaces; D2 has the same surface area but equal proportions of semi-natural and crop; D3 delimits a much larger domain including D1 and D2.

We use a simplified representation of the landscape as a tessellation of 2D space with polygon-shaped cells. Linear segments (e.g., hedgerows) correspond to polygon edges. To achieve a partition of space through polygon-shaped patches and to align hedgerows with polygon edges, we preprocess the landscape toward a polygon tessellation of 2D space (Boots, Okabe and Sugihara (1992)), based on a heuristic loss criterion measuring the distance between original and transformed landscape (Adamczyk-Chauvat et al. (2020)). Figure 2 illustrates that preprocessing modifications for domain D2 are mostly minor. For simplicity, we here attribute always one of the two categories of "crop" or "semi-natural" to the patches in the three study domains. Specifically, we gather several types of seminatural habitat into a single category, including some patches with built structures (farms, greenhouses...). In principle, the subsequent modeling would also allow for parts of the landscape with unspecified category. While we here consider tessellations as a fixed support for linear element attribution and crop rotation, tessellation simulation algorithms for agricultural landscapes (Kiêu et al. (2013), Papaïx et al. (2014), Poggi et al. (2018)) would enable the generation of new, synthetic but realistic supports for our models.

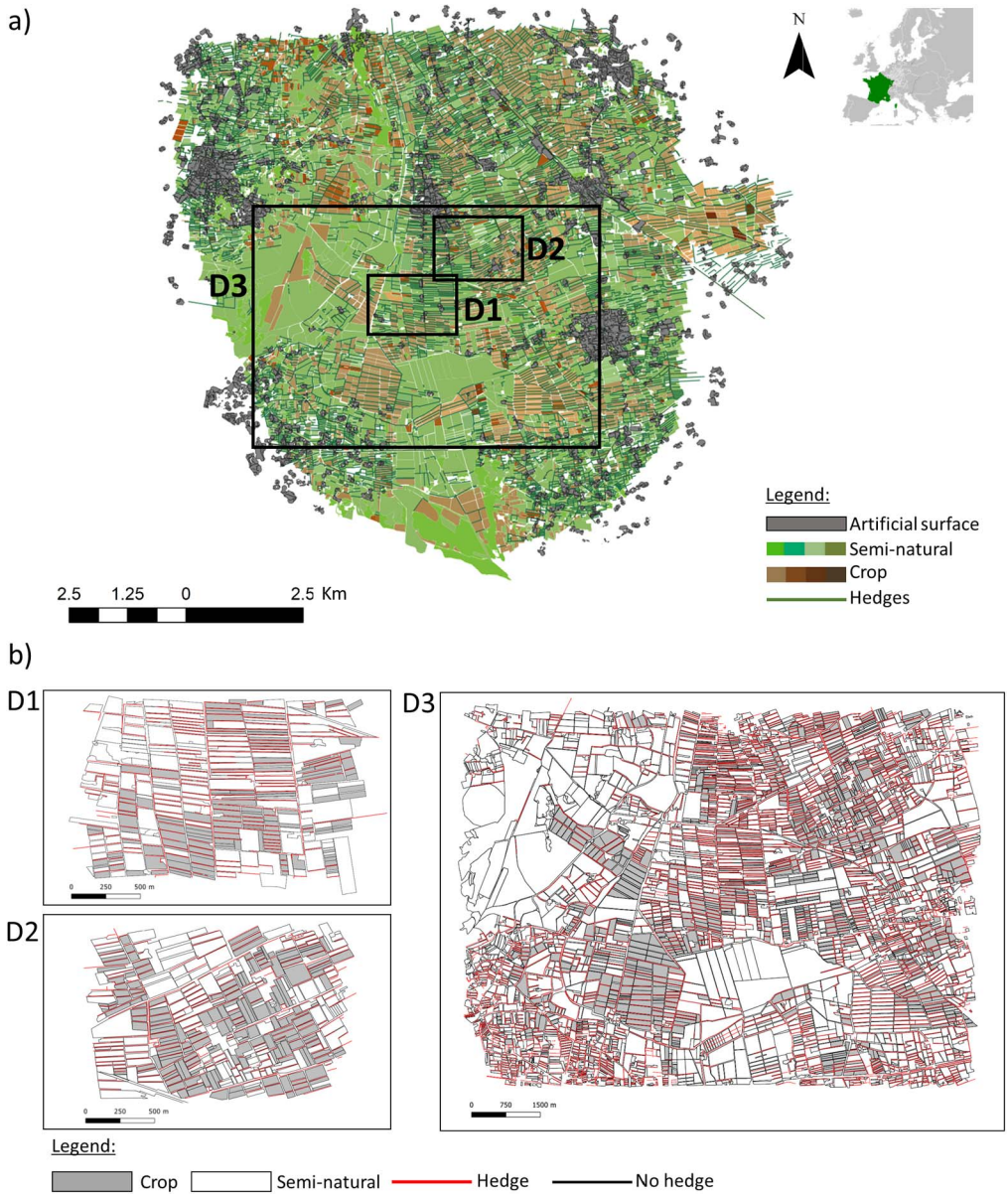

FIG. 1. *Lower Durance Valley study area.* (*a*) *Full area with three subdomains.* (*b*) *Subdomains D*1, *D*2, *D*3. *The Lower Durance Valley is characterised mainly by agricultural cover*: *Green-shaded patches represent semi–natural area* (*i.e.*, *woods*, *open area*, *grassland*); *brown-shaded patches represent* 34 *different cultures* (*e.g.*, *apple*, *pear*, *vineyards*). *Artificial surface* (*dark gray*) *consists of built structures and urbanized area. The area is rich in linear elements* (*i.e.*, *segments*), *including small water courses*, *roads and hedges* (*Panel* (*a*)). *In the selected domains* (*Panel* (*b*)), *we selected as "crop" the category of "apple/pear orchard," as it is the most abundant culture* (*gray patches*), *and we simplify the rest of the landscape surface as seminatural area* (*white patches*) *in order to establish a continuous cover with two categories. Patch boundaries are presented as linear elements which are marked in red when hedges are present.*

3. Stochastic modeling and simulation of landscape allocation.

3.1. *Mathematical landscape representation.* We propose to represent a landscape as a collection $\mathcal{O} = \{o_1, \ldots, o_n\}$ of n geometric objects as follows:

$$o_i = (x_i, z_i), \quad x_i \in \mathcal{X}_i = \{0, 1, \ldots, \ell_i - 1\}, i = 1, \ldots, n, \tag{1}$$

TABLE 1
Summary of selected subregions of the Lower Durance Valley study area; see Figure 1

	D1	D2	D3
Area (km^2)	3.37	2.3	41.13
% of Semi-natural	73	50	76
% of Crop	27	50	24
Hedgerows (km)	44.64	33.61	386.36
No. of patches	368	468	4379
No. of linear segments	1105	1405	12517

where each element is composed of two sets of data, x_i and z_i. The information in $\boldsymbol{z} = (z_1, \ldots, z_n)$ represents the geometrical structure of the landscape, determining object dimension and their organisation, and it is considered as being fixed. The vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ represents categories that we allocate to the geometric elements in the landscape, such as land-use types among hedges, water courses (for linear elements) or crop, grassland (for patches), and that we aim to model. We suppose that $x_i \in \mathcal{X}_i$ with a finite space $\mathcal{X}_i$ of $\ell_i \geq 1$ possible categories for the ith element, where the index 0 usually represents a baseline category. The objects $o_i = (x_i, z_i)$ could represent different geometric types, such as polygons (i.e., habitat patches) or linear segments (i.e., linear landscape elements); see Figure 3(a). For polygon objects the data component z_i could contain this type information as well as the geographical coordinates of its vertices, its surface area, and potentially other exogenous covariates. For instance, we could allocate each polygon with a category among the following three options: *crop* ($x_i = 1$), *(semi-)natural habitat* ($x_i = 2$), *other* ($x_i = 0$). A linear segment could be allocated with a category among *hedgerow* ($x_i = 1$) or *no hedgerow* ($x_i = 0$). In the case $\ell_i = 1$ with only a single category $x_i = 0$, no choice of allocation has to be made. The space of all possible combinations of allocations is $\mathcal{X} = \mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \cdots \otimes \mathcal{X}_n$. This finite collection contains $|\mathcal{X}| = \ell_1 \times \ell_2 \times \cdots \times \ell_n$ possible allocations. If the geometric structure contained in z_i may vary through time, it is possible to describe temporal dynamics (if present) by the sequence $x_{i,\tau}$, $\tau = 1, 2, \ldots$ of categories allocated within patches over discrete time.

3.2. *Network model of landscape.* We use a graphical representation of landscape to capture spatial or functional adjacency of landscape elements such as patches or linear segments in Figure 3(b). Adjacency of objects is modeled through a multilayer or multiplex network,

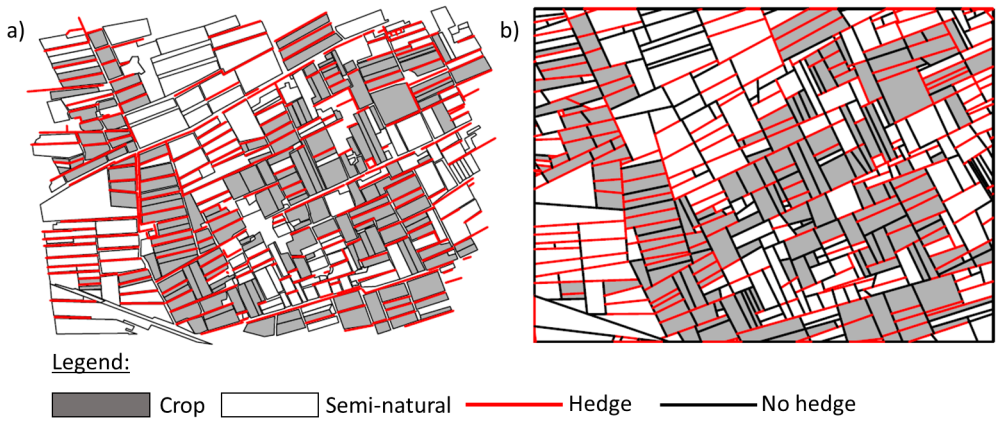

FIG. 2. *Preprocessing of domain D2. (a) Original digitalized shapefile; (b) Preprocessed landscape tessellation defined over a rectangular domain.*

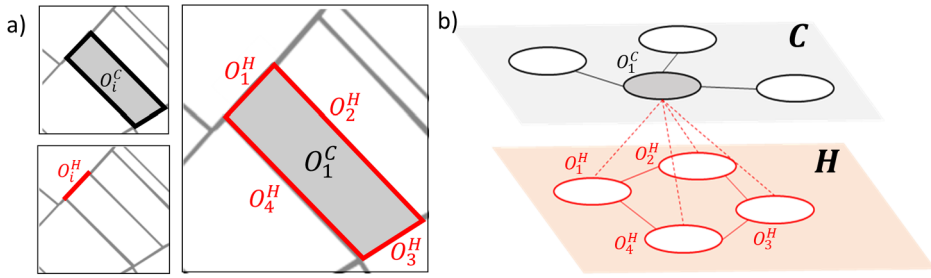

FIG. 3. *Landscape representation.* (*a*) *Polygon objects* (*patches*, *in grey*) *and linear segment objects* (*in red*). (*b*) *Multilayer network of connections. Layer C*: *Single network of connections between patches*; *layer H*: *Single network of connections between linear elements*; *links between C and H represent connections of patches and linear elements.*

that is, a set of single network layers with some nodes connected between layers (Boccaletti et al. (2014), Kivelä et al. (2014)). Each layer in this graphical representation corresponds to an object type; nodes stand for individual objects; edges in single network layers represent adjacency of objects of the same type; edges between different network layers represent adjacency of objects of different type. There are two types of networks that can be considered. First, for the specification of the Markov models we develop, we need a network that represents the fixed landscape support where the layers correspond to patches and to linear segments, respectively, in our setting. The probability of landscape category allocations in *pairwise Markov models* is then constructed from individual contributions of the nodes and edges in this network. Second, given a landscape allocation, we can consider the network where each layer corresponds to a specific allocation category. In the case of categories "crop" and "semi-natural" for patches, and of categories "hedge" and "no hedge" for linear segments, we obtain four layers. This second network type is useful for calculating landscape metrics taking into account allocation.

We illustrate the structure of the first network used for constructing Markov models. We define a collection of objects with two types, $\mathbf{o} = (\mathbf{o}^C, \mathbf{o}^H)$ (see Figure 3(a)), where $o_i^C = (x_i^C, z_i^C)$, $i = 1, \ldots, n^C$ represent patches (layer C), and $o_i^H = (x_i^H, z_i^H)$, $i = 1, \ldots, n^H$ represent linear segments (layer H); see Figure 3(b). We express that two distinct objects o_1 and o_2 are directly connected through an edge in the graph (i.e., they are adjacent) using the following notation:

$$o_1 \sim o_2, \quad o_1, o_2 \in \mathcal{O}. \tag{2}$$

For the models in this paper, we assume that two patches o_i^C, o_j^C are connected, $o_i^C \sim o_j^C$, if they are adjacent, that is, if they share part of their physical boundary; two linear elements are connected if they intersect or have a vertex in common; finally, interlayer connections $o_i^C \sim o_j^H$ arise if the linear element o_j^H is located on the boundary of patch o_i^C. This structure is similar to the nearest-neighbour relations discussed in Section 4 of Baddeley and Møller (1989) with respect to Markov properties.

For mathematical operations based on the network structure, we encode the object interactions in the *network matrix* (or *adjacency matrix*) $\mathcal{A}$,

$$\mathcal{A} = \begin{pmatrix} A^C & A^{CH} \\ A^{HC} & A^H \end{pmatrix}, \qquad \mathcal{A}_{i,j} = \begin{cases} 1, & o_i \sim o_j, \\ 0, & o_i \nsim o_j, \end{cases} \quad i, j \in \{1, \ldots, n^C + n^H\}, \tag{3}$$

where $A^C \in \mathbb{R}^{n^C \times n^C}$ and $A^H \in \mathbb{R}^{n^H \times n^H}$ represent the network matrices of intralayer connections of C and H, respectively, and $A^{CH} \in \mathbb{R}^{n^C \times n^H}$ encodes interlayer connections among C and H. For simplicity, we here assume symmetric connections with binary weights, such

that $\mathcal{A}_{ij} \in \{0, 1\}$ and $\mathcal{A} = \mathcal{A}^T$, but the extension to asymmetric and directed connections with $\mathcal{A}_{ij} \in \mathbb{R} \setminus \{0\}$ if $o_i \sim o_j$ would be straightforward. Nonbinary weights could be based on distance or sizes of connected elements.

Based on this landscape representation, we develop parametric probability distributions over the allocations $\boldsymbol{x} \in \mathcal{X}$, conditional on the (fixed) information in $\boldsymbol{z} = (z_1, \ldots, z_n)$ and $\mathcal{A}$. We put focus on Markov models where we assume conditional independence of category x_i with respect to $\boldsymbol{z}$ and the categories of objects not directly connected with o_i through the $\sim$-relation of adjacency.

We adopt notations such as $\boldsymbol{o}_{-i}$ to refer to the set $\mathcal{O} \setminus \{o_i\}$. Therefore, we make the following assumption of equality of conditional distributions:

$$x_i \mid (\boldsymbol{z_i}, \boldsymbol{o_{-i}}) \overset{d}{=} x_i \mid (\boldsymbol{z}, \{o_j \in \mathcal{O} \mid o_i \sim o_j\}), \quad i = 1, \ldots, n^C + n^H. \tag{4}$$

This framework allows for flexible dependence structures expressed through the adjacency matrix $\mathcal{A}$ with sparse structure, that is, with a relatively small proportion of nonzero entries.

3.3. *Probabilistic mechanistic models for landscape descriptors.* We utilize Gibbs energies to define probabilistic models of mechanistic nature, including Markov processes; see, for example, Cressie (1991), van Lieshout (2019). We construct a model using m functions $T_k : \mathcal{X} \to (-\infty, \infty)$, $k = 1, \ldots, m$ that each measure the value $T_k(\boldsymbol{x} \mid \boldsymbol{z})$ of a summary statistic for the allocations in $\boldsymbol{x}$ given the fixed information in $\boldsymbol{z}$. In the following we often omit $\boldsymbol{z}$ for notational simplicity when no confusion arises; for example, we simply write $T_k(\boldsymbol{x})$. We refer to the T_k as *landscape descriptors* and use them as sufficient statistics of the model by defining the probability of observing an allocation $\boldsymbol{x}$ as follows, with coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^m$:

$$p(\boldsymbol{x}) = \frac{1}{c(\boldsymbol{\beta})} \exp\left(- \sum_{k=1}^{m} \beta_k T_k(\boldsymbol{x}) \right), \quad \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\beta} \in \mathbb{R}^m. \tag{5}$$

The normalizing constant $c(\boldsymbol{\beta}) > 0$, also known as the *partition function*, ensures that probabilities in (5) sum up to 1.

Since the number of possible configurations $|\mathcal{X}|$ is finite, the normalizing constant is finite, and the model is well defined. In practice, the number of configurations is usually very large, such that numerical computation of the constant $c(\boldsymbol{\beta})$ is not tractable. If all descriptors T_k can be represented as sums of terms for single objects or two objects linked in the network through the $\sim$-relation in (2), the Markov property (4) holds.

We will also explore extensions beyond pairwise interactions but related to connected components where the descriptor is still defined through the graph structure. Instead of the *global specification* in equation (5), we now consider a *local specification*, that is, the allocation of x_i conditional to fixed information z_i and the rest of the landscape. Therefore, we determine the probability of observing category x_i given $\boldsymbol{z}$ and the allocations $\boldsymbol{x}_{-i}$ of all the other elements, where we use the notation $(\boldsymbol{x}_{-i}, x) = (x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)$ to indicate an (arbitrary) category x attributed to the object o_i. Then, the normalizing constant $c(\boldsymbol{\beta})$ cancels out in the conditional probability

$$p(x_i \mid \boldsymbol{x}_{-i}) = \frac{p(\boldsymbol{x})}{\sum_{y \in \mathcal{X}_i} p(\boldsymbol{x}_{-i}, x)} = \frac{\exp(-\sum_{k=1}^{m} \beta_k T_k(\boldsymbol{x}))}{\sum_{x \in \mathcal{X}_i} \exp(-\sum_{k=1}^{m} \beta_k T_k(\boldsymbol{x}_{-i}, x))}, \tag{6}$$

where the denominator adds up the probabilities over all landscape configurations obtained when varying the category of o_i but keeping the rest of the landscape fixed. In the two-level case with $x_i \in \{0, 1\}$, we show in Section 4.1 how parameters β_k can be estimated through classical logistic regression.

3.4. *Examples of parametric models*. Landscape descriptors are intended to capture important landscape characteristics. In *composition terms* such functions are the sum of contributions of individual objects; in *configuration terms* (or *interaction terms*) we add up contributions that evaluate the spatial interaction of two or more objects. An example specification is as follows, with three generic spatial landscape descriptors given by

$$T^{C}_{\text{act}}(\boldsymbol{x}) = \sum_{i=1}^{n^C} t\big(x_i^C\big), \qquad T^{CC}_{\text{adj}}(\boldsymbol{x}) = \sum_{o_i^C \sim o_j^C} t\big(x_i^C, x_j^C\big), \tag{7}$$
$$T^{CH}_{\text{adj}}(\boldsymbol{x}) = \sum_{o_i^C \sim o_j^H} t\big(x_i^C, x_j^H\big).$$

Then, T_{act} is a composition term, T^{CC}_{adj} an interaction term for network layer C and T^{CH}_{adj} is an interaction term for interlayer interactions of C and H. Figure 4 illustrates landscape descriptor evaluation on a subarea of D1 using the network to represent landscape interactions. The adjacency network of landscape elements is fixed for all landscape allocations and is based on the information in $\boldsymbol{z}$. It is defined among all the objects of the same layer C (Figure 4(a)), layer H (Figure 4(b)) and the interlayer of C and H (Figure 4(c)). It represents all pairwise interactions (i.e., all the interactions between adjacent patches (4(a)), adjacent linear segments (Figure 4(b)) and adjacent linear elements and patches (Figure 4(c))). Next, given an allocation of this landscape support, the second network type (also called *active network*) is used to represent the adjacency of objects allocated with the same category (e.g., *crop* category for patches, *hedge* for linear elements). To calculate the landscape descriptors T_k, we illustrate the additive contribution of a single object provides in Figure 5. The fixed information in $\boldsymbol{z}$ characterises the objects through their geometrical properties and assesses how the category allocation could be influenced by features such as the size of the patch (Figure 5(b), (c)), the length or orientation of the linear elements and determines the adjacency with respect to other objects (Figure 5(d), (e)).

Table 2 illustrates relevant choices of landscape descriptors involving C and H, that is, patches and linear elements, with *two* allocation categories (i.e., $x_i \in \{0, 1\}$): *crop* ($x_i^C = 1$) or *natural habitat* ($x_i^C = 0$), and *hedgerow* ($x_i^H = 1$) or *no hedgerow* ($x_i^H = 0$). In the Supplementary Material (Section 2) (Zamberletti et al. (2021)), a temporal descriptor for crop rotation is illustrated.

We employ the label of *activity terms* for composition terms where $T(\boldsymbol{x})$ is the count of the number of objects of a specific category. To ensure identifiability, we fix a reference category (e.g., $x_i^C = 0$ for objects of type C) and specify the activity term and its coefficient $\beta^C_{x_i^C} \in \mathbb{R}$ only for categories $x_i^C \neq 0$ such that it is expressed relative to $x_i^C = 0$, and, implicitly we have $\beta^C_{x_i^C=0} = 0$. A positive coefficient $\beta^C_{x_i^C=1} > 0$ gives relative preference to category 1 over category 0 such that landscapes tend to have more objects of category 1 than of category 0 for type C, provided that the energy terms of other landscape descriptors do not conversely influence the proportion of categories. Markov models with only two-level categories and terms for activity and pairwise interaction can be viewed as variants of the classical Ising model (Gallavotti (1999)) and, more generally, of autologistic regression models; see Besag (1972), Hammersley and Clifford (1971) and Section 3 of van Lieshout (2019).

Instead, we could also consider landscape descriptors providing a more global perspective on the graph. As an example, we study a global descriptor related to the notion of *connected components* of landscape elements of the same category (see also Møller and Waagepetersen (1998) which highlight Markov-like properties in this case). By definition, a connected component is a subgraph in which any two vertices are linked to each other along paths of graph

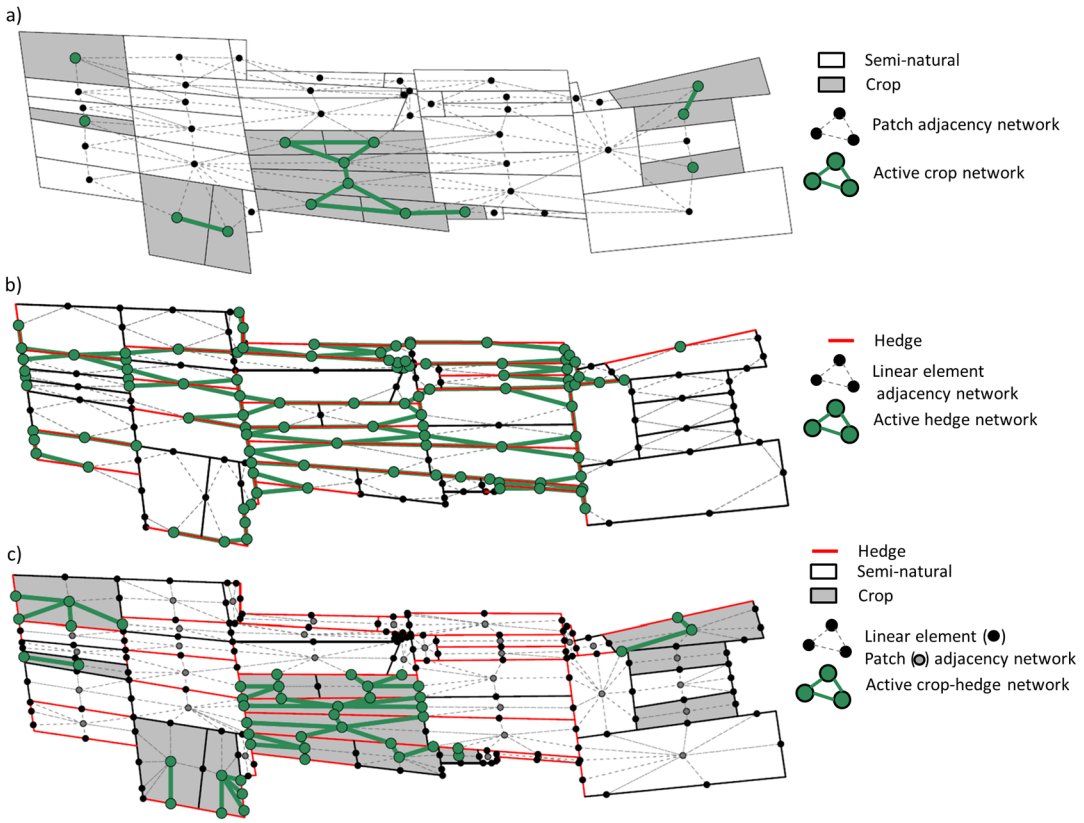

FIG. 4. *Example of landscape descriptor evaluation over a small portion of D1 for the network layer C of crop allocation (Panel (a)), the network layer H of hedge allocation (Panel (b)), the multilayer network connecting layer C and layer H (Panel (c)). The landscape is simplified in potential networks, where all connections among adjacent objects are possible, and an active network, where only the connections among allocated objects of the same type or different type are maintained, depending on their categories. The landscape descriptors for this landscape small sample for the layer C are evaluated as:* $T^{C}_{\text{act}} = 14$, $T^{C}_{\text{area},0.25} = 12$, $T^{CH}_{\text{adj}} = 168$, $T^{CC}_{\text{adj}} = 58$, *and for the layer H are evaluated as:* $T^{H}_{\text{act}} = 101$, $T^{H}_{\text{length}} = 49$, $T^{HH}_{\text{adj}} = 379$, $T^{H}_{\text{orient}} = 74$. *Formulations for computing these landscape descriptors are expressed in equation (7) and in Table 2.*

edges defined by the $\sim$-related in (2), while there are no connections to any other vertices in the complementary graph. A connected component could represent a cluster of patches or linear elements allocated with the same category. The number of connected components in a landscape conveys global information about spatial clustering of an allocation category, and it can be evaluated through dedicated algorithms (Hopcroft and Tarjan (1973)). Formally, we define the landscape descriptor T_{cluster} as the minimum possible number of sets in any partition $S_1, S_2, \ldots, S_K$ of $\boldsymbol{O}$ satisfying the following property: $o_i, o_j \in S_k$ if a path along edges between objects in S_k exists from o_i to o_j.

3.5. *Simulation examples.* Iterative simulation of Gibbs random fields with finite state spaces through Markov chain Monte Carlo techniques is, in general, relatively straightforward and stable (see, e.g., Section 3.6 of van Lieshout (2019)). We here implement the Gibbs sampler; see the Supplementary Material for details (Zamberletti et al. (2021)), where we also check MCMC convergence diagnostics such as trace plots of descriptors.

We show several simulations for the domain D1 to visually explore the influence of parameters β_k in (5); see Figure 6. We focus on three types of Markov interactions: *crop–crop adjacency* (β^{CC}_{adj}), *hedge–hedge adjacency* (β^{HH}_{adj}) and *crop–hedge adjacency* (β^{CH}_{adj}), as defined in

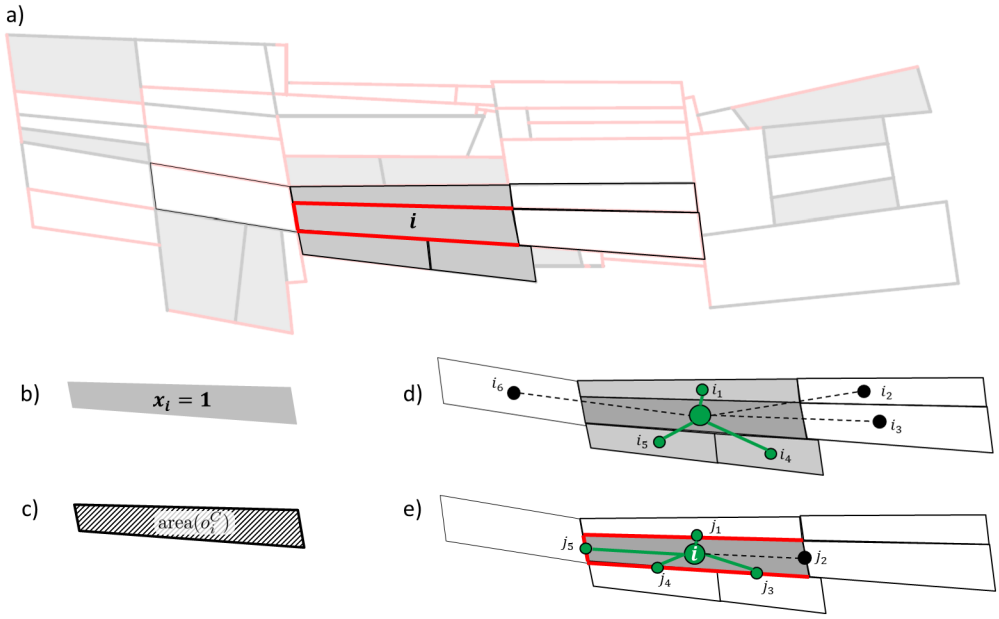

FIG. 5. *Given Figure 4, we focus on a single object* $o_i = (z_i, x_i)$, *where* $z_i = \text{coords}(o_i), \text{area}(o_i)$ *and* $x_i = crop, x_i = 1$) (*Panel* (*a*)), *with detailed landscape descriptor specification* (*Panels* (*b*)–(*e*)). *Specification for object* O_i *are evaluated as*: (*b*) $t^C_{\text{act}}(x_i) = 1$, (*c*) $t^C_{\text{area},0.25}(x_i) = 0$, (*d*) $t^{CH}_{\text{adj}}(x_i) = 4$, (*e*) $t^{CC}_{\text{adj}}(x_i) = 3$. *Formulations for the evaluation specific landscape descriptors are found in Table 2.*

TABLE 2

Examples of landscape descriptors (*top*) *and model configurations* (*bottom*). *Notations*: *C and H refer to patches and linear elements, respectively*; $\mathbb{I}$ *is the indicator function*; $\mathbb{Q}_p$ *is the* (*empirical*) *p-percent quantile* ($p \in (0, 1)$); $\mathbb{E}$ *is the* (*empirical*) *expected value. The function* angle *returns the radians angle of a linear segment in* $[-\pi/2, \pi/2)$ *with respect to the West–East direction* (*i.e., the axis* $(0, 1)^T$). *Landscape models show descriptors related to crop patches in network C, and to hedges in linear element network H*

Examples of landscape descriptors			
Composition	Activity term	$T^C_{\text{act}}, T^H_{\text{act}}$	$t(x_i^C) = \mathbb{I}(x_i^C = 1)$
	Patch area	$T^C_{\text{area,p}}$	$t(x_i^C; p) = 1(x_i^C = 1, \text{area}(o_i^C) \le \mathbb{Q}_p(\text{area}(o_i^C)))$
	Long segments	T^H_{length}	$t(x_i^H) = 1(x_i^H = 1, \text{length}(o_i^H) \ge \mathbb{E}[\text{length}(o_i^H)])$
	Horizontal segments	T^H_{orient}	$t(x_i^H) = 1(x_i^H = 1, \text{angle}(o_i^H) \in [0, \frac{\pi}{6}] \cup [\frac{5\pi}{6}, 2\pi])$
Interaction (Adjacency)	Patch-patch	T^{CC}_{adj}	$t(x_i^C, x_j^C) = a^C_{ij}$
	Segment-segment	T^{HH}_{adj}	$t(x_i^H, x_j^H) = a^H_{ij}$
	Patch-segment	T^{CH}_{adj}	$t(x_i^C, x_j^H) = a^{CH}_{ij}$

Landscape models	C	H
M1	$T^C_{\text{act}}, T^C_{\text{area},0.25}, T^C_{\text{area},0.75}, T^{CH}_{\text{adj}}, T^{CC}_{\text{adj}}$	$T^H_{\text{act}}, T^H_{\text{length}}, T^H_{\text{orient}}, T^{HH}_{\text{adj}}$
M2	cf. M1	$T^H_{\text{act}}, T^H_{\text{orient}}, T^{HH}_{\text{adj}}$
M3	$T^C_{\text{act}}, T^H_{\text{area},0.25}, T^{CH}_{\text{adj}}, T^{CC}_{\text{adj}}$	cf. M1
M4	$T^C_{\text{act}}, T^H_{\text{area},0.25}, T^C_{\text{area},0.75}, T^{CH}_{\text{adj}}, T^{\text{cluster}}_C$	cf. M1

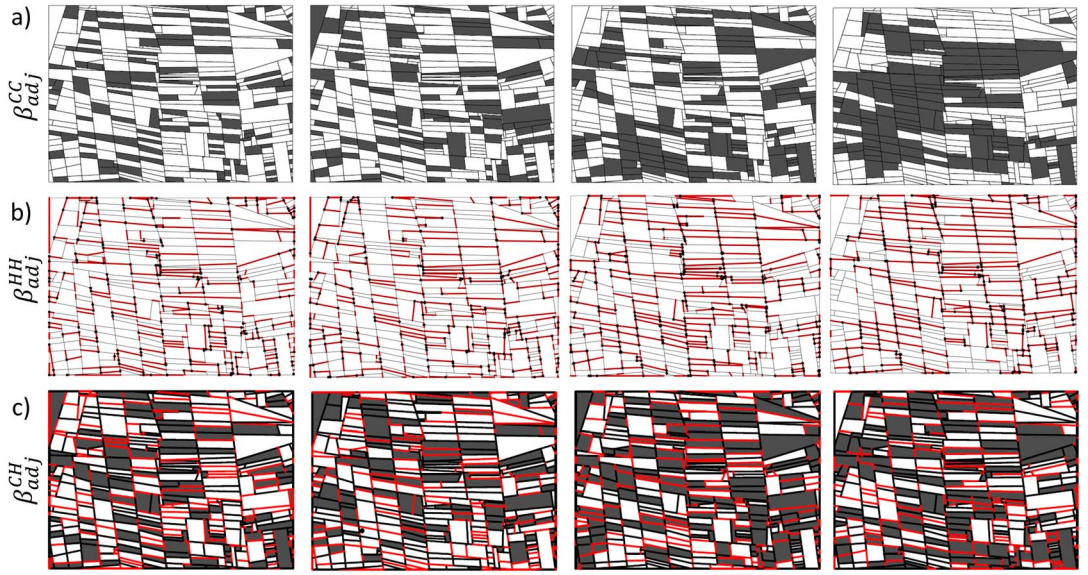

FIG. 6. *Landscape simulations on D*1. *Panel* (*a*): *Varying crop–crop adjacency*; *Panel* (*b*): *Varying hedge–hedge adjacency*; *Panel* (*c*): *Varying crop–hedge adjacency*. *Columns from left to right*: *coefficient* $-1, -0.33, 0.33, 1$.

Table 2. In each simulation run we set only one of the coefficients to a nonzero value among $\{-1, -1/3, 1/3, 1\}$; other descriptors are not controlled in the model. The MCMC simulation runs take from several seconds (D1, D2) to several minutes (D3) before approximately reaching the stationary distribution. For all simulations in this paper, we have fixed relatively large numbers of burn-in steps of $N_0 = 10^4$ (D1, D2) and of $N_0 = 10^6$ (D3) to ensure that chains always reach the stationary distribution. Negative coefficients produce fragmented allocation structures of the two corresponding categories, while a positive coefficient results in clustered configurations of categories. In Figure 6(c) a negative coefficient of the *crop–hedge adjacency* leads to many hedges being located away from crop-patch boundaries, while they tend to concentrate on such boundaries for positive coefficients.

4. Statistical inference and model validation.

4.1. *Parameter inference*. We infer the allocation mechanism of real landscapes by first estimating the parameter vector $\boldsymbol{\beta}$ of candidate models and then studying significance and other diagnostics. The likelihood function is not tractable in practice due to the normalizing constant $c(\boldsymbol{\beta})$ in the probability mass function (5). Instead, we use a pseudo-likelihood based on conditional distributions; see Besag (1972, 1974), Møller and Waagepetersen (1998), Stoehr (2017), van Lieshout (2000) and, particularly, Section 3.5 of van Lieshout (2019). Given n objects $\boldsymbol{x} = (x_1, \ldots, x_n)$ with their allocation categories, we define the pseudo-likelihood as the product of the conditional probability of the category x_i given all the other variables $\boldsymbol{x}_{-i}$; that is, it is the composite likelihood (Varin, Reid and Firth (2011)) of conditional distributions given as

$$\mathcal{L} = \prod_{i=1}^{n} p(x_i \mid \boldsymbol{x}_{-i}, \boldsymbol{z}), \tag{8}$$

where the conditional probability $p(x_i \mid \boldsymbol{x}_{-i}, \boldsymbol{z})$ is defined in equation (6) and does not depend on the normalizing constant $c(\boldsymbol{\beta})$; in Markov models it depends only on information from adjacent objects in $\boldsymbol{o}_{-i}$.

For binary $x_i \in \mathcal{X}_i = \{0, 1\}$, we write $\tilde{\boldsymbol{x}}$ for $\boldsymbol{x}$, with x_i replaced by the alternative level; then, (6) is equivalent to the logistic regression equation

$$\log \frac{p(x_i \mid \boldsymbol{x}_{-i})}{1 - p(x_i \mid \boldsymbol{x}_{-i})} = \sum_{k=1}^{m} \beta_k \big(T_k(\boldsymbol{x}) - T_k(\tilde{\boldsymbol{x}})\big). \tag{9}$$

Parameter estimation of $\boldsymbol{\beta}$ can then be carried out using standard software for logistic regression (if $\ell_i = 2$) or using the more general pseudo-likelihood framework (if $\ell_i > 2$).

The maximum pseudo-likelihood estimator $\widehat{\boldsymbol{\beta}}$ is asymptotically consistent and normal when independent replicates of the spatial process have been observed (Jensen and Møller (1991), Varin, Reid and Firth (2011)).

However, in the single-replicate setting it is difficult to obtain standard errors and confidence bounds since standard asymptotic theory for maximum likelihood and maximum pseudo-likelihood estimation is based on replicated data structures. Moreover, one should check if estimation bias arises with finite-sample data. Block-bootstrapping procedures using a spatial partition of the study area (Lahiri (2003)) could provide a solution in cases of very large study areas, especially with Markov models whose spatial dependence strength decays relatively fast for larger distances; however, such bootstrap schemes would be difficult to implement on moderately large domains, such as D1, D2 and D3 in Figure 1, and handling the multiplex networks may be awkward. Instead, we propose to resort to a parametric bootstrap, using MCMC simulation of the estimated model to generate pseudo-replicates of the observed data which then allows us to check if estimation is stable and unbiased to obtain confidence intervals and, more specifically, to check if a descriptor is significant.

We proceed as follows: generate n_{boot} independent simulations (e.g., $n_{\text{boot}} = 99$) of the fitted model using $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, and reestimate the coefficient vector for each simulation to obtain a sample of the pseudo-likelihood estimator; then, use this sample to check for estimation bias, and derive Monte-Carlo confidence intervals. For a test of the null hypothesis of $\beta_k = 0$ for fixed $k \in \{1, \ldots, m\}$, that is, to check if the landscape descriptor T_k is significant, we implement a Monte-Carlo test where we repeatedly simulate the fitted model, but with the modification $\hat{\beta}_k = 0$. Then, we here reject the null if the value $\hat{\beta}_k$ does not lie within the one-sided Monte–Carlo confidence interval of $\hat{\beta}_k$, that is, if less than $\alpha\%$ (e.g., $\alpha = 5$) of the β_k-values estimated for the simulations have the same sign and higher absolute value than the value estimated for the data (see, e.g., Davison and Hinkley (1997)).

4.2. *Pseudo-likelihood-based model selection.* We propose approaches to statistically compare models with different landscape descriptor configurations and to assess their goodness-of-fit.

As first criterion we consider the maximum pseudo-loglikelihood value, denoted by mpll. We have to rank models based on information criteria that take the model complexity (i.e., the number of parameters) into account to avoid overfitting and to identify parameter configurations that are both parsimonious and informative. Relevant information criteria, such as the composite likelihood information criterion (CLIC), are hard to calculate in a single-replicate setting such that we only use them to compare models with the same number of parameters; that is, we directly compare mpll values to rank CLICs in this case.

The second criterion, applicable to compare any types of models, is the mean squared error (MSE) based on k-fold cross-validation. In k-fold cross-validation we partition the original dataset into k groups (e.g., $k = 5$), and reestimate the model repeatedly by holding out one fold a time. For each hold-out set we then summarise the skill of the model by a prediction score, here chosen as the MSE between the predicted probability of an allocation category and its actual value. The MSE is also known as the *Brier score* (Brier (1950)), and it is a proper score function, as defined by Gneiting and Raftery (2007). We evaluate scores separately on

each hold-out set and then average the k resulting values to obtain the global score. We here propose to generate a random partition into k folds of the same size separately for each layer of the network (e.g., patches, edges).

4.3. *Model diagnostics using landscape summaries.* We propose to check if the fitted model is able to appropriately reproduce three types of summaries of the real landscape: (1) landscape descriptors used in the model (i.e., sufficient statistics); (2) variograms (using a raster representation), as defined in the following, to measure the variability of crop, hedge and crop–hedge structures with respect to Euclidean distance in space; (3) general landscape metrics commonly used in landscape ecology, based on vector or raster representations.

Type 1 concerns statistical validation: the theoretical distribution of a landscape descriptor should be in line with its observed value; we check this through Monte Carlo samples of the fitted model.

Regarding type 2 variograms (Cressie (2015), van Lieshout (2019)), we adopt a geostatistical perspective (Saura and Martinez-Millan (2000)) that focuses on the variability and geographic scales of the landscape which has already proven useful to characterize land use properties (Garrigues et al. (2006, 2008)). We here define two variogram variants. The first variant, called *one-category variogram*, explores the spatial variability of the presence $Z_c(s) \in \{0, 1\}$ (1 for present, 0 for absent) of a category c at any location s in the study domain D. The second variant, called *two-category variogram*, explores the spatial interaction of two distinct categories c_1, c_2 (e.g., crop and hedges). We define $Z_{c_1,c_2}(s) \in \{0, 1\}$ only for locations s where either c_1 or c_2 is present (other locations are considered as not being part of the domain of the process), and we set $Z_{c_1,c_2}(s) = 1$ if c_1 is present at s, otherwise the value is 0. For both variants, denoted by γ_c and γ_{c_1,c_2}, respectively, we calculate experimental variograms, assuming stationarity and isotropy, according to the empirical counterpart of $\gamma(h) = \mathbb{E}(Z(s_1) - Z(s_2))^2$, where $\|s_2 - s_1\| = h$ for distances $h \geq 0$.

Regarding type 3 of summaries, various metrics have been used to assess if simulated landscape patterns appropriately represent landscape functionality and ecological relevancy (Kupfer (2012), Frazier and Kedron (2017)); some metrics are known to be strongly correlated in practice. We assess if models of type (5), endowed with a small number of landscape descriptors, are able to generate metric values close to the observed one. Simulation of fitted models is used to generate a representative sample of the theoretical model-based distributions of metrics.

Some commonly used metrics require landscapes to be represented as a mosaic of discrete habitat patches, as in our case. Many other metrics have been developed for landscapes conceptualized as environmental gradients (i.e., for raster representations; see McGarigal and Marks (1995), Cushman et al. (2010)). Here, we assess how data patterns are reproduced by models through metrics based on graph theory (*network metrics* Minor and Urban (2008), Urban and Keitt (2001), Urban et al. (2009), Lü et al. (2016)) or *raster metrics* (McGarigal and Marks (1995)), where we transform our vector-based patch-mosaic representation into a raster; see Table 3 for a summary.

We focus on standard network metrics (Urban and Keitt (2001), Minor and Urban (2008)), evaluated either at node scale (with one value per node) or at network scale. These active networks associated to a specific allocation have one layer for each allocation category (i.e., crop network C and hedge for network H); recall Section 3.3, and there are edges if two adjacent objects ($(x_1, z_1) = o_1 \sim o_2 = (x_2, z_2)$) have been allocated the same category, that is, $x_1 = x_2$. Node scale helps to identify vital nodes associated with structural or functional objectives (Lü et al. (2016)), while network scale summarizes the global topology (Urban and Keitt (2001), Calabrese and Fagan (2004)). For metrics based on gradient theory, we follow Cushman, McGarigal and Neel (2008) and choose those metrics identified as "highly

TABLE 3

Landscape metrics. A star ⋆ indicates metrics normalized with the number of nodes. Metric type is either "node" (node-scale network metrics), "network" (global network metrics), or "raster." Listed references are as follows: [1] *Urban et al.* (2009), [2] *Lü et al.* (2016), [3] *Latora and Marchiori* (2001), [4] *McGarigal and Marks* (1995), [5] *Cushman, McGarigal and Neel* (2008)

Name	Description	Support	Range	Reference
Degree⋆	Number of connected nodes	node	[0, 1]	[1],[2]
Coreness	K-shell decomposition for a node's spreading influence	node	[0,∞)	[1],[2]
Degree grade 2⋆	Number of connected nodes at most 2 nodes away	node	[0, 1]	[1],[2]
Eccentricity⋆	Maximum shortest path to connected nodes	node	[0, 1]	[1],[2]
Closeness	Reciprocal of total length of shortest paths to connected nodes	node	[0,∞)	[1],[2]
Betweenness⋆	Potential power to control information flow	node	[0, 1]	[1],[2]
Diameter	Longest path	network	[0,∞)	[1]
Efficiency	Efficiency of information exchange	network	[0,∞)	[2],[3]
Cluster avg.	Proportion of interconnected adjacent nodes of a vertex	network	[0,1]	[2],[3]
PLAND [%]	Percentage of a habitat in the landscape	raster	[0,100]	[4],[5]
PD [# / ha × 100]	Patch density	raster	[0,∞)	[4],[5]
ENN [m]	Mean Euclidean nearest neighbor distance	raster	[0,∞)	[4],[5]
PARA [/]	Perimeter-area ratio of contiguous habitat	raster	[0,∞)	[4],[5]
IJI [%]	Interspersion/juxtaposition index measuring spatial intermixing of different habitats	raster	[0,100]	[4],[5]
CLUMPY [/]	Clumpiness index measuring deviation from randomness	raster	[-1,1]	[4],[5]

universal and consistent class-level landscape structure components." We use the `R` package `raster` (Hijmans et al. (2015)) to transform vector objects (i.e., polygons and linear segments) into rasters with categorical values and the package `landscapemetrics` to evaluate raster metrics (Hesselbarth et al. (2019)). In the case of two polygon types (crops/hedges) and two edge types (presence/absence of hedge), we obtain *three* pixel categories in the raster, also called habitats: crop, seminatural and hedge; absence of hedges is not a class in itself.

5. Application to the Lower Durance Valley in southern France. We fit parametric stochastic models for the category allocation mechanism of crops and hedges in the domains D1, D2 and D3 using the logistic regression equation (9), and we discuss descriptor selection by assessing a moderate number of landscape descriptors.

5.1. *Structure of descriptors and models*. We allow for two allocation categories of both patches and linear elements: *crop* or *seminatural area* (network C); presence or absence of a hedgerow for (network H).

We consider four models for (5), denoted M1–M4 and summarized in Table 2, to test different combinations of landscape descriptors in the general model (5). To avoid collinearity of descriptors, we first check correlations between the covariates arising in the logistic regression (9) for each spatial domain; see Figure 5 in the Supplementary Material (Zamberletti et al. (2021)). Strong negative correlation is observed between T^{CC}_{adj} (crop adjacency) and T^{C}_{cluster} (number of connected crop components) such that we avoid including both of them in the same model, and we seek to assess which of the two descriptors better captures crop-to-crop interaction (M1/M3 vs. M4).

The patch area distribution shows high variance, and we include descriptors for the effect of patch area in network C. While the behavior of large patches can strongly influence the proportions of crop and seminatural habitat, small field sizes may benefit biodiversity through

TABLE 4

Values of mpll *and MSE in five-fold cross-validation for each combination of model* (*M*1–*M*4, *Crop/Hedge network*) *and spatial domain* (*D*1–*D*3). *Highest* mpll *and lowest MSE are highlighted in bold for each domain and Crop/Hedge*

		Landscape descriptors	D1	D2	D3
C	M1	$T^C_{\text{act}}, T^C_{\text{area,0.25}}, T^C_{\text{area,0.75}}, T^{CH}_{\text{adj}}, T^{CC}_{\text{adj}}$	**−198.9**, **0.184**	**−236.4**, **0.171**	**−1778**, **0.130**
	M3	$T^C_{\text{act}}, T^C_{\text{area,0.25}}, T^{CH}_{\text{adj}}, T^{CC}_{\text{adj}}$	−205.6, 0.191	−239.5, 0.173	−1782, **0.130**
	M4	$T^C_{\text{act}}, T^C_{\text{area,0.25}}, T^C_{\text{area,0.75}}, T^{CH}_{\text{adj}}, T^C_{\text{cluster}}$	−202.1, 0.191	−243.5, 0.182	−1970, 0.151
H	M1	$T^H_{\text{act}}, T^H_{\text{length}}, T^H_{\text{orient}}, T^{HH}_{\text{adj}}$	**−608.5**, 0.186	**−633.3**, **0.143**	**−6399**, **0.169**
	M2	$T^H_{\text{act}}, T^H_{\text{orient}}, T^{HH}_{\text{adj}}$	−608.6, **0.185**	−640.7, 0.144	−6405, **0.169**

easier access to adjacent fields with complementary resources (Sirami et al. (2019)). Therefore, we use a patch area condition using $T^C_{\text{area,p}}$ in Table 2 with $p = 0.25$ and $p = 0.75$ (M1), and we check redundancy by removing $T^H_{\text{area,0.75}}$ in M3.

For hedges, Figure 5 in the Supplementary Material shows strong positive correlation between T^H_{length} (long hedges, where we count the number of hedges positioned on edges longer than the average edge length) and T^H_{orient} (horizontal hedges) which is related to the windbreak function of many hedges against strong Mistral winds blowing from the north; to check redundancy of these two descriptors, we include only T^H_{length} in M2, in contrast to M1–M3 (Zamberletti et al. (2021)).

We present a detailed analysis of model M1 in D1, denoted as M1–D1, and we point out some salient results of the comparison of M1, M3 and M4, and of other domains D2 and D3. Detailed results can be found in the Supplementary Material (Zamberletti et al. (2021)).

5.2. *Likelihood-based model comparison*. Table 4 reports mpll values, and mean-squared errors (MSE) obtained through *five*-fold cross-validation; recall Section 4.2. In the network H the hedge length descriptor T^H_{length}, included in M1 but not M2, does not provide notable improvements, except for the domain D2 where the correlation between the logistic regression covariates related to T^H_{length} and T^H_{orient} is relatively weaker. In the network C the comparison of the crop models in M1, M3 and M4 reveals that M1 consistently performs best for both criteria; that is, explicit control over large patches is required, and the Markov interaction model based on adjacency provides better results than direct control over the number of connected crop components.

5.3. *Estimated parameters*. In Table 5, we report coefficient estimates of β_k as well as standard errors and significance with respect to the null $\beta_k = 0$, based on Monte–Carlo procedures using 100 simulations; see Section 4.1. Figure 7 in the Supplementary Material shows boxplots of the parametric bootstrap estimations, and we detect no bias in the estimators (Zamberletti et al. (2021)). All estimated parameters are significant for the Markov interaction in the networks C and H (positive coefficient of C–C, H–H), for the area descriptor (negative coefficient of *Small area* and of *Large area*), for the hedge orientation descriptor (positive coefficient of *Horizontal H*) and for the activity terms. No strong signal is found for a dominance of long hedge segments (*Long H*) in D1, confirming results in Section 5.2 and of Markov interaction between C and H. All descriptors are significant for the large domain D3. The signs of estimates are the same across D1–D3 for all significant effects, implying structurally similar behavior. The different sign in M4 for C–C interaction is due to different specification as a global descriptor using the number of connected components. Overall, estimated parameters tend to have comparable magnitudes across D1–D3.

TABLE 5

Parameter estimates for crop-related (C) and hedge-related (H) descriptors. Crop-hedge interaction (C–H), Crop-crop interaction (C–C). "SD" values and significance (at the 95% level, indicated through bold face) are based on 100 parametric bootstrap simulations. C–C is of Markov type (using T^{CC}_{adj}) in M1/M3 and is global (using T^{C}_{cluster}) in M4

		Crop					Hedge			
		Activity	Small area	Large area	C–H	C–C	Activity	Long H	H–H	Horizontal H
M1–D1	Estimate	**−1.01**	**−1.52**	**−1.13**	0.03	**0.40**	**−2.38**	−0.10	**0.77**	**1.58**
	SD	0.34	0.35	0.29	0.07	0.10	0.19	0.19	0.07	0.19
M3-D1	Estimate	**−1.08**	**−1.24**	–	−0.05	**0.37**	**−2.38**	−0.10	**0.77**	**1.58**
	SD	0.37	0.34	–	0.09	0.10	0.19	0.19	0.07	0.19
M4–D1	Estimate	−0.16	**−1.80**	**−1.15**	0.06	**−0.99**	**−2.38**	−0.10	**0.77**	**1.58**
	SD	0.25	0.39	0.32	0.08	0.47	0.19	0.19	0.07	0.19
M1–D2	Estimate	**−1.58**	**−1.13**	**−0.68**	−0.04	**0.65**	**−3.38**	**−0.58**	**0.77**	**3.65**
	SD	0.32	0.32	0.28	0.07	0.08	0.23	0.17	0.08	0.2
M1–D3	Estimate	**−1.82**	**−1.44**	**−0.25**	**−0.14**	**0.66**	**−3.01**	**−0.16**	**0.95**	**1.96**
	SD	0.09	0.14	0.07	0.02	0.02	0.06	0.05	0.02	0.05

We interpret these results as follows: given the parameter estimates of our model, the crop category is usually not allocated on relatively small and relatively large fields; later results will confirm the superior performance of M1 over M3/M4, the latter without the large area descriptor. Crop fields and hedges tend to cluster in space, that is, they tend to be allocated on adjacent patches and linear elements, respectively, such that they provide relatively large and contiguous habitats and relatively long continuous movement corridors. Moreover, the connected component descriptor in M4 has negative coefficients, that is, the study domains seem to be characterised by relatively few large crop clusters. There is a dominating horizontal orientation of hedges for protecting against strong winds. Crop-hedge adjacency has negative coefficients and is significant only for the large domain M1–D3, suggesting a slight tendency of hedges to not being directly adjacent to crop fields. In M1–D2 we discern a particularly strong signal of *Long H*, indicating many short, strongly horizontally oriented hedges.

5.4. *Summary diagnostics of observed and simulated landscape.* We check if landscape descriptors as well as one- and two-category variograms and graph- and raster-based metrics, introduced in Section 4.3, are appropriately reproduced by the models for crop allocation to patches in D1. We focus on M1 which was found to show good relative performance in the preceding diagnostics. We generate 100 independent simulations of the fitted model. For scalar metrics and for fixed distances in the variogram analysis, we report results for approximate two-sided Monte-Carlo test procedures at 95% confidence level with respect to the null hypothesis that the observed summary could have been generated by the fitted model. Results for the hedge network and for other domains (M1–D2, M1–D3) are structurally similar; they are reported in the Supplementary Material (Zamberletti et al. (2021)).

5.4.1. *Landscape descriptors.* Figure 7 shows observed and simulated landscape descriptors, that is, sufficient statistics for the estimated coeffients. Models M1, M3, M4 tend to produce realistic values, especially M1.

5.4.2. *Variogram analysis.* Figure 8 shows empirical one-category (Crop) and two-category (Crop-Hedge) variograms with pointwise simulation envelopes. All variograms show a relatively steep slope at the origin and tend to flatten for larger distances such that the

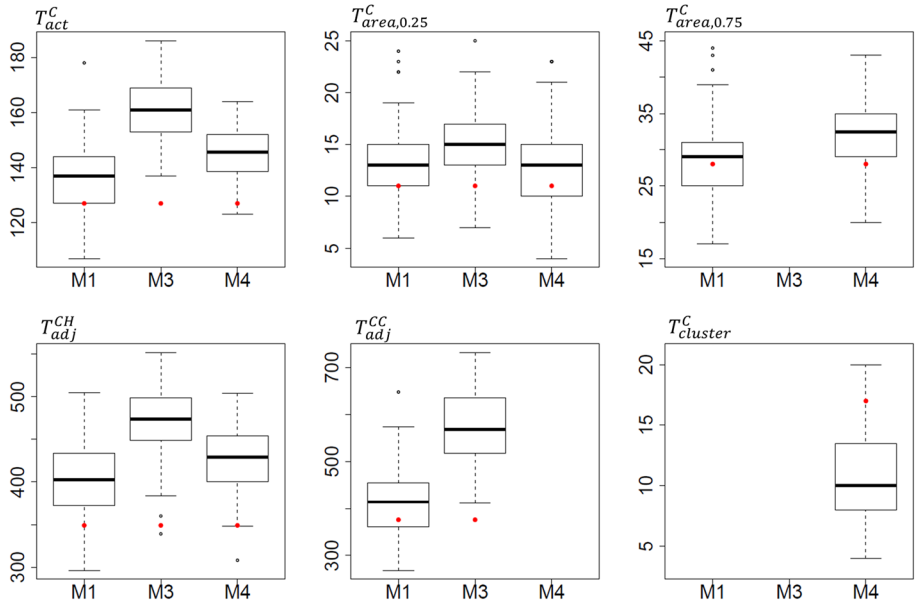

FIG. 7. *Landscape descriptors for domain D*1 *and network C* (*Crop*) *in M*1, *M*3, *M*4. *Boxplots summarize* 100 *simulations of fitted models. Red dots are observed values.*

general shape of the empirical data variogram is well reproduced by the models. In several cases, especially with M3, empirical variograms of the dataset clearly fall outside the envelope, such that the observed variability of landscape features with distance is not appropriately captured. In general, the structure of M1 (with the large patch area descriptor, and Markov interaction for crops) improves the match between data and model variograms—in contrast to M4 using the global interaction descriptor based on the number of connected components. One-category variograms for hedges are appropriately captured by models (Supplementary Material (Zamberletti et al. (2021))).

5.4.3. *Network metrics*. Figure 9 shows observed and simulated network metrics. For node-scale metrics (two top rows), we observe good overlap of boxplots of observed and simulated values for the crop and hedgerow network, with the exception of *Betweenness* (number of the shortest paths going through a node when connecting any two other nodes), where we tend to simulate too large values for crop and too small values for hedges. Some outlying values are not shown since *Betweenness* is very heavy-tailed, due to high variability among different networks which may explain the mismatch between observed and simulated values. Heavy-tailedness highlights that few crop patches serve as bridges connecting different crop clusters which is fundamental to global connectivity of the landscape (Belgrano, Woodward and Jacob (2015), Estrada and Bodin (2008), Urban et al. (2009)); this property is preserved in the fitted model.

For network-scale metrics (last row of Figure 9), we show the real landscape value within the boxplot of simulated values. Observed metrics fall within or close to the interquartile range of the simulated ones for the crop network, while they lie outside the boxplot whiskers for the hedge network but are still of the same order of magnitude. Moreover, we define and report so-called pseudo-p-values in Section 7 of the Supplementary Material to allow for automatic screening of network and raster metrics (Zamberletti et al. (2021)).

Model M1 does not directly control the number or dimension of clusters, only local interactions through the Markov model. This explains better performance for *neighborhood-based centralities* in comparison to *path-based centralities* and metrics. However, using a global

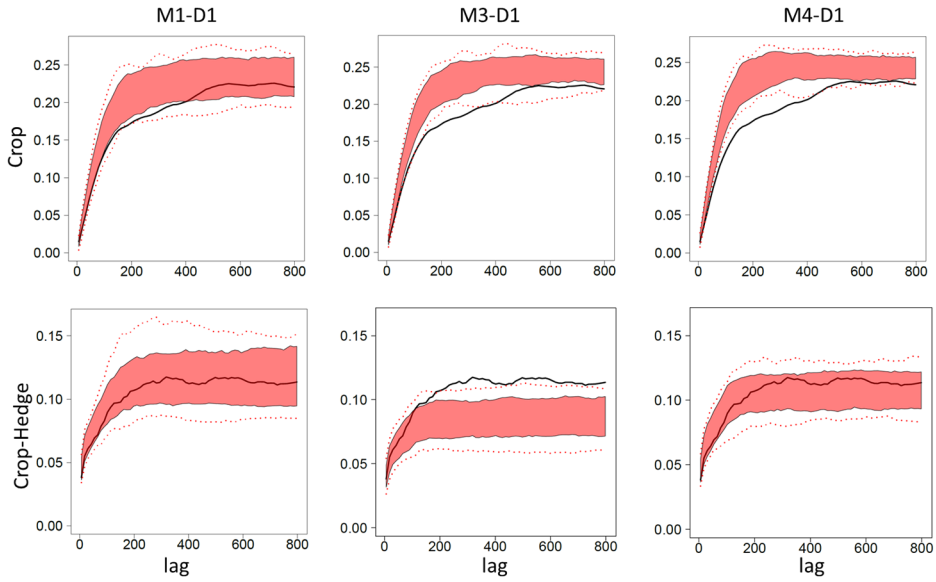

FIG. 8. *Variogram analysis of models M*1, *M*3, *M*4 *for domain D*1. *One-category variogram for crop* (*top row*); *two-category variogram for crop and hedges* (*bottom row*). *Empirical variogram of observed landscape* (*black line*); *pointwise simulation envelopes* (*red-shaded area*: 5%-95%; *dotted red lines*: *minimum/maximum*).

descriptor instead of a local descriptor in M4 does not substantially improve performance for path-based centralities; see Section 7 in the Supplementary Material (Zamberletti et al. (2021)).

5.4.4. *Raster metrics*. Figure 10 shows the raster-based landscape metrics of FRAGSTAT; see Section 4.3 and Table 3.

In most cases, observed metric fall within the whiskers of the boxplots, and in the other cases the order of magnitude is still relatively well captured by the fitted model.

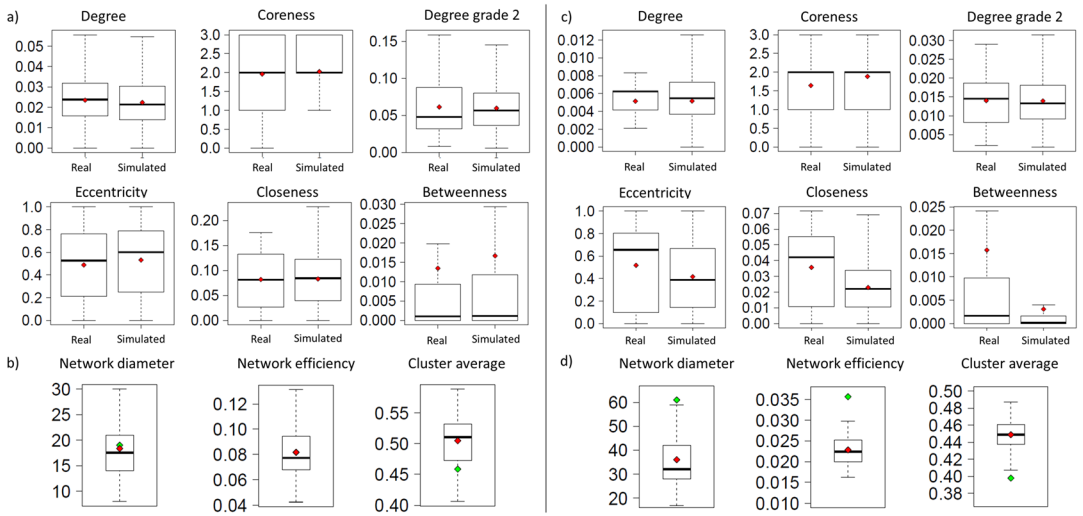

FIG. 9. *Validation metrics M*1–*D*1 *for crop network C* (*left*) *and hedge network H* (*right*). *Panels* (*a*), (*c*): *Metrics at node scale* (*red dots*: *mean values*). *Panels* (*b*), (*d*): *Metrics at network scale* (*boxplots*: *simulations*; *red dots*: *mean values of simulations*; *green dots*: *observed values*).

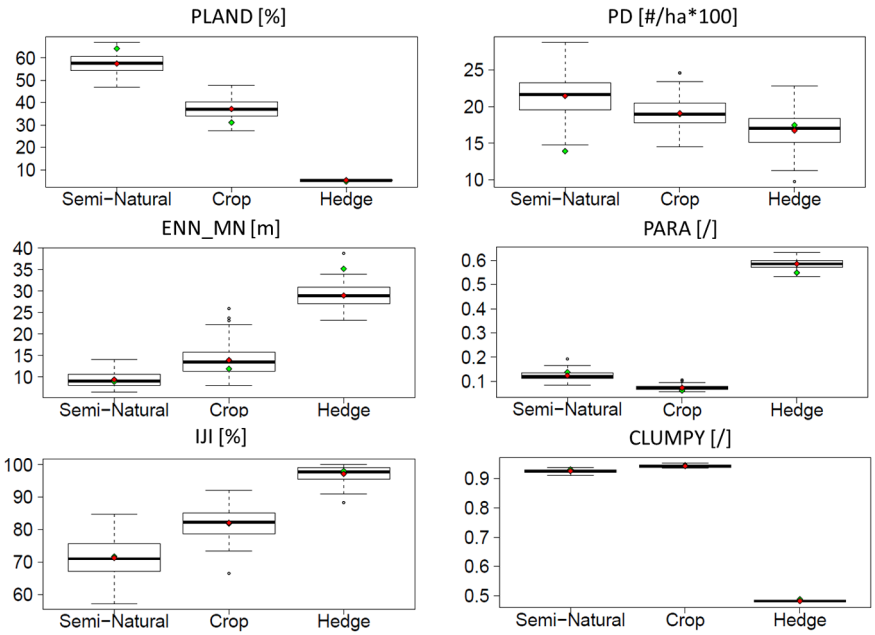

FIG. 10. *Raster-based metrics for M*1–*D*1. *Simulated values* (*boxplots*); *mean of simulated values* (*red dots*); *observed value* (*green dots*).

5.4.5. *Correlation analysis of landscape summaries.* Different landscape summaries (descriptors, metrics) may comprise similar information, and strong correlation may arise among such variables. If we seek a realistic representation of a specific metric through the model, the landscape descriptors included in the model (or combinations of them) should be strongly correlated with this metric. To assess such relationships, that is, if the model may allow us to target specific values of metrics of interest, we generate a sample of size 100 of the model through the MCMC approach from Section 3.5. We then use linear regression, with the landscape descriptors as predictors and one landscape metric at a time as dependent variable, and then consider the part of the standard deviation of the response not explained by the predictors. For illustration, we analyse differences among the models M1 and M3 (where M1 has an additional descriptor $T^C_{\text{area},0.75}$ related to the allocation of large patches with crop) for domain D1 through the correlation analysis in Figure 11(a). The descriptor $T^C_{\text{area},0.75}$ is generally more strongly correlated with other metrics for crop patches than $T^C_{\text{area},0.25}$, which tends to substantially reduce residual standard deviation not explained by the descriptors of the model, as shown in Figure 11(b).

6. Conclusion. We have developed stochastic agricultural landscape models and statistical inference with a focus on the land-use allocation mechanism of patches and linear elements, using network models as an intuitive and flexible tool for direct control and interpretation with respect to local behavior. We have focused on descriptors based on single objects or pairwise Markov interactions, which leads to robust modeling, estimation and simulation procedures, while we found it generally difficult to improve models by the use of more globally specified interaction descriptors. Overall, the descriptors of the model and other landscape metrics were satisfactorily reproduced by simulations from models (especially M1) fitted to Lower Durance Valley data. We highlighted the flexibility of the approach by comparing outcomes of different models over the same domain, and we also tested models over domains having different and relatively large size. The generality of calibrated models was evaluated

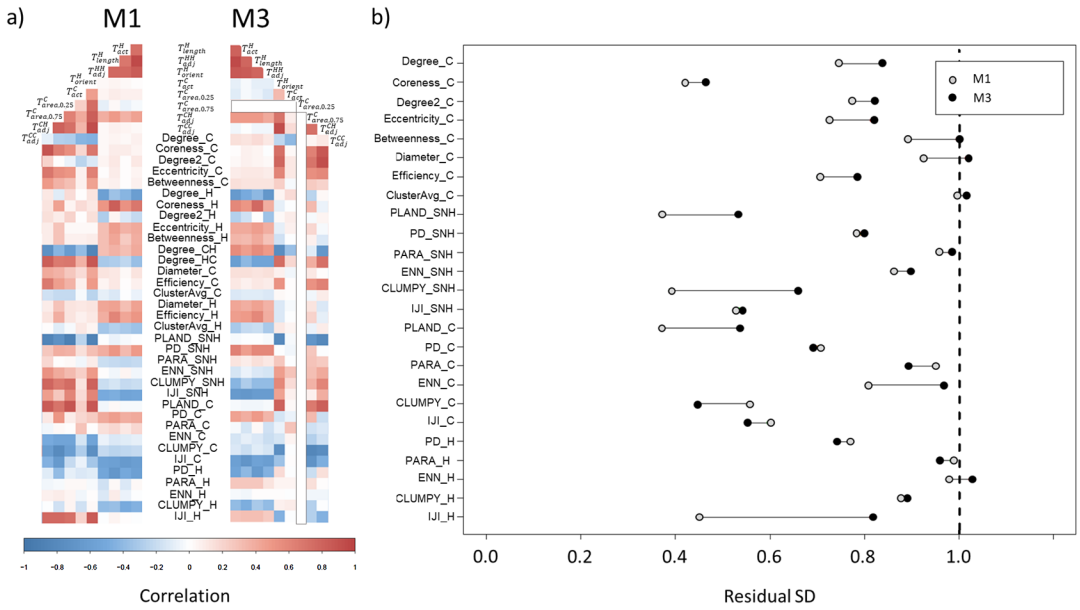

FIG. 11. *Correlation analysis for M*1–*D*1, *M*3–*D*1. *Panel* (*a*): *Correlations among landscape descriptors and metrics* (*notation*: *C*—*patch network*; *H*—*linear element network*; *CH*—*patch to linear element connections*; *SNH*—*seminatural habitat* (*raster*)). *Panel* (*b*): *Comparison of patch-related metrics between M*1–*D*1 *and M*3–*D*1 *based on residual standard deviation*.

through variograms and metrics whose values are not explicitly encoded into the model structure. Time dynamics, such as crop rotation, cannot be estimated for the dataset due to lack of dynamic land-cover allocation data. The integration of temporal descriptors, as illustrated through the simulations in the Supplementary Material, would be an interesting perspective for future development of such classes of models (Zamberletti et al. (2021)). The proposed model class succeeds in capturing key patterns of configuration and composition in real landscapes.

The developed approach focuses on the task of classification, that is, of attributing to each landscape element one class among a finite number of possible classes. The use of Gibbs energies could be extended to more general numeric labels (e.g., continuous variables) associated with landscape elements, for instance, the crop yield in a field or the proportions of a crop field used for specific crop types when several crops are planted in the same field in some small-scale-alternating way. Then, the proposed approach could be extended to more general models of exponential family type (e.g., Brown (1986)).

We have provided a set of diagnostic and inferential tools to assess model performance from different perspectives and select an appropriate candidate. Not all relevant metrics can be reproduced through our model without bias, especially on raster scale where the grid discretization of space may produce instabilities in treating small-scale small-area patterns, especially those related to linear segments. Linear element allocation also showed some discrepancy between model and data for large-scale clustering properties. To remedy the issue of appropriately simulating an important landscape summary that is not directly controlled by the model, we can add constraints during simulation, using techniques such as simulated annealing (e.g., Papaïx et al. (2014)).

We outline the potential of approximate Bayesian computation (ABC) for parameter estimation and likelihood-free model selection using Bayes factors (e.g., Grelaud et al. (2009)). Using landscape descriptors for the ABC target summaries yields asymptotically consistent estimators under mild conditions since descriptors are sufficient statistics. However, rather long computation times may arise with this method.

Vector-based models such as ours are more parsimonious and meaningful from an ecological point of view (Gaucherel et al. (2012), Bonhomme et al. (2017)), and they enable explicitly handling different spatial and temporal scales. In raster-based approaches an appropriate representation of small-surface elements, such as hedges, would require a very high and unwieldy resolution, where a homogeneous large-surface patch would be made up of a very large number of pixels, instead of a single geometric object in our model. Our multiplex network structure assures low computational cost and memory requirements.

We do not directly model human action in the temporal dynamics of agricultural environments (Bonhomme et al. (2017), Poggi et al. (2018)), for which we would have to couple our model with a decision tool. Future developments also comprise the integration of our allocation model with (existing) generative parametric tessellation models for the geometrical support (Kiêu et al. (2013)).

Acknowledgments. We are thankful to Claire Lavigne and Katarzyna Adamczyk for help and their wise suggestions for landscape data processing and for the discussion part. We are grateful for the comments and suggestions of two anomymous reviewers and the Associate Editor that helped to significantly improve the quality of the manuscript.

SUPPLEMENTARY MATERIAL

Technical report with extended results (DOI: 10.1214/21-AOAS1447SUPP; .pdf). Additional details for the simulation algorithm and results for the case study in Section 5.

REFERENCES

ADAMCZYK-CHAUVAT, K., KASSA, M., KIÊU, K., PAPAÏX, J. and STOICA, R. S. (2020). Gibbsian t-tessellation model for agricultural landscape characterization. Preprint. Available at arXiv:2007.16094.

BADDELEY, A. and MØLLER, J. (1989). Nearest-neighbour Markov point processes and random sets. *International Statistical Review/Revue Internationale de Statistique* 89–121.

BELGRANO, A., WOODWARD, G. and JACOB, U. (2015). *Aquatic Functional Biodiversity*: *An Ecological and Evolutionary Perspective*. Academic Press, San Diego, CA.

BESAG, J. E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *J. Roy. Statist. Soc. Ser. B* **34** 75–83. MR0323276

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BOCCALETTI, S., BIANCONI, G., CRIADO, R., DEL GENIO, C. I., GÓMEZ-GARDEÑES, J., ROMANCE, M., SENDIÑA-NADAL, I., WANG, Z. and ZANIN, M. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* **544** 1–122. MR3270140 https://doi.org/10.1016/j.physrep.2014.07.001

BONHOMME, V., CASTETS, M., IBANEZ, T., GÉRAUX, H., HÉLY, C. and GAUCHEREL, C. (2017). Configurational changes of patchy landscapes dynamics. *Ecol. Model.* **363** 1–7.

BOOTS, B., OKABE, A. and SUGIHARA, K. (1999). Spatial tessellations. *Geographical Information Systems* **1** 503–526.

BRIER, G. W. (1950). Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **78** 1–3.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. MR0882001

BÜTTNER, G. and MAUCHA, G. (2006). *The Thematic Accuracy of CORINE Land Cover* 2000. *Assessment Using LUCAS* (*Land Use/Cover Area Frame Statistical Survey*). European Environment Agency, Copenhagen.

CALABRESE, J. M. and FAGAN, W. F. (2004). A comparison-shopper's guide to connectivity metrics. *Front. Ecol. Environ.* **2** 529–536.

CRESSIE, N. A. C. (1991). *Statistics for Spatial Data*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1127423

CRESSIE, N. A. C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. Wiley, New York. MR3559472

CUSHMAN, S. A., MCGARIGAL, K. and NEEL, M. C. (2008). Parsimony in landscape metrics: Strength, universality, and consistency. *Ecol. Indicators* **8** 691–703.

CUSHMAN, S. A., GUTZWEILER, K., EVANS, J. S. and MCGARIGAL, K. (2010). The gradient paradigm: A conceptual and analytical framework for landscape ecology. In *Spatial Complexity, Informatics, and Wildlife Conservation* 83–108. Springer, New York.

DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. MR1478673 https://doi.org/10.1017/CBO9780511802843

ESTRADA, E. and BODIN, Ö. (2008). Using network centrality measures to manage landscape connectivity. *Ecol. Appl.* **18** 1810–1825.

FIENBERG, S. E. (2010). Introduction to papers on the modeling and analysis of network data. *Ann. Appl. Stat.* **4** 1–4. MR2758081 https://doi.org/10.1214/10-AOAS346

FORESIGHT, U. (2011). *The Future of Food and Farming*. *Final Project Report*. The Government Office for Science, London.

FRAZIER, A. E. and KEDRON, P. (2017). Landscape metrics: Past progress and future directions. *Current Landscape Ecology Reports* **2** 63–72.

GAETAN, C. and GUYON, X. (2010). *Spatial Statistics and Modeling*. *Springer Series in Statistics*. Springer, New York. MR2569034 https://doi.org/10.1007/978-0-387-92257-7

GALLAVOTTI, G. (1999). *Statistical Mechanics*: *A Short Treatise*. *Texts and Monographs in Physics*. Springer, Berlin. MR1707309 https://doi.org/10.1007/978-3-662-03952-6

GARDNER, R. H. (1999). RULE: Map generation and a spatial analysis program. In *Landscape Ecological Analysis* 280–303. Springer, New York.

GARDNER, R. H. and URBAN, D. L. (2007). Neutral models for testing landscape hypotheses. *Landsc. Ecol.* **22** 15–29.

GARDNER, R. H., MILNE, B. T., TURNEI, M. G. and O'NEILL, R. V. (1987). Neutral models for the analysis of broad-scale landscape pattern. *Landsc. Ecol.* **1** 19–28.

GARRIGUES, S., ALLARD, D., BARET, F. and WEISS, M. (2006). Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sens. Environ.* **103** 81–96.

GARRIGUES, S., ALLARD, D., BARET, F. and MORISETTE, J. (2008). Multivariate quantification of landscape spatial heterogeneity using variogram models. *Remote Sens. Environ.* **112** 216–230.

GAUCHEREL, C., FLEURY, D., AUCLAIR, D. and DREYFUS, P. (2006a). Neutral models for patchy landscapes. *Ecol. Model.* **197** 159–170.

GAUCHEREL, C., GIBOIRE, N., VIAUD, V., HOUET, T., BAUDRY, J. and BUREL, F. (2006b). A domain-specific language for patchy landscape modelling: The Brittany agricultural mosaic as a case study. *Ecol. Model.* **194** 233–243.

GAUCHEREL, C., BOUDON, F., HOUET, T., CASTETS, M. and GODIN, C. (2012). Understanding patchy landscape dynamics: Towards a landscape language. *PLoS ONE* **7** e46064. https://doi.org/10.1371/journal.pone.0046064

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 https://doi.org/10.1198/016214506000001437

GREEN, P. J., HJORT, N. L. and RICHARDSON, S. (2003). *Highly Structured Stochastic Systems*, *Volume* 27. Oxford Univ. Press, Oxford.

GRELAUD, A., ROBERT, C. P., MARIN, J.-M., RODOLPHE, F. and TALY, J.-F. (2009). ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Anal.* **4** 317–335. MR2507366 https://doi.org/10.1214/09-BA412

HAMMERSLEY, J. M. and CLIFFORD, P. (1971). Markov fields on finite graphs and lattices. 46. Unpublished manuscript.

HESSELBARTH, M. H., SCIAINI, M., WITH, K. A., WIEGAND, K. and NOWOSAD, J. (2019). Landscapemetrics: An open-source R tool to calculate landscape metrics. *Ecography* **42** 1648–1657.

HIJMANS, R. J., VAN ETTEN, J., CHENG, J., MATTIUZZI, M., SUMNER, M., GREENBERG, J. A., LAMIGUEIRO, O. P., BEVAN, A., RACINE, E. B. et al. (2015). Package 'raster'. R package.

HOPCROFT, J. and TARJAN, R. (1973). Algorithm 447: Efficient algorithms for graph manipulation. *Commun. ACM* **16** 372–378.

INKOOM, J. N., FRANK, S., GREVE, K. and FÜRST, C. (2017). Designing neutral landscapes for data scarce regions in West Africa. *Ecol. Inform.* **42** 1–13.

JENSEN, J. L. and MØLLER, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **1** 445–461. MR1111528

KIÊU, K., ADAMCZYK-CHAUVAT, K., MONOD, H. and STOICA, R. S. (2013). A completely random T-tessellation model and Gibbsian extensions. *Spat. Stat.* **6** 118–138.

KIVELÄ, M., ARENAS, A., BARTHELEMY, M., GLEESON, J. P., MORENO, Y. and PORTER, M. A. (2014). Multilayer networks. *J. Complex Netw.* **2** 203–271.

KUPFER, J. A. (2012). Landscape ecology and biogeography: Rethinking landscape metrics in a post-FRAGSTATS landscape. *Progress in Physical Geography* **36** 400–420.

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. Springer, New York. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

LANGHAMMER, M., THOBER, J., LANGE, M., FRANK, K. and GRIMM, V. (2019). Agricultural landscape generators for simulation models: A review of existing solutions and an outline of future directions. *Ecol. Model.* **393** 135–151.

LATORA, V. and MARCHIORI, M. (2001). Efficient behavior of small-world networks. *Phys. Rev. Lett.* **87** 198701.

LE BER, F., LAVIGNE, C., ADAMCZYK, K., ANGEVIN, F., COLBACH, N., MARI, J.-F. and MONOD, H. (2009). Neutral modelling of agricultural landscapes by tessellation methods—application for gene flow simulation. *Ecol. Model.* **220** 3536–3545.

LEFEBVRE, M., FRANCK, P., TOUBON, J.-F., BOUVIER, J.-C. and LAVIGNE, C. (2016). The impact of landscape composition on the occurrence of a canopy dwelling spider depends on orchard management. *Agriculture, Ecosystems & Environment* **215** 20–29.

LIN, Y., DENG, X., LI, X. and MA, E. (2014). Comparison of multinomial logistic regression and logistic regression: Which is more efficient in allocating land use? *Front. Earth Sci.* **8** 512–523.

LÜ, L., CHEN, D., REN, X.-L., ZHANG, Q.-M., ZHANG, Y.-C. and ZHOU, T. (2016). Vital nodes identification in complex networks. *Phys. Rep.* **650** 1–63. MR3543857 https://doi.org/10.1016/j.physrep.2016.06.007

MAALOULY, M., FRANCK, P., BOUVIER, J.-C., TOUBON, J.-F. and LAVIGNE, C. (2013). Codling moth parasitism is affected by semi-natural habitats and agricultural practices at orchard and landscape levels. *Agriculture, Ecosystems & Environment* **169** 33–42.

MARTIN, E. A., DAINESE, M., CLOUGH, Y., BÁLDI, A., BOMMARCO, R., GAGIC, V., GARRATT, M. P. D., HOLZSCHUH, A., KLEIJN, D. et al. (2019). The interplay of landscape composition and configuration: New pathways to manage functional biodiversity and agroecosystem services across Europe. *Ecol. Lett.* **22** 1083–1094. https://doi.org/10.1111/ele.13265

MCGARIGAL, K. and MARKS, B. J. (1995). FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-GTR-351. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 122 p., 351.

MINOR, E. S. and URBAN, D. L. (2008). A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conserv. Biol.* **22** 297–307.

MØLLER, J. and WAAGEPETERSEN, R. P. (1998). Markov connected component fields. *Adv. in Appl. Probab.* **30** 1–35. MR1618872 https://doi.org/10.1239/aap/1035227989

PAPAÏX, J., ADAMCZYK-CHAUVAT, K., BOUVIER, A., KIÊU, K., TOUZEAU, S., LANNOU, C. and MONOD, H. (2014). Pathogen population dynamics in agricultural landscapes: The ddal modelling framework. *Infect. Genet. Evol.* **27** 509–520.

POGGI, S., PAPAÏX, J., LAVIGNE, C., ANGEVIN, F., LE BER, F., PARISEY, N., RICCI, B., VINATIER, F. and WOHLFAHRT, J. (2018). Issues and challenges in landscape models for agriculture: From the representation of agroecosystems to the design of management strategies. *Landsc. Ecol.* **33** 1679–1690.

POWER, A. G. (2010). Ecosystem services and agriculture: Tradeoffs and synergies. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **365** 2959–2971.

RICCI, B., FRANCK, P., TOUBON, J.-F., BOUVIER, J.-C., SAUPHANOR, B. and LAVIGNE, C. (2009). The influence of landscape on insect pest dynamics: A case study in southeastern France. *Landsc. Ecol.* **24** 337–349.

SAURA, S. and MARTINEZ-MILLAN, J. (2000). Landscape patterns simulation with a modified random clusters method. *Landsc. Ecol.* **15** 661–678.

SCIAINI, M., FRITSCH, M., SCHERER, C. and SIMPKINS, C. E. (2018). NLMR and landscapetools: An integrated environment for simulating and modifying neutral landscape models in R. *Methods Ecol. Evol.* **9** 2240–2248.

SIRAMI, C., GROSS, N., BAILLOD, A. B., BERTRAND, C., CARRIÉ, R., HASS, A., HENCKEL, L., MIGUET, P., VUILLOT, C. et al. (2019). Increasing crop heterogeneity enhances multitrophic diversity across agricultural regions. *Proc. Natl. Acad. Sci. USA* **116** 16442–16447. https://doi.org/10.1073/pnas.1906419116

STOEHR, J. (2017). A review on statistical inference methods for discrete Markov random fields. Preprint. Available at arXiv:1704.03331.

URBAN, D. and KEITT, T. (2001). Landscape connectivity: A graph-theoretic perspective. *Ecology* **82** 1205–1218.

URBAN, D. L., MINOR, E. S., TREML, E. A. and SCHICK, R. S. (2009). Graph models of habitat mosaics. *Ecol. Lett.* **12** 260–273.

VAN LIESHOUT, M. N. M. (2000). *Markov Point Processes and Their Applications*. Imperial College Press, London. MR1789230 https://doi.org/10.1142/9781860949760

VAN LIESHOUT, M. N. M. (2019). *Theory of Spatial Statistics*: *A Concise Introduction*. CRC Press, Boca Raton, FL.

VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21** 5–42. MR2796852

WITH, K. A. and KING, A. W. (1997). The use and misuse of neutral landscape models in ecology. *Oikos* **79** 219–229.

ZAMBERLETTI, P., PAPAÏX, J., GABRIEL, E. and OPITZ, T. (2021). Supplement to "Markov random field models for vector-based representations of landscapes." https://doi.org/10.1214/21-AOAS1447SUPP.