Noname manuscript No.
(will be inserted by the editor)

SpICE: An interpretable method for spatial data

Natalia da Silva · Ignacio Alvarez-Castro · Leonardo Moreno · Andrés Sosa

Received: / Accepted:

Abstract Statistical learning methods are widely utilised in tackling complex problems due to their flexibility, good predictive performance and ability to capture complex relationships among variables. Additionally, recently developed automatic workflows have provided a standardised approach for implementing statistical learning methods across various applications. However, these tools highlight one of the main drawbacks of statistical learning: the lack of interpretability of the results. In the past few years, a large amount of research has been focused on methods for interpreting black box models. Having interpretable statistical learning methods is necessary for obtaining a deeper understanding of these models. Specifically in problems in which spatial information is relevant, combining interpretable methods with spatial data

Corresponding author: Natalia da Silva
Instituto de Estadística (IESTA), Universidad de la República, Montevideo, Uruguay.
E-mail: natalia.dasilva@fcea.edu.uy
ORCID:0000-0002-6031-7451

Ignacio Alvarez-Castro
Instituto de Estadística (IESTA), Universidad de la República, Montevideo, Uruguay.
E-mail: ignacio.alvarez@fcea.edu.uy
ORCID:0000-0003-1633-2432

Leonardo Moreno
Instituto de Estadística (IESTA), Universidad de la República, Montevideo, Uruguay.
E-mail: leonardo.moreno@fcea.edu.uy
ORCID:0000-0003-1630-1361

Andrés Sosa
Instituto de Estadística (IESTA), Universidad de la República, Montevideo, Uruguay.
E-mail: andres.sosa@fcea.edu.uy
ORCID:0000-0002-6007-4373

can help to provide a better understanding of the problem and an improved interpretation of the results.

This paper is focused on the individual conditional expectation plot (ICE-plot), a model-agnostic method for interpreting statistical learning models and combining them with spatial information. An ICE-plot extension is proposed in which spatial information is used as a restriction to define spatial ICE (SpICE) curves. Spatial ICE curves are estimated using real data in the context of an economic problem concerning property valuation in Montevideo, Uruguay. Understanding the key factors that influence property valuation is essential for decision-making, and spatial data play a relevant role in this regard.

Keywords Interpretable method · spatial data · statistical learning · real estate

1 Introduction

Statistical learning methods have been used successfully in many fields and for different kinds of complex research problems. Some of the reasons for the extensive use of statistical learning methods are their flexibility, good predictive performance and ability to capture complex relationships among variables. In recent years, several computational tools to automate the workflow of the implementation of statistical learning models have been proposed, such as `caret` (Kuhn 2021), `h2o` (LeDell et al. 2023) and `tidymodels` (Kuhn and Wickham 2020), among others. These tools make it possible to train and tune many different algorithms, and the best one can be chosen based on a selected performance measure in a standardised way, reducing some common implementation errors. At the same time, models generated by the automatic workflow typically exhibit very good predictive performance; however, explaining the model's decisions remains challenging. These tools make more evident the inability to explain key aspects of the problem under consideration of statistical learning methods. Fortunately, there is a growing amount of research focused on methods for interpreting black box models. Broadly speaking, *interpretability* is the degree to which a human can understand decisions or predictions from a statistical method (Miller 2019). Having interpretable methods can be useful for detecting bias, understanding model errors, improving the model performance and understanding hidden relationships discovered by the algorithm, among other things. These reasons are relevant even when the objective is purely predictive.

Interpretable methods can be divided into model-specific and model-agnostic methods. Model-specific methods are intrinsically interpretable based on model characteristics, such as regression coefficients. In this case, since the interpretation is intrinsic to the model, it can be difficult to compare different models. On the other hand, model-agnostic methods are used after model fitting; they apply techniques that make it possible to analyse the results of any subsequent model to be trained. A review of the different methods for improving interpretability can be found in Molnar et al. (2020) or Maksymiuk et al. (2020). In particular, this paper is mainly concerned with the individual conditional expectation plot (ICE-plot), which was proposed by Goldstein et al. (2015). ICE curves are used to visualise the relationship between a response variable and a specific feature for each individual observation. The ICE-plot might help to represent heterogeneous effects in a problem but it may not work well in big data scenarios.

The real estate market has a key role in the economic activity of a nation and plays a central role in economic and financial crises (Mooya 2016). The global financial crisis of 2007–2008 pointed out how the real estate market can create massive financial instability. This financial market differs from other markets due to property value heterogeneity. Understanding the spatial variability of property prices is a relevant economic problem for many financial and government organisations. An adequate model for property valuation is

an important tool in the decision-making process in the public and private sectors (Osland 2010; Case et al. 2004).

Several statistical methods have been used to identify significant patterns in home pricing, from traditional hedonic models (Rosen 1974) to advanced methods that include spatial dependency and non-linear relationships. Many proposed models assume that the property price can be decomposed linearly as the sum of its determinants. These models are easy to explain and to use to obtain the joint impact of the variables on the price. However, they are also often not as good with respect to the predictive performance, because the relationships that can be learned are very restricted and generally oversimplify reality. For these reasons, the use of statistical learning methods and non-linear models has grown significantly over the few last years in the real estate industry (Limsombunchai 2004; Yoo et al. 2012; Park and Bae 2015; Goyeneche et al. 2017). Although statistical learning methods have proven useful in real estate modelling for predictive purposes, the lack of transparency and interpretation limits their use.

Usually, in these models, two groups of explanatory factors are considered. In the first place, spatial information relative to the location based on the geographical coordinates of the real estate is considered. These variables are known to be key determinants of a piece of real estate's value (Kiel and Zabel 2008). The neighbourhood, general service access, value of nearby properties, distance to special points of interest (downtown, a coast, etc.) and crime rate are some examples of factors that have a large impact on the property price. Additionally, a complementary set of explanatory variables includes features such as the number of bedrooms, number of bathrooms, and total area of the property; these are known in the property literature as hedonic variables.(Sirmans et al. 2005).

In problems in which the data contain a spatial structure, this structure can be used to improve the interpretability of statistical learning models in a natural way. The main goal of this paper is to combine interpretable statistical learning methods, specifically ICE curves, with spatial information. An extension of the ICE-plot is proposed that takes into account spatial restrictions to group the ICE curves based on the hierarchical clustering algorithm with spatial restrictions proposed in Chavent et al. (2018) combined with a Sobolev distance function that considers the distance between ICE curves. Spatial ICE curves make it easier to interpret a model's results in problems in which the number of observations is large and the spatial information is relevant.

The paper is structured as follows. In Section 2, we introduce the interpretable machine learning methods that we use in this paper, and we present the so-called spatial ICE curves. In Section 3, the results are presented. We start the section with a description of the real estate data in Montevideo (Uruguay) that were used. Using the data, several statistical learning models were trained, and the main interpretable methods were applied. Finally, we compare the results with the spatial ICE (SpICE) curves proposed. Some final remarks and further work are discussed in Section 4.

2 ICE curves in spatial problems

2.1 Basic interpretability problem

Let us consider a supervised problem, such that the goal of statistical learning models is to approximate

$$\mathbb{E}(Y|X = x) = f(x) \approx \hat{f}(x),$$

where $X = (X_1, X_2, \ldots, X_q)$ is a vector of q-variables, Y is the response variable and $\hat{f}$ is the fitted model that predicts the scalar Y as a function of X. In this context, one goal of interpretable methods is to characterise the dependence of the 'main effects' on $f(x)$ for each explanatory variable. It is also possible to analyse the 'low-order' dependence between pairs of variables.

Let us assume that the main goal is to understand the effect of a set of explanatory variables denoted as X_S on the response variable in a model-agnostic way ($S \in \{1, 2, \ldots, q\}$, and denote by the subset C the complement of S). One of the earliest methods for this is the partial dependence plot (PD-plot) (Friedman 2001). The PD-plot describes the change of the response variable in a model as a function of the marginal effect of one or more variables (subset X_S) when the effects of the other explanatory variables (complementary subset X_C) are averaged.

The main advantages of the PD-plot are that its estimation is very intuitive and that it presents a causal interpretation of the results of any learning model (Zhao and Hastie 2021). The main drawbacks are that it hides heterogeneous effects, it might rely on an unrealistic set of observations and it is computationally expensive. Alternatives to the PD-plot have been proposed to overcome its problems, such as the accumulated local effects plot (ALE-plot) (Apley and Zhu 2020) and the individual conditional expectation plot (ICE-plot) (Goldstein et al. 2015).

2.2 Individual conditional expectation plot (ICE-plot)

The ICE-plot is proposed as an extension of the PD-plot to visualise the dependence of the prediction on a feature for each sample separately, with one curve per observation. The method attempts to capture the dependency of the response variable on a set of variables, allowing heterogeneous effects.

For each observation, the curve $\hat{f}_S^{(i)}(x_S)$ is obtained by varying x_S in the function $\hat{f}(x_S, x_C^{(i)})$ while the variables $x_C^{(i)}$ remain constant. That is to say,

$$\hat{f}_S^{(i)}(x_S) = \hat{f}(x_S, x_C^{(i)}). \tag{1}$$

It is worth noting that averaging the ICE curves corresponds to the definition of the PD-plot. Thus, it is possible to interpret the ICE curves similarly to the classic PD-plot, but it is also possible to pick interactions when visualising the N plots (something that vanishes when all ICE curves are averaged into

one curve). A visual example is added to illustrate the ICE interpretation. Figure 1 shows ICE curves corresponding to individual observations from the dataset described in Section 3; only three curves are presented to simplify the example. Each curve examines the association between the price per square metre and the corresponding total area of the apartment. Having individual curves for each observation makes it possible to visualise potential heterogeneous effects; in particular, the solid ICE line suggests a small effect at around $e^{3.5}$ square metres, while the two dashed ICE curves show a larger negative impact, with a pattern closer to linearity.

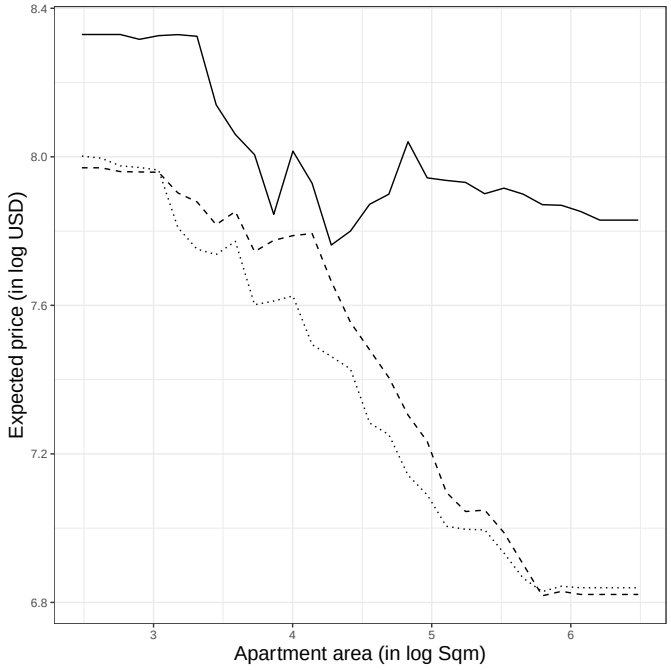

Fig. 1 Individual ICE curves, where each line represents one selected observation.

2.3 Spatial ICE (SpICE) curves

The interpretation of ICE curves can be challenging when a large number of observations is available. A solution based on clustering ICE curves, which takes into account two sources of information, is proposed. The main source of information corresponds to the covariate effect profile described by the ICE curve itself. This is relevant to maintaining the original goal of the individual ICE curves. Additionally, when observations contain individual spatial information, location information can be used to improve the interpretability of ICE curves. Spatial data is data where the location of the measurements is

a key element. This often includes data that references a specific geographical area. In this paper, SpICE refers to clustering ICE curves with spatial contiguity constraints.

In order to construct clusters of ICE curves, it is necessary to define a distance or dissimilarity measure between functions. There are many approaches that can be used to solve this problem; a typically used metric for functional clustering is the L2 distance:

$$d_{L2}(i,j) = \sqrt{\int |\hat{f}_S^{(i)}(x) - \hat{f}_S^{(j)}(x)|^2 dx},$$

where $d_{L2}(i,j)$ represents the L2 distance between ICE curve i and ICE curve j. In Appendix A, a model example in which the L2 distance is not adequate to distinguish ICE curves with different predictor effects profiles is presented. Figure 7 presents two scenarios in which $d_{L2}(\cdot,\cdot)$ produces the same value but the covariate effect is not the same.

Moreover, Hitchcock and Greenwood (2015) point out that in many applications, information about the derivatives of the function is even more relevant than the function itself. In these cases, one approach could be to use the same L2 distance on the derivative functions. However, in clustering ICE curves, the key idea is to group observations with similar predictor effects on the response; thus, considering both the level and variation is important. For these reasons, it is appropriate to consider a distance that takes into account not only the value of the function but also its growth/decrease. Therefore, for the ICE function $\hat{f}_S^{(i)}(x_S)$, it is possible to use the Sobolev $W^{1,2}(\mathbb{R})$ distance (Adams and Fournier 2003):

$$d_{Sob}(i,j) = \sqrt{\int |\hat{f}_S^{(i)}(x) - \hat{f}_S^{(j)}(x)|^2 dx + \int |\hat{f}'^{(i)}_S(x) - \hat{f}'^{(j)}_S(x)|^2 dx}. \quad (2)$$

The Sobolev distance has recently been used in some functional data applications as a way to highlight complex functional patterns. Cremona and Chiaromonte (2023) and Ehsani and Drabløs (2020) report good performances when this type of distance is used in different statistical applications. Particularly in clustering functional data, using the Sobolev distance might improve the clustering solution. In Appendix B, results from a small simulation study are presented. A classical functional data example (Hitchcock et al. 2007) is used to compare the proportions of correctly matched pairs of curves when the L2 or Sobolev distance is used. The Sobolev distance shows better results in at least 95% of the 500 replications.

However, some statistical learning models (such as tree-based methods) produce nondifferentiable predictor functions; thus, the ICE curves for these models inherit the nondifferentiability property, which makes the computation of $d_{Sob}(,)$ impossible. One way to overcome this issue and obtain estimates in a C^1 space is to consider the curve $\widetilde{f}_{S,ICE}^{(i)}(x_S)$, which is the convolution of the

function $\hat{f}(x_S, x_C^{(i)})$ with a Gaussian kernel. That is,

$$\widetilde{f}_S^{(i)}(x_S) := \hat{f}(x_S, x_C^{(i)}) * K_h(x_S), \tag{3}$$

where K_h is the Gaussian kernel, h is the smoothing parameter and $*$ is the convolution operation. It is worth noting that this operation can be thought of as a pre-smoothing of the functional data, which has been shown to be beneficial in the cluster analysis of functions (Hitchcock et al. 2007). Additionally, when the true $f(x)$ is a smooth function, the above transformation of the ICE curve (the convolution) will not distort the ICE estimator in a relevant way.

When considering ICE curves, each curve is associated with a specific observation. In applications in which the location is a key element, i.e. spatial problems, it is possible to link such curves with the location information of the corresponding observation. Combining these two sources of information might improve the interpretability of ICE curves.

Chavent et al. (2018) propose an algorithm to construct clusters of multivariate data with spatial restrictions of contiguity. Let us consider a partition $\mathcal{P}_K = (C_1, \ldots, C_K)$ into K clusters. Using two dissimilarity matrices, D_0 and D_1, and a mixing parameter $\alpha \in [0, 1]$, the pseudo-inertia in cluster k can be defined as

$$I_\alpha\left(C_k^\alpha\right) = (1-\alpha) \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} d_{0,ij}^2 + \alpha \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} d_{1,ij}^2,$$

where $d_{0,ij}$ and $d_{1,ij}$ represent dissimilarities between observations i and j in D_0 and D_1, respectively. Then, a Ward-like method (Ward 1963) is used for the construction of the clusters, and a data-driven procedure to assist in the choice of the α value is proposed.

In this paper, this clustering algorithm is adapted to cluster ICE curves (functional data). Mainly, the dissimilarity matrix D_0 represents distances among ICE curves instead of multivariate data. The Sobolev distance is used to compute dissimilarities among pre-smoothed ICE curves and to construct the D_0 matrix, i.e $d_{0,ij} = d_{sob}(i, j)$ (see Equation (2)). As mentioned earlier, this distance takes into account the covariate effect profile in terms of both the level and variation and could be beneficial for the clustering solution. In addition to ICE information, the geographical distance between observations is used for the D_1 matrix. The `SpICE` R package implements the computation and visualisation of SpICE curves; it is available at https://github.com/natydasilva/SpICE.

3 Application results

3.1 Dataset: Apartment prices in Montevideo

The data in this paper are from an eCommerce platform called Mercado Libre, where apartments and houses are offered for sale and for rent. The complete dataset contains asking price information for properties in Montevideo,

the capital city of Uruguay, from February 2018 to January 2019 (Picardo 2019). There are 92,832 observations; however, only apartment data will be considered for the analysis, which results in 70,817 observations concerning apartments for sale in Montevideo.

Figure 2 shows the distribution of the asking price for apartments in Montevideo, where each dot represents one apartment for sale and the colour represents the asking price of the apartment per square metre in US dollars. Figure 2 suggests that locations close to the coast are associated with a higher price range. The asking price ranges from 541 to 5,000 US dollars per square metre in Montevideo, and the median value is close to 2,500 US dollars.

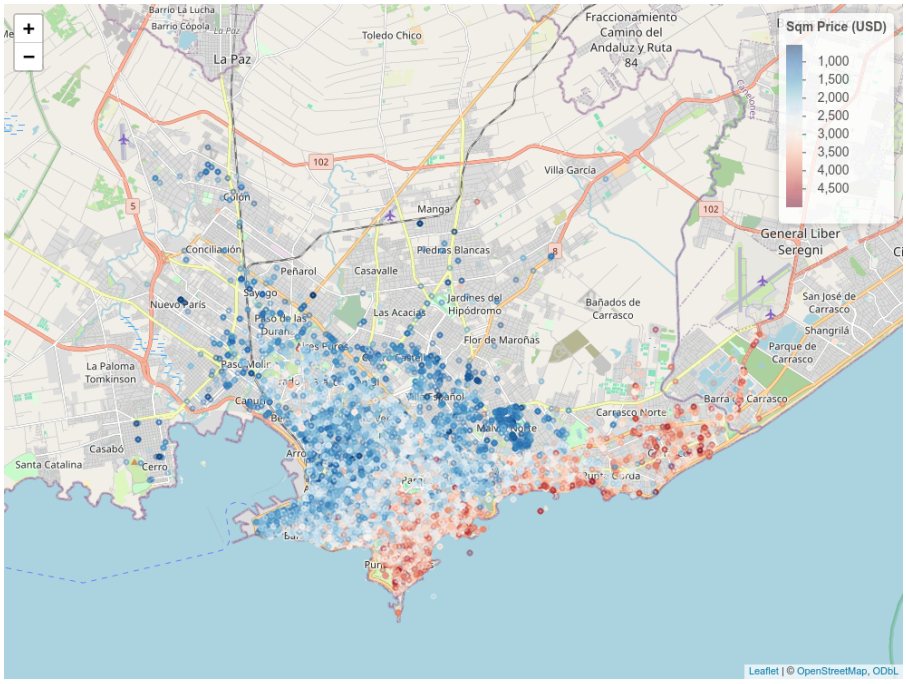

Fig. 2 Distribution of the asking price for apartments in Montevideo. Each dot represents one apartment, and the colour represents the asking price offered by the seller in US dollars per square metre.

The complete dataset contains 116 explanatory variables representing apartment features commonly used in real estate modelling. With the available information, an additional two explanatory variables were created. First, a variable indicating the distance from the apartment to the beach was computed since it appears to be a relevant feature based on data exploration (see Figure 2). Second, many of the variables indicate the presence/absence of one specific amenity; thus, a variable indicating the total number of amenities present in an apartment was computed.

Table 1 Variable description

Variable	Description
`lpricem2`	Log of asking price in US dollars per square metre.
`amenities`	Total number of amenities present in the apartment.
`bedroom`	Bedroom quantity. Reduced to values between 0 (studio) and 3.
`bathroom`	Bathroom quantity. Reduced to values between 1 and 3.
`elevators`	Elevator quantity. Reduced to values between 0 and 2.
`condition`	Property condition (new/used). Properties that were less than 1 year old were marked as 'new'.
`expenses`	Numerical value representing monthly expenses in Uruguayan pesos (local currency).
`garage`	Whether or not there is a garage. Reduced to values between 0 ('No') and 1 ('Yes').
`ldistance_beach`	Minimum distance (Euclidean) between the property and the beach (on a log scale).
`larea_apt`	Log of the apartment area in square metres. Values over 2,000 or under 9 square metres were removed from the data.
`neighborhoodgr`	Montevideo neighbourhoods grouped by proximity in 12 regions.
`lat`	Property latitude coordinate.
`long`	Property longitude coordinate.

A reduced number of variables were selected to improve the data quality of the complete dataset. Table 1 presents the selected variables, indicating the name of each variable in the dataset and a brief description. All the models were fitted using the natural logarithm of the price per square metre (`lpricem2`) as the response variable. Finally, the geographical coordinates of each property were not used as explanatory variables in the models but were used as additional information to improve interpretability using the SpICE curves described earlier.

3.2 Predictive models

Several models were trained using the automatic machine learning (`autoML`) procedure from the `h2o` R package (LeDell et al. 2023) to predict the apartment price as an alternative to classical methods. `autoML` estimates well-tuned models in four families: penalised linear models (glm), random forest (drf), extreme gradient boosting (xgboost) and fully connected multi-layer artificial neural networks (deeplearning). Additionally, stacked ensembles (stackedensemble) of individual models are trained; this includes the combination of all the models and ensembles using subsets of trained models. In the rest of the paper, the best model of each family and the best-performing stacked ensemble are used.

Table 2 shows the performance measures for the selected predictive models. The root mean square error (RMSE) and the R^2 value are used to evaluate the model performance. These values are computed with the response variable in logs (as this was done for the training of every model). Additionally, the mean absolute error (MAEo) and mean absolute percentage error (MAPEo) are both

computed on the original scale of the response variable, so they have units of dollars per square metre. The four measures are computed using a testing dataset different from the training samples (2/3 training and 1/3 testing).

Table 2 Predictive performance measures by model

	model	rMSE	R2	MAEo	MAPEo
1	stackedensemble	0.14	0.81	242.68	9.58
2	drf	0.14	0.80	251.08	9.94
3	xgboost	0.15	0.80	254.48	10.06
4	deeplearning	0.20	0.63	395.84	15.18
5	glm	0.22	0.55	427.72	17.13

In terms of the predictive performance, the stacked ensemble, random forest and xgboost algorithms show similar performances; they are somewhat better than the deep learning method or the penalised linear model. It is worth noting that the stacked model combines seven tree-based models. The best model obtained an average error in the asking price of $243 per square metre, or around 9.6% of the observed price.

3.3 Variable importance measures

The first approach to interpreting model results for statistical learning methods is to compute variable importance measures. The variable importance measures were scaled so that the most important variable in each model has a value of 1, which simplifies the model comparison. In Figure 3, we show the results by model for the variable importance. The ordering of the predictors is similar in all models. The apartment area (`larea_apt`) and the neighbourhood (`neighborhoodgr`) are the two most important variables for predicting the apartment price. All the tree-based methods (drf, xgboost and stackedensemble) show a third relevant variable, which is the distance from the apartment to the beach (`ldistance_beach`). The rest of the predictor variables are not relevant for prediction.

3.4 Partial effect of apartment area

In subsection 3.3, the variable importance was presented; these results provide a ranking of variables according to their relevance for predicting the response variable. However, they do not provide information on the effect that the individual variables have on the response variable. In order to characterise the average effect of apartment features on the price, the PD-plot and the ALE-plot are estimated. The algorithms used for this estimation are from the R packages `pdp` (Greenwell 2017) and `ALEPlot` (Apley 2018).

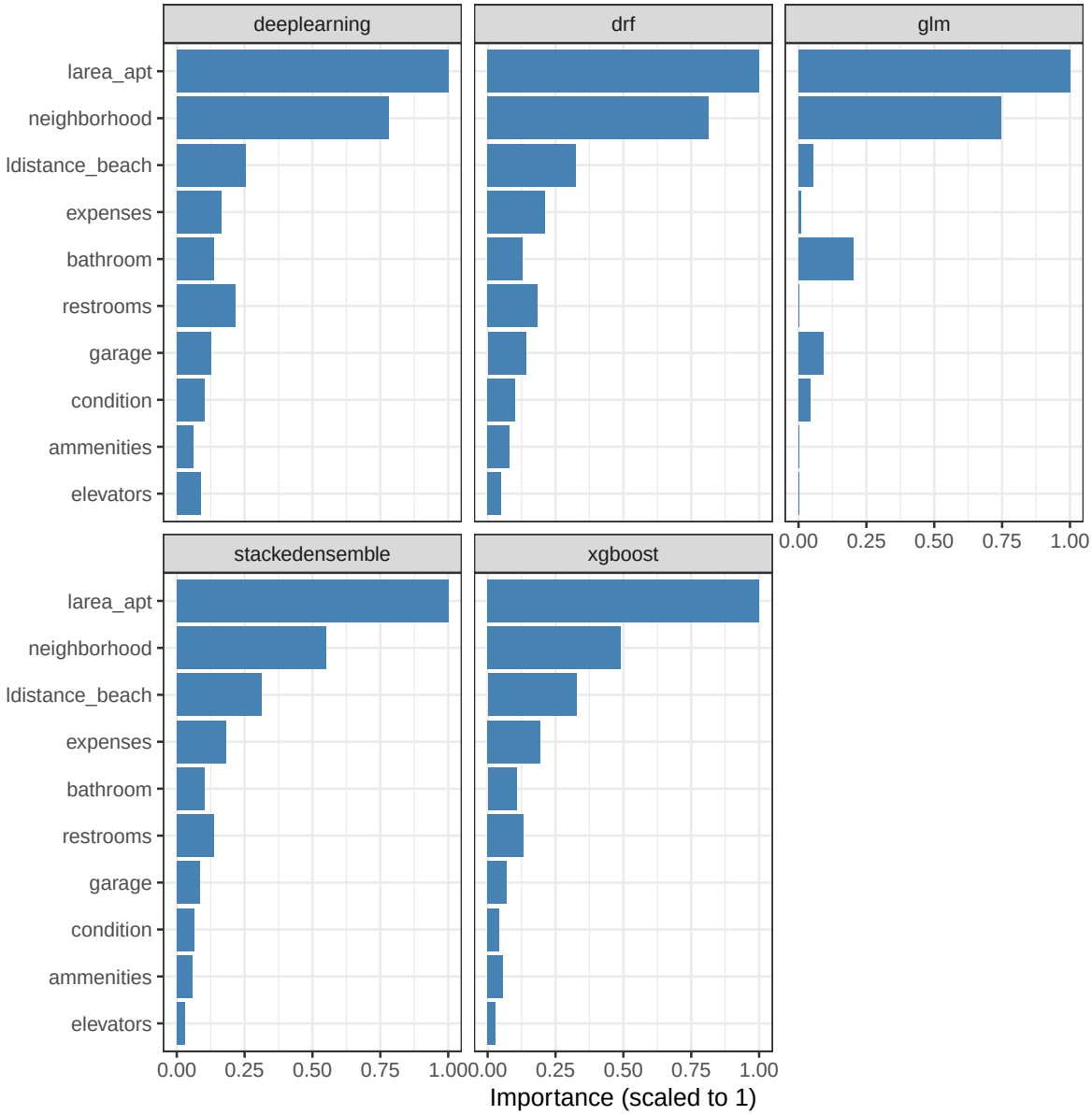

Fig. 3 Variable importance. Each panel represents a model, and the y-axis shows the variables included in all the models. The bar length represents the scaled variable importance measure used to predict the apartment price.

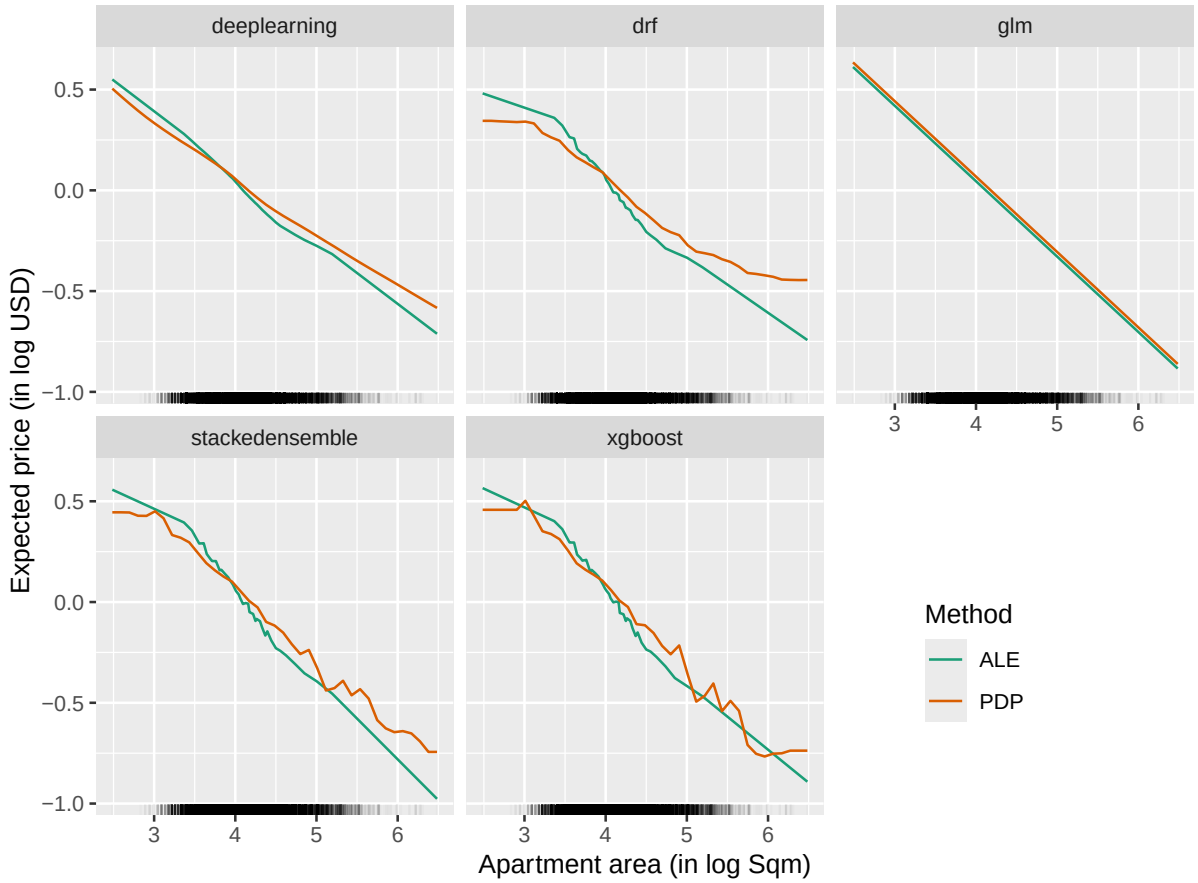

Fig. 4 Effect of the `larea_apt` variable in different models. Each panel corresponds to a predictive model and the colour represents the interpretable method (ALE-plot or PD-plot).

The apartment area (`larea_apt`) is the most relevant feature in every model. The effect of this variable on the response is described with the PD-plots and ALE-plots shown in Figure 4, where each panel corresponds to a model. A negative effect is suggested by these plots, with a similar effect in every model. Especially, this is true in the middle of the range for the apartment area, where most of the observed samples lie. Some differences can be seen in very large or very small apartments, where random forest shows smaller effects on prices.

Specifically, Figure 4 shows the average effects of the variable `larea_apt` in the sample. In a scenario in which the impact of an explanatory variable on the response presents heterogeneity among the observations, the PD-plot and ALE-plot methods hide the variability of effects. An alternative method for interpretability that can be used to tackle this issue is the ICE-plot. The value of visualising the individual curves that compose the PD-plot lies in exploring other patterns in the effects, rather than just the mean value.

When big datasets are involved, a major disadvantage of ICE-plots is overplotting, which makes it difficult to answer anything. This is especially relevant for ICE curves since their main purpose is to look for heterogeneity patterns in the predictor effect. The use of graphical solutions such as transparency or 2-dimensional histograms is not suitable for plotting lines. In the `h2o` implementation (Hall et al. (2017)), a few ICE curves for decile values (in the observed response) are displayed, so no matter how large the sample data are, only ten ICE curves are displayed, resulting in an oversimplified plot. An alternative is to plot a stratified sample of curves. This method can be combined with line transparency to allow one to reduce overplotting while making it possible to see the different patterns of the ICE curves in the data. Figure 5 shows the individual conditional expectation plot for the apartment area (`larea_apt`) for the random forest model. The other predictive models shows similar results, as can be seen in Figure 10 in Appendix C.

The results illustrate the negative effect of the variable `larea_apt` in each individual curve. However, it is relevant to note that the effect presents heterogeneity for the different properties. Figure 5 suggests that for properties with a high predicted price, the apartment area presents a small and linear effect, while cheap properties show a non-linear and larger impact of the apartment area.

3.5 SpICE curve effects

Finally, to explore the connections among ICE curves and the geographical locations of properties, the SpICE curve results are presented. A range of three to five clusters is considered; for each value, an optimal value for the α parameter is determined as a compromise between the internal homogeneity of the clusters in terms of the ICE curves and the geographical information, approximately. The results are shown in Figure 11 in Appendix C, using four groups and $\alpha = 0.5$.

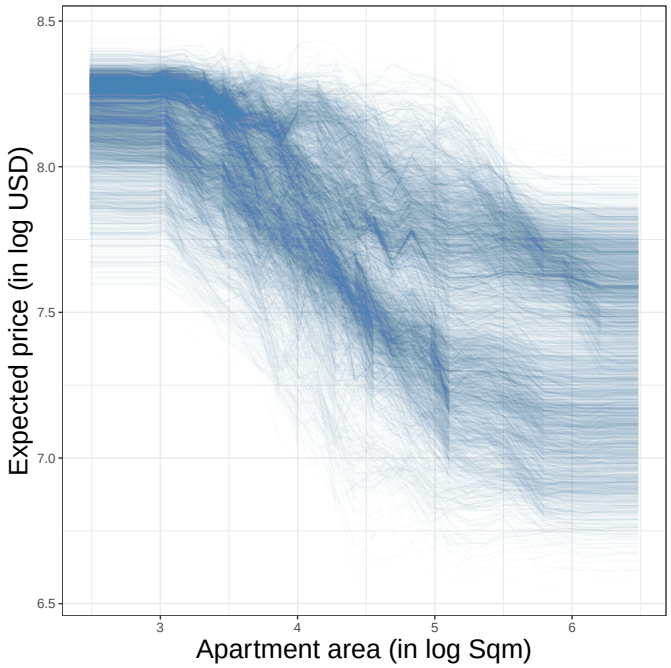

Fig. 5 ICE-plot for the `larea_apt` variable. Each panel corresponds to a predictive model. Five thousand randomly stratified selected curves are shown.

Figure 6 show the results of the geographically constrained clusters; an apartment for sale in Montevideo is represented as a point on the map that is linked to an ICE curve in the bottom panel. The colours of the points and curves indicate the cluster that they belong to. The cluster locations suggest a layout in the north-west/south-east direction, similar to the price gradient present in the data (see Figure 2). The apartments in the west and on the north-west side (green cluster) correspond to low-income neighbourhoods, while the east side of the city (red cluster) represents the zone with the highest income in Montevideo.

Focusing on the SpICE curves presented in Figure 6, there are distinct patterns in the relationship between the apartment area and the asking price per square metre. Across all clusters, there is a negative relationship between the price per square metre and the apartment area. However, the effect of the apartment area on the price per square metre differs between the green and blue clusters (associated with medium- and low-income neighbourhoods) and the red and violet clusters (associated with high-income neighbourhoods). In the red and violet clusters, an increase in the apartment area results in a smaller decrease in the price per square metre compared to the green and blue clusters. This suggests that, in high-income neighbourhoods, the total apartment price (not the price per square metre) is very sensitive to changes in the apartment area. Conversely, in low-income neighbourhoods, the total

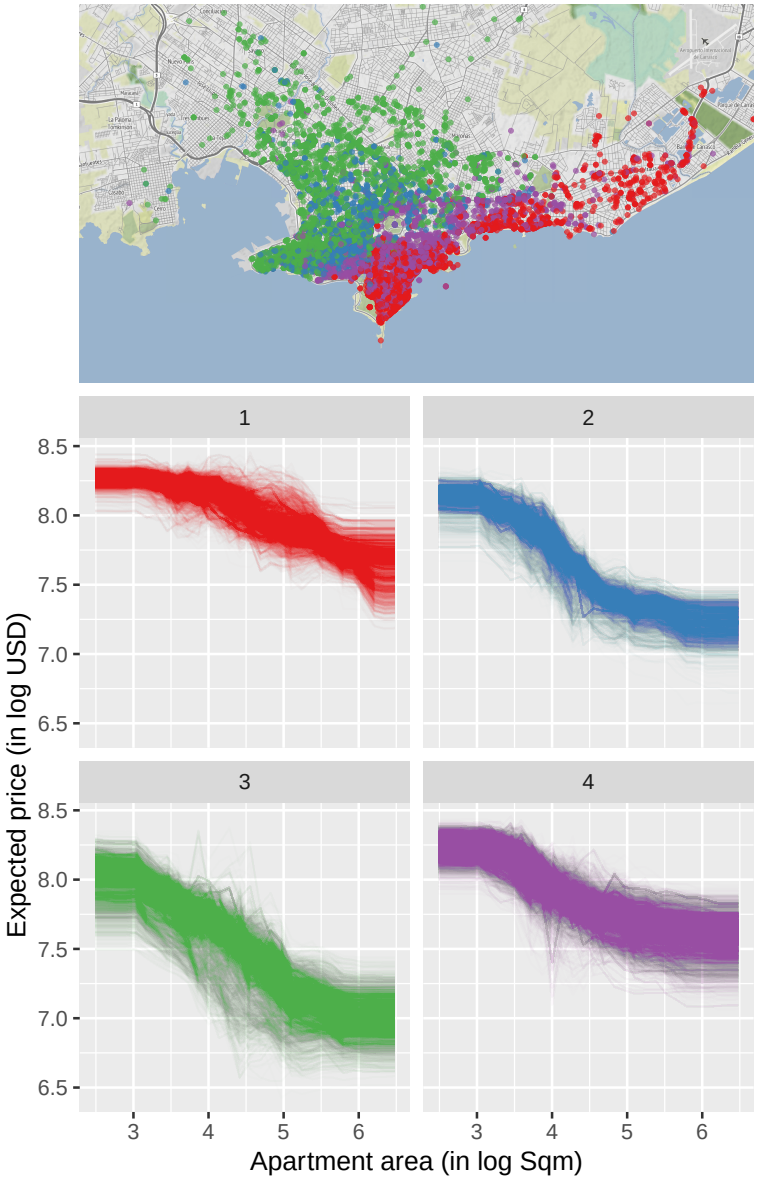

Fig. 6 SpICE curves and geographical locations of clusters.

apartment price is inelastic to changes in the apartment area. Consequently, additional square metres in the apartment have no impact on the total apartment price, leading to a decline in the price per square metre.

4 Discussion

In this paper, interpretable methods for measuring heterogeneous covariate effects for black box models, functional data clustering and spatial information were combined to improve interpretability in spatial applications. In particular, geographically constrained clusters of ICE curves called SpICE curves were constructed, combining two sources of information. On the one hand, the covariate effect profile is considered by using the Sobolev distance as a dissimilarity measure for the ICE curves. The Sobolev distance is useful from an interpretability point of view since it summarises the main characteristics of the covariate effect (level and variation). On the other hand, the location information of observations is used to obtain geographically contiguous clusters.

Similarly to the ICE-plot, the SpICE curves can show the heterogeneous effects of a predictor variable in a black box model. SpICE curves are easier to work with in big data applications since it is possible to interpret each cluster instead of each individual curve. In specific problems, the spatial contiguity of the cluster provides more interpretable information associated with other relevant aspects that may not be present in the ICE curves.

In the motivating example, SpICE curves were constructed to profile the effect of the total area of a property on the price per square metre in Montevideo apartments. Five statistical learning methods were selected from a list of fitted models using `h2o` with the `autoML` procedure based on the predictive performance. The predictor variable 'total area of the apartment' was (unsurprisingly) the most important feature in all models. Then, using a fitted model with the random forest algorithm, the ICE-plot and SpICE curves for each apartment were computed. The spatial information of the properties was combined with ICE curves to gain interpretability. Four property clusters were selected, based on a distance that combines the functional distance between ICE curves and the geographical distance between apartment coordinates. The blue and green clusters mainly represent properties located in medium- or low-income zones in Montevideo, and these clusters present a large, negative, non-linear effect of the apartment area on the price per square metre. On the other hand, the red and violet clusters mainly correspond to high-income neighbourhoods, and they show a small, close-to-linear effect of the property area.

There are several aspects of this paper that could be explored in future work. First, the choice of the distance between the ICE curves could be improved; specifically, it is relevant to analyse more deeply in which scenarios the Sobolev distance results in a better clustering solution for functional data. Second, instead of constructing fixed clusters of ICE curves, it is possible to consider the local average of the ICE curves in a nearest neighbours fashion, where the distance used to determine the neighbours could combine the structure of the ICE curves and the geographical distance. Finally, a summary of the clustered ICE curves could be based on the ALE-plot instead of the average of the clustered ICE curves.

5 Supplementary material

This article was written with the R packages `knitr` (Xie 2023), `ggplot2` (Wickham 2016), `leaflet` (Cheng et al. 2022), `tidyverse` (Wickham et al. 2019), `h2o` (LeDell et al. 2023), `ClustGeo` (Chavent et al. 2021) and `KernSmooth` (Wand 2021). The files needed to reproduce the article and the results are available at https://github.com/natydasilva/SpICE_COST. Additionally, the `SpICE` R package implements the computation and visualisation of SpICE curves proposed in this paper and it is available at https://github.com/natydasilva/SpICE.

References

Adams, R. A. and Fournier, J. J. F. (2003). *Sobolev spaces*. Elsevier/Academic Press.

Apley, D. (2018). *ALEPlot: Accumulated Local Effects (ALE) Plots and Partial Dependence (PD) Plots*. R package version 1.1.

Apley, D. W. and Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):1059–1086.

Case, B., Clapp, J., Dubin, R., and Rodriguez, M. (2004). Modeling spatial and temporal house price patterns: A comparison of four models. *The Journal of Real Estate Finance and Economics*, 29(2):167–191.

Chavent, M., Kuentz, V., Labenne, A., and Saracco, J. (2021). *ClustGeo: Hierarchical Clustering with Spatial Constraints*. R package version 2.1.

Chavent, M., Kuentz-Simonet, V., Labenne, A., and Saracco, J. (2018). Clustgeo: an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 33(4):1799–1822.

Cheng, J., Karambelkar, B., and Xie, Y. (2022). *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 2.1.1.

Cremona, M. A. and Chiaromonte, F. (2023). Probabilistic k-means with local alignment for clustering and motif discovery in functional data. *Journal of Computational and Graphical Statistics*, pages 1–12.

Ehsani, R. and Drabløs, F. (2020). Robust distance measures for k nn classification of cancer data. *Cancer Informatics*, 19:1176935120965542.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232.

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1):44–65.

Goyeneche, J. J., Moreno, L., and Scavino, M. (2017). Predicción del valor de un inmueble mediante técnicas agregativas. *Serie Documento de Trabajo (17/1)*.

Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. *The R Journal*, 9(1):421–436.

Hall, P., Gill, N., Kurka, M., and Phan, W. (2017). Machine learning interpretability with h2o driverless ai. *Open source, distributed machine learning platform*.

Hitchcock, D. B., Booth, J. G., and Casella, G. (2007). The effect of pre-smoothing functional data on cluster analysis. *Journal of Statistical Computation and Simulation*, 77(12):1043–1055.

Hitchcock, D. B. and Greenwood, M. C. (2015). Clustering functional data. In Hennig, C., Meila, M., Murtagh, F., and Rocci, R., editors, *Handbook of Cluster Analysis*, chapter 13, pages 286–309. Chapman and Hall/CRC.

Kiel, K. A. and Zabel, J. E. (2008). Location, location, location: The 3l approach to house price determination. *Journal of Housing Economics*, 17(2):175–190.

Kuhn, M. (2021). *caret: Classification and Regression Training*. R package version 6.0-90.

Kuhn, M. and Wickham, H. (2020). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.* R package.

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., and Malohlava, M. (2023). h2o: R interface for the 'h2o' scalable machine learning platform. R package version 3.40.0.1.

Limsombunchai, V. (2004). House price prediction: hedonic price model vs. artificial neural network. In *New Zealand Agricultural and Resource Economics Society Conference*, pages 25–26.

Maksymiuk, S., Gosiewska, A., and Biecek, P. (2020). Landscape of r packages for explainable artificial intelligence. *arXiv preprint arXiv:2009.13248*.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38.

Molnar, C., Casalicchio, G., and Bischl, B. (2020). Interpretable machine learning–a brief history, state-of-the-art and challenges. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 417–431. Springer.

Mooya, M. M. (2016). Standard theory of real estate market value: Concepts and problems. *Real Estate Valuation Theory: A Critical Appraisal*, pages 1–21.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, 32(3):289–320.

Park, B. and Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of fairfax county, Virginia housing data. *Expert Systems with Applications*, 42(6):2928–2934.

Picardo, P. (2019). Predicción de precios de la vivienda aprendizaje estadístico con datos de ofertas y transacciones para montevideo. *Tesis de Maestría en Economía, FCEA-UDELAR*.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Sirmans, S., Macpherson, D., and Zietz, E. (2005). The composition of hedonic pricing models. *Journal of Real Estate Literature*, 13(1):1–44.

Wand, M. (2021). *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*. R package version 2.23-20.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Xie, Y. (2023). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.42.

Yoo, S., Im, J., and Wagner, J. E. (2012). Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, New York. *Landscape and Urban Planning*, 107(3):293–306.

Zhao, Q. and Hastie, T. (2021). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1):272–281.

A Minimal example

This section presents a model example in which the L2 distance is not a good choice for finding dissimilarities among ICE curves. Let us assume the regression model

$$\begin{aligned} Y_i &= f(X_{1,i}, X_{2,i}) + \epsilon_i \\ &= -C(1 - X_{1,i})X_{1,i} + \tfrac{1}{2}X_{2,i}(2 - X_{1,i} - X_{1,i}^2) + X_{2,i}^{1+X_{1,i}\mathbb{1}_{\{X_{1,i}\geq 0\}}} + \epsilon_i, \end{aligned}$$

where $X_1 \sim U(-1,1)$, $X_2 \sim U(0,2)$ and ϵ_i represents white noise with a standard deviation $\sigma > 0$. When the true $f()$ function is known, the ICE curves at three chosen points are $f(-1,x2) = 2C + 2x_2$, $f(0,x2) = 2x_2$ and $f(1,x2) = x_2^2$. If $C = -\sqrt{\frac{\int_0^2 (x^2-2x)^2 dx}{8}}$, then

$$d_2\left(f(-1,\cdot), f(0,\cdot)\right) = d_2\left(f(1,\cdot), f(0,\cdot)\right).$$

Figure 7 shows these three ICE curves in two panels. The left panel (case 1) contains $f(-1,\cdot)$ and $f(0,\cdot)$, two parallel linear functions representing the same effect of $X2$ on the response variable. Meanwhile, the right panel (case 2) shows the ICE curves for $f(1,\cdot)$ and $f(0,\cdot)$; here, the two curves represent different effects of $X2$.

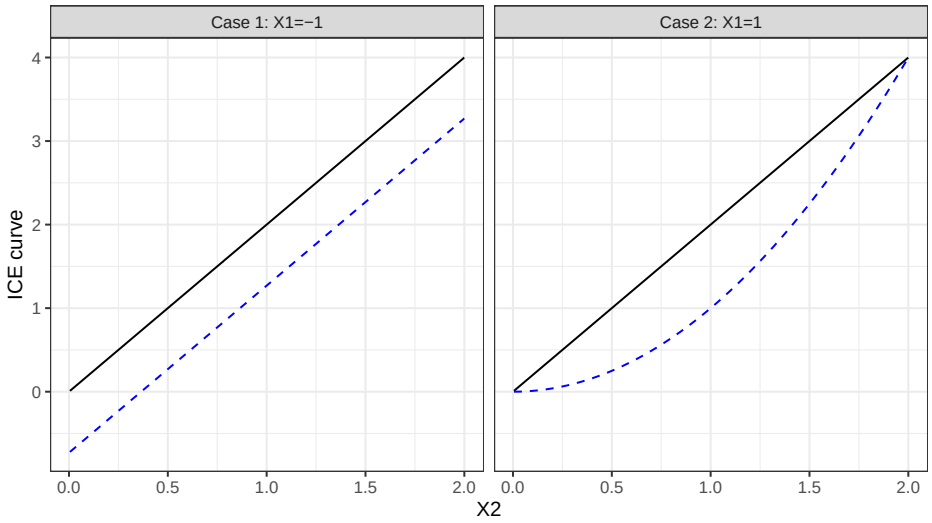

Fig. 7 ICE curves from the model example.

From an interpretability point of view, it is important to differentiate between the two cases plotted in Figure 7; however, the L2 distance is the same in these two cases.

B Simulation study

A small simulation study comparing the performances of the L2 and Sobolev distances in clustering functional data is performed, following the design choices proposed in Hitchcock et al. (2007).

Functional data are simulated with a four-cluster structure, based on four functions that represent the mean of each group. Figure 8 shows the true mean functions, $\mu_g(x)$, $x \in [0,5]$, where $g = 1,2,3,4$ indicates the cluster. According to Hitchcock et al. (2007), the $\mu_g(\cdot)$ *'... were intentionally chosen to be similar enough to provide a good test for the clustering methods that attempted to group the curves into the correct clustering structure, yet different enough that they represented four clearly distinct processes.'*

Synthetic data are obtained by adding random noise to $\mu_g(x)$:

$$y_{ig} = \mu_g(x_i) + u_g + \epsilon_{ig},$$

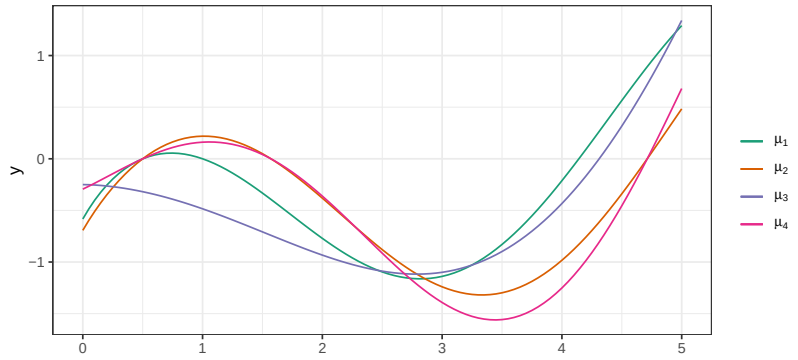

Fig. 8 True mean functions.

where $u_g \sim N(0,1)$ and $\epsilon_{ig} \sim N(0, 0.1^2)$. Figure 9 shows a simulated dataset, consisting of $N = 40$ curves (10 curves from each group) represented in a discretised form, with $n = 50$ values per curve, equally spaced in $[0, 5]$.

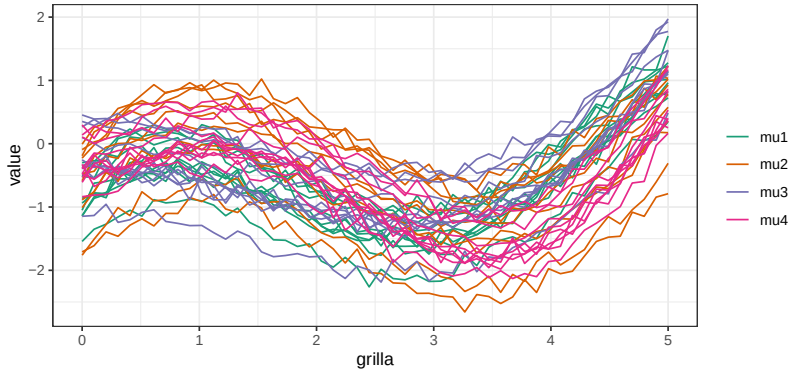

Fig. 9 A simulated dataset.

The clusters are estimated using the k-medoids method with the `pam()` function in `R`, with pre-smoothed data as the main input. Two cluster solutions are obtained for each simulated dataset, using the L2 and Sobolev distances as the dissimilarity function. The cluster solution performance is evaluated based on the proportion of pairs of curves that are correctly matched in the same group. The process is replicated 500 times.

Table 3 presents results from the 500 replications. Using the Sobolev distance provides a larger proportion of correctly matched pairs on average; additionally, the 5% quantile of

the ratio is slightly larger than 1, so the Sobolev distance resulted in a better performance for at least 95% of the simulations.

Table 3 Summary of results

	mean	sd	Q0.05	Q0.95
L2	0.48	0.06	0.37	0.58
Sob	0.59	0.09	0.47	0.74
Ratio	1.26	0.20	1.01	1.63

C Supplementary figures

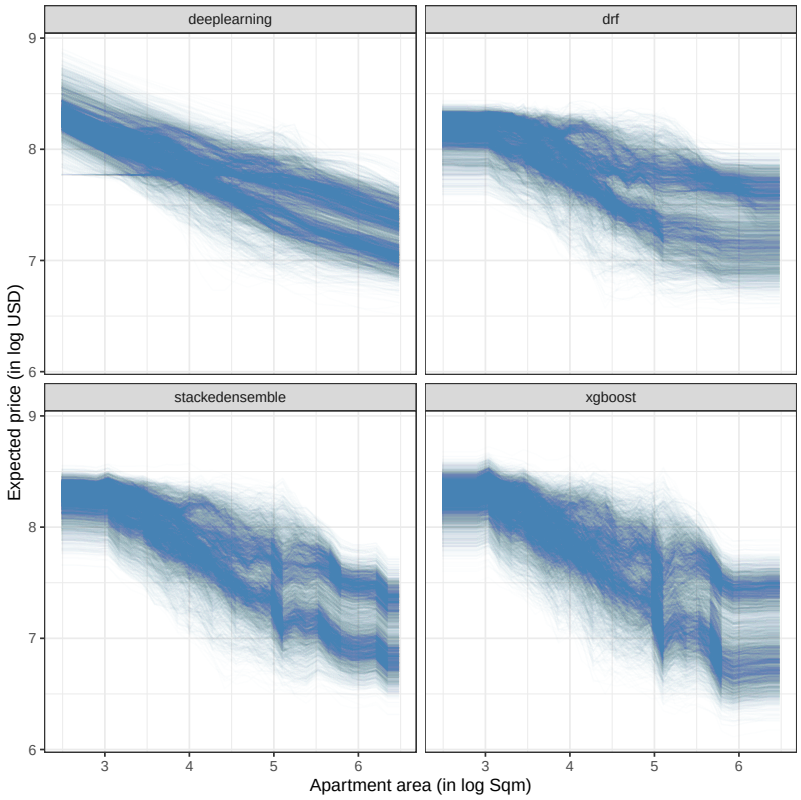

Fig. 10 ICE-plot for the *log apartment area* variable. Each panel corresponds to a predictive model. Five thousand randomly stratified selected curves are shown.

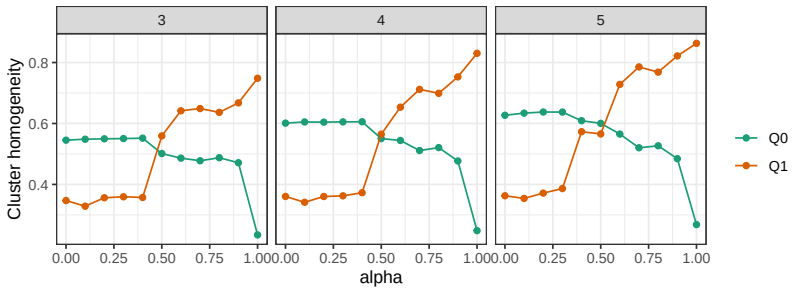

Fig. 11 Optimal α for different groups.