Predicting environmental features by learning spatiotemporal embeddings from social media

Shelan S. Jeawak[a,b,*], Christopher B. Jones[a] and Steven Schockaert[a]

[a] *Cardiff University, School of Computer Science and Informatics, Cardiff, UK*
[b] *Al-Nahrain University, Department of Computer Science, Baghdad, Iraq*

ARTICLE INFO

Keywords:
Social media
Flickr
Text mining
Vector space embeddings
Spatiotemporal data
Ecology

ABSTRACT

Spatiotemporal modelling is an important task for ecology. Social media tags have been found to have great potential to assist in predicting aspects of the natural environment, particularly through the use of machine learning methods. Here we propose a novel spatiotemporal embeddings model, called SPATE, which is able to integrate textual information from the photo-sharing platform Flickr and structured scientific information from more traditional environmental data sources. The proposed model can be used for modelling and predicting a wide variety of ecological features such as species distribution, as well as related phenomena such as climate features. We first propose a new method based on spatiotemporal kernel density estimation to handle the sparsity of Flickr tag distributions over space and time. Then, we efficiently integrate the spatially and temporally smoothed Flickr tags with the structured scientific data into low-dimensional vector space representations. We experimentally show that our model is able to substantially outperform baselines that rely only on Flickr or only on traditional sources.

1. Introduction

With the popularity of social media, a large amount of user generated textual data that is grounded in time and space has become available. As an example, Flickr[1], a photo-sharing platform, hosts more than 10 billion photographs[2], most of which are associated with short textual descriptions in the form of tags to describe what is depicted in the photograph. In addition, the time at which these photographs were taken and their geographical coordinates are available as meta-data for many photographs. The tags associated with such georeferenced photographs often describe the location where they were taken and Flickr can thus be regarded as a source of environmental information. The use of Flickr for modelling urban environments has already received considerable attention. For instance, various approaches have been proposed for modelling urban regions (Cunha and Martins, 2014), and for identifying points-of-interest (Van Canneyt et al., 2013a) and itineraries (De Choudhury et al., 2010; Quercia et al., 2014). However, using Flickr for modelling the natural environment has so far received only limited attention. Other social media such as Instagram and Facebook have had very limited applications due to the restrictions of data access (Ghermandi and Sinclair, 2019).

Many recent studies have highlighted the fact that Flickr captures valuable ecological information (Ghermandi and Sinclair, 2019), which can complement more traditional sources. A shortcoming of most of these existing methods is that they rely on manual evaluations, with little automated exploitation of the associated tags, and fail therefore to exploit the full potential of the data (Richards and Friess, 2015; ElQadi et al., 2017). This motivates us to automate methods that can utilize Flickr as a supplementary source of environmental information. In previous work (Jeawak et al., 2017), we introduced a method for modelling locations, and hence inferring environmental phenomena, using georeferenced Flickr tags. Our focus was on comparing the predictive power of Flickr tags with that of structured environmental data from more traditional sources for the task of predicting a range of environmental phenomena. We found that Flickr was generally competitive with the structured environmental data for prediction, being sometimes better and sometimes worse. However, combining Flickr tags with the existing environmental data sources consistently improved the results, which suggests that Flickr can indeed be considered as complementary to traditional sources. This method represents each location as a concatenation of two feature vectors: a bag-of-words representation derived from Flickr and a feature vector encoding the numerical and categorical features obtained from the structured dataset. Following on from that approach we experimented in (Jeawak et al., 2019) with the EGEL (Embedding Geographic Locations) model, which learns vector space embeddings of geographic locations by integrating the textual information derived from Flickr with the numerical and categorical information derived from environmental datasets. We found that this approach led to more accurate predictions than the previous approach from (Jeawak et al., 2017) that concatenated the bag of words data with the structured data.

A bag-of-words (BOW) representation is a sparse vector of occurrence counts of individual words. Technically a "bag-of-words" contains frequency scores, such as, in our case, the number of times that an individual tag is used. In our work we convert the counts of tag occurrences to weights, based on a form of Pointwise Mutual Information (PMI) that attaches more significance to tags that are less

*Corresponding author
✉ JeawakSS@cardiff.ac.uk (S.S. Jeawak)
ORCID(s): 0000-0002-4753-8997 (S.S. Jeawak)

[1] http://www.flickr.com

[2] http://expandedramblings.com/index.php/flickr-stats

common and more closely correlated with a particular location. Such representations are often still called bag-of-words models. Note that each dimension of the BOW corresponds to an individual tag of which there can be millions of distinct values. An embedding is a mapping from such a high-dimensional vector representation into a relatively low dimensional representation (e.g. 300-dimensions). Unlike the BOW, the individual dimensions in the vector space embedding typically have no specific meaning. They represent the overall patterns of distance between objects (i.e. locations) by placing semantically similar objects close together in the embedding space.

In this paper, we extend our approach from (Jeawak et al., 2019) by considering a spatiotemporal representation of regions. In particular, we learn a vector space embedding for each geographic region and each month of the year, which allows us to capture environmental phenomena that may depend on monthly or seasonal variation. Apart from extending our main model, we also introduce a new smoothing method to deal with the sparsity of Flickr tags. This is motivated by the fact that when fine grained regions are used and data may be sparse, the number of times that a tag is used in a particular region and month is not a reliable indicator by itself of the relevance of that tag. For evaluation, we consider the problem of predicting climate features and predicting the distribution of species in a given location and a given month. The proposed method has proven to be advantageous, in particular when we have a very small training data set. We also qualitatively evaluate the proposed model by generating similarity maps for a number of selected locations (the details of which can be found in Section 6.5) .

The remainder of this paper is organized as follows. In the next section, we provide a discussion of the related work. Section 3 and Section 4 present our methodology for spatiotemporal modelling using Flickr tags and using structured data respectively. Section 5 then describes our spatiotemporal embeddings model. In Section 6 we provide a detailed discussion about the experimental results as well as the qualitative evaluations. Finally, Section 7 summarizes our conclusions.

2. Related Work

2.1. Spatiotemporal analysis and modelling

Spatiotemporal analysis and modelling has been a major interest in many research areas. Examples include environmental science (Shaddick and Zidek, 2015; McLean, 2018), social science (Brunsdon et al., 2007; Hu et al., 2018), and business (Fotheringham et al., 2015a,b). Fotheringham et al. (2015b) developed a geographical and temporal weighted regression (GTWR) model to account for the variations in time and space when modelling house prices in London from 1980 to 1998. The model is based on a spatiotemporal kernel function using a Gaussian distribution. Similar to our work, they allocated each spatial point to a time interval. However, while they model time on a linear scale, we use a circular scale since our focus is on modelling seasonality. Brunsdon et al. (2007) proposed a spatiotemporal kernel density estimation method (STKDE) which is based on multiplying the spatial kernel function and the temporal kernel function. It is a space-time cube method that extends the 2-dimensional grid used in the spatial kernel to a 3-dimensional cube and computes density values at cube centres with overlapping space-time cylinders. Time was represented on a circular scale that uses a Von Mises distribution as the time kernel. STKDE has shown promising results in many applications such as crime hot-spot detection (Hu et al., 2018) and disease patterns detection (Delmelle et al., 2014). In this paper, we use the STKDE method (Brunsdon et al., 2007) to smooth the distribution of Flickr tags over space and time, as a way of alleviating the sparsity of Flickr tags.

Within a broader context, kernel-based methods have also been used for estimating geographic locations of unstructured text documents. For example, Adams and Janowicz (2012) proposed a method based on kernel density estimation (KDE) and topic modelling to estimate the locations of documents from Wikipedia and a travel blog. Hulden et al. (2015) used kernel density estimation (KDE) to smooth relevant features on a geodesic grid to address the problem of data sparsity. These features were then used for georeferencing text documents. They show that using KDE significantly improves the results. The aim of using KDE in this latter work is similar in spirit to our motivation for using KDE.

2.2. Analyzing Flickr data

Many studies have focused on analyzing Flickr data to extract useful information in domains such as linguistics (e.g. Eisenstein et al. (2010)), geography (e.g. Cunha and Martins (2014); Grothe and Schaab (2009)) and ecology (e.g. Barve (2015); Jeawak et al. (2018)). In the context of ecology, Barve (2015) examined Flickr biodiversity data quality by analysing its metadata and comparing it with ground-truth data, using Snowy owls and Monarch butterflies as a case study. They concluded that Flickr data has the potential to add to the knowledge of these species in terms of geographic, taxonomic, and temporal dimensions, which tends to be complementary to the information contained in other available sources. In (Richards and Friess, 2015), the content of the Flickr photos was analysed manually to assess the quality of cultural ecosystem services and derive useful information to manage Singapore's mangroves. Wang et al. (2013) analysed the visual features of the photographs on Flickr (in an automated way) to observe natural world features such as snow cover and particular species of flowers. In (Zhang et al., 2012) photos from Flickr were used to estimate snow cover and vegetation cover, and to compare these estimations with fine-grained ground truth collected by earth-observing satellites and ground stations. Both the text associated with Flickr photographs and their visual features were used in Leung and Newsam (2012) to perform land-use classification. The approach was evaluated on two university campuses and three land-use classes were considered: Academic, Residential, and Sports. In (Estima et al., 2014; Estima and Painho, 2014), they classified a sample of georeferenced Flickr photos according to CORINE land cover classes. They also evaluated the use of Flickr photos in supporting Land Use/Land Cover (LULC) classification for the city of Coimbra in Portugal and for comparison with Corine Land Cover (CLC) level 1 and level 2 classes. Note that their approach did not use machine learning and the results were evaluated manually by experts. Their results suggest that Flickr photos cannot be used as a single source to achieve this purpose but they could be helpful if combined with other sources of data.

In our previous work (Jeawak et al., 2017), we found that the tags of georeferenced Flickr photos can effectively supplement traditional environmental data in tasks such as predicting climate features, land cover, species occurrence, and human assessments of scenicness. To encode locations, we combined a bag-of-words representation of geographically nearby tags with a feature vector that encodes the associated structured data. We found that the predictive value of Flickr tags is roughly on a par with that of standard commonly available environmental datasets, and that combining both types of information leads to significantly better results than using either of them alone. In (Jeawak et al., 2019), we proposed the EGEL (Embedding GEographic Locations) model that integrates both Flickr and environmental data into low-dimensional vector space embeddings. This model was found to outperform the bag-of-words model for all the evaluation experiments. The main difference between the SPATE (SPAtioTemporal Embeddings) model proposed in this paper and EGEL is that EGEL used a location-based embedding model while here we also take into account the time of year. The resulting model also handles the data sparsity problem in a more robust way than the EGEL model.

2.3. Vector space embeddings

The use of low-dimensional vector space embeddings for representing objects has already proven effective in a large number of applications, including natural language processing (NLP), image processing, and pattern recognition. In the context of NLP, the most prominent example is that of word embeddings (Mikolov et al., 2013; Pennington et al., 2014; Grave et al., 2017), which represent word meaning using vectors of typically around 300 dimensions. These vectors are derived from associated words that occur in the context of the target word. A large number of different methods for learning such word embeddings have already been proposed, including Skip-gram and the Continuous Bag-of-Words (CBOW) model (Mikolov et al., 2013), GloVe (Pennington et al., 2014), and fastText (Grave et al., 2017). They have been applied effectively in many NLP tasks such as sentiment analysis (Tang et al., 2014), part of speech tagging (Qiu et al., 2014; Liu et al., 2016a), and text classification (Lilleberg et al., 2015; Ge and Moh, 2017). The model we consider in this paper builds on GloVe, which was designed to capture linear regularities of word-word co-occurrence. In GloVe, there are two word vectors w_i and $\tilde{w}_j$ for each word in the vocabulary (i.e. the set of words for which we want to learn a vector representation), that are learned by minimizing the following objective:

$$J = \sum_{i,j=1}^{V} f(x_{ij})(w_i.\tilde{w}_j + b_i + \tilde{b}_j - \log x_{ij})^2$$

where x_{ij} is the number of times that word i appears in the context of word j, V is the vocabulary size, b_i is the target word bias, $\tilde{b}_j$ is the context word bias. The weighting function f is used to limit the impact of rare terms. It is defined as 1 if $x > x_{max}$ and as $(\frac{x}{x_{max}})^\alpha$ otherwise, where x_{max} is usually fixed to 100 and α to 0.75. Intuitively, the target word vectors w_i correspond to the actual word representations which we would like to find, while the context word vectors $\tilde{w}_j$ model how occurrences of j in the context of a given word i affect the representation of this latter word. In this paper we will use a similar model, which will however be aimed at learning spatiotemporal cell vectors instead of the target word vectors.

Beyond word embeddings, various methods have been proposed for learning vector space representations from structured data such as knowledge graphs (Bordes et al., 2013; Yang et al., 2015; Trouillon et al., 2016), social networks (Grover and Leskovec, 2016; Wang et al., 2017) and taxonomies (Vendrov et al., 2015; Nickel and Kiela, 2017). The idea of combining a word embedding model with structured information has also been explored by several authors, for example to improve the word embeddings based on information coming from knowledge graphs (Xu et al., 2014; Speer et al., 2017). Along similar lines, various lexicons have been used to obtain word embeddings that are better suited at modelling, for example, sentiment (Tang et al., 2014) and antonymy (Ono et al., 2015). The method proposed by (Liu et al., 2015) imposes the condition that words that belong to the same semantic category should be closer together than words from different categories, which is somewhat similar in spirit to how we will model categories in our model.

2.4. Embedding Spatiotemporal information

The problem of representing geographic locations using embeddings has also attracted some attention. An early example is (Saeidi et al., 2015), which used principal component analysis and stacked autoencoders to learn low-dimensional vector representations of city neighbourhoods based on census data. They use these representations to predict attributes such as crime, which is not included in the given census data, and find that in most of the considered evaluation tasks, the low-dimensional vector representations lead to more faithful predictions than the original high-dimensional census data.

Some existing works combine word embedding models with geographic coordinates. For example, in (Cocos and Callison-Burch, 2017) an approach is proposed to learn word embeddings based on the assumption that words which tend to be used in the same geographic locations are likely to be similar. Note that their aim is dual to our aim in this paper: while they use geographic location to learn word vectors, we use textual descriptions to learn vectors representing geographic locations.

Several methods also use word embedding models to learn representations of Points-of-Interest (POIs) that can be used for predicting user visits (Feng et al., 2017; Liu et al., 2016b; Zhao et al., 2017). These works use the machinery of existing word embedding models to learn POI representations, intuitively by letting sequences of POI visits by a user play the role of sequences of words in a sentence. In other words, despite the use of word embedding models, many of these approaches do not actually consider any textual information. For example, in (Liu et al., 2016b) the Skip-gram model is utilized to create a global pattern of users' POIs. Each location was treated as a word and the other locations visited before or after were treated as context words. They then use a pair-wise ranking loss (Weston et al., 2010) which takes into account the user's location visit frequency to personalize the location recommendations. The methods of (Liu et al., 2016b) were extended in (Zhao et al., 2017) to use a temporal embedding and to take more account of geographic context, in particular the distances between preferred and non-preferred neighbouring POIs, to create a "geographically hierarchical pairwise preference ranking model". Similarly, Yang and Eickhoff (2018) developed a method for modeling places, neighbourings, and users from social media check-ins. They treat the check-ins as sentences to generate the embeddings which encode the geographical, temporal, and functional (e.g. College & University, Event, Residence [3]) aspects. In (Yao et al., 2017) the CBOW model was trained with POI data. They ordered POIs spatially within the traffic-based zones of urban areas. The ordering was used to generate characteristic vectors of POI types. Zone vectors, represented by averaging the vectors of the POIs contained in them, were then used as features to predict land use types. Yan et al. (2017) proposed a method that uses the Skip-gram model to represent POI types, based on the intuition that the vector representing a given POI type should be predictive of the POI types found in nearby places of that type. In the CrossMap method, Zhang et al. (2017a) learned unsupervised embeddings for spatio-temporal hotspots obtained from social media data of locations, times and text. In one form of embedding, intended to enable reconstruction of records, neighbourhood relations in space and time were encoded by averaging hotspots in a target location's spatial and temporal neighborhoods. They also proposed a graph-based embedding method with with different nodes for modelling location, time and text. The concatenation of the location, time and text vectors was then used to predict peoples' activities in urban environments. In another work, Zhang et al. (2017b) proposed the ReAct model, which is similar to CrossMap. However, while the CrossMap model is unsupervised and handles static data, ReAct is a semi-supervised model and handles continuous online data to learn the activity models.

In NLP research, embedding methods have been used to measure the language variation across geographical regions as well as over time (Bamman et al., 2014; Kim et al., 2014; Kulkarni et al., 2016; Phillips et al., 2017; Hovy and Purschke, 2018). For instance, Bamman et al. (2014) and Kulkarni et al. (2016) present methods to learn geographically situated word embeddings from geo-tagged tweets. They used cosine similarities between the generated embeddings to measure the spatial variation of language across English speaking countries. Hovy and Purschke (2018) used the Doc2Vec method (Le and Mikolov, 2014) to learn document embeddings from online posts in German speaking regions. These embeddings have been used to study language variation in German. To study the temporal variation of language, Kim et al. (2014), among others, trained the Skip-gram model on text from the Google Books corpus for the period from 1900 to 2009. They also used cosine similarity to measured the change in word meaning between the embeddings of the same words learned in different time periods. Phillips et al. (2017) used the CBOW model to learn spatiotemporal embeddings from geo-tagged tweets. They first split the data into 8 hour windows (i.e. the temporal granularity) for each separate country (i.e. the spatial granularity). For each time window they then trained a joint embedding using tweets from all countries and used it to initialize the country specific embeddings.

Despite the considerable progress that has been made on embedding social media data, the problem of embedding Flickr tags has so far received very limited attention. To the best of our knowledge, (Hasegawa et al., 2018) is the only work that generated embeddings for Flickr tags. However, their focus was on learning embeddings that capture word meaning which has been evaluated on word similarity tasks.

Our work is different from all these studies, as our focus is on spatiotemporal embeddings based on text descriptions (in the form of Flickr tags), along with numerical and categorical features from environmental datasets.

3. Spatiotemporal Modelling Using Flickr tags

3.1. Data Acquisition

The first source of information that we consider in this work is a collection of georeferenced time-stamped social media postings from the photo-sharing website Flickr[4]. We used the Flickr API to collect the metadata of approximately 12 million geocoded Flickr photographs within the UK and Ireland (which is the region our experiments will focus on), all of which were uploaded to Flickr from January 2004 to September 2015. Our analysis in this paper will focus on the tags in addition to the spatial and temporal information associated with these photographs. We only considered the photographs in which the difference between the time at which the photo was taken and the upload time is less than 6 months to, as much as possible, avoid photographs with an incorrect time stamp.

[3] https://developer.foursquare.com/docs/resources/categories

[4] http://www.flickr.com

3.2. Data Preprocessing

With the objective of using Flickr tags for spatiotemporal modelling, we split the target spatial area into $10km \times 10km$ grid cells. Furthermore, we discretize the time stamps with a granularity of 1 month. We thus view the overall dataset as 12 separate grid layers, each layer corresponding to a month of the year. Thus there are 12 instances for each spatial cell as illustrated in Figure 1. The choice of the $10km \times 10km$ spatial granularity and the one-month temporal granularity is to balance between resolution and computation time. Let $c_1, ..., c_n$ be the spatiotemporal grid cells, each represented by a triple (lat, lon, m) where lat is the latitude coordinate of the centre of c, lon is the longitude coordinate of the centre of c, and m is the month of the year. We associate each such a cell with a histogram of Flickr tags, reflecting how many times each tag has been added to a photograph whose coordinates and time stamp fall within the cell. However, to reduce the impact of bulk uploading, following Van Canneyt et al. (2013b), we count a tag occurrence only once for all photos by the same user at the same location and on the same date. Let *f(t,c)* be the number of times tag t (from the set of all tags T) occurs in the cell c. We then use Positive Pointwise Mutual Information (PPMI) to weight how strongly tag t is associated with cell c. In particular, PPMI compares the actual number of occurrences with the expected number of occurrences, considering how many tags occur overall in c and how common the tag t is. Then the *PPMI* weight is given by:

$$PPMI(t, c) = \max\left(0, \log\left(\frac{P(t, c)}{P(c)P(t)}\right)\right) \tag{1}$$

where:

$$P(t, c) = \frac{f(t, c)}{N} \qquad P(t) = \frac{\sum_{c' \in C} f(t, c')}{N}$$

$$P(c) = \frac{\sum_{t' \in T} f(t', c)}{N} \qquad N = \sum_{t' \in T} \sum_{c' \in C} f(t', c')$$

Each cell c can thus be represented as a sparse vector $v_f(c)$ which is defined as $(PPMI(t_1, c), ..., PPMI(t_k, c))$, where $t_1, ..., t_k$ is an enumeration of the tags in T.

3.3. Tag Selection

Our aim here is to select tags whose occurrence is correlated with specific times of the year (e.g. Summer) or with photos that occur in particular geographic regions (e.g. forests). When constructing the feature representation $v_f(c)$, we then only consider those tags that have been selected. The aim of this step is to reduce the impact of tags that are not relevant for modelling the environment, such as tags which only relate to a given individual or a group of users. Intuitively, to determine whether a given tag is time and/or location specific, we assess to what extent the distribution of its occurrences across all spatiotemporal cells diverges from the overall distribution of all tag occurrences. To this end, we use a method based on Kullback-Leibler (KL) divergence, which was previously found to be effective in Van Laere et al. (2014) and Jeawak et al. (2019). In particular, we select those tags $T_{KL} \subseteq T$ which maximize the following score:

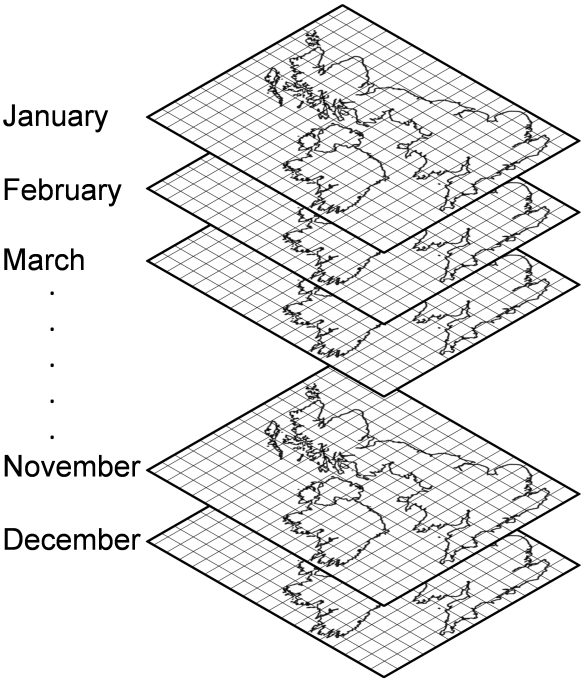


Figure 1: Spatiotemporal grid cells.

$$D_{KL}(t) = \sum_{i=1}^{n} Ps(c_i|t) \log \frac{Ps(c_i|t)}{Q(c_i)} \tag{2}$$

where $Ps(c_i|t)$ is the probability that a photo with tag t has a location and time in c_i and $Q(c_i)$ is the probability that an arbitrary tag occurrence is assigned to a photo in c_i. Since $Ps(c_i|t)$ has to be estimated from a small number of tag occurrences, it is estimated using Bayesian smoothing as follows:

$$Ps(c_i|t) = \frac{f(t, c_i) + \delta \cdot Q(c_i)}{\left(\sum_{j=1}^{n} f(t, c_j)\right) + \delta}$$

where n is the total number of cells, δ is a parameter controlling the amount of smoothing, which will be tuned in the experiments. $Q(c_i)$ is estimated using maximum likelihood, as more data is available for estimating these probabilities:

$$Q(c_i) = \frac{1}{N} \sum_{t' \in T} f(t', c_i)$$

We will use the notation $v_{KL}(c)$ for the sparse vector representation of cell c encoding the *PPMI* weight of those tags in T_{KL} only.

3.4. Spatiotemporal Smoothing

The vector representation $v_{KL}(c)$ encodes which tags are most strongly correlated with the spatiotemporal grid cell c.

However, these scores are computed from sometimes very limited amounts of data, and for some cells we may not have any photographs at all. To tackle this problem, we used kernel density estimation to smooth the *PPMI* weight of each tag in T_{KL} over a larger region. For this purpose, we used the spatiotemporal kernel density estimation method that was introduced in Brunsdon et al. (2007). In particular, we define the smoothed weight of tag t in cell c as follows:

$$KDE(t,c) = \frac{\hat{s}(t,c)}{\max\left(\hat{s}(t,c)\right)} \tag{3}$$

where n_t is the number of cells c with tag t, h_s is the spatial smoothing bandwidth of tag t, and h_m is the temporal smoothing bandwidth of tag t. The reason for normalising the *KDE* value is to keep the weight of all the tags within the same range and avoid the impact of the dominant tags. The $\hat{s}(t,c)$ value is computed as:

$$\hat{s}(t,c) = \sum_{i=1}^{n_t} PPMI(t,c_i) \cdot K_s\left(\Lambda_{lat^i}, \Lambda_{lon^i}\right) \cdot K_m\left(\Lambda_{m^i}\right) \tag{4}$$

where $\Lambda_{lat^i} = \frac{c_{lat}-c_{lat^i}}{h_s}$, $\Lambda_{lon^i} = \frac{c_{lon}-c_{lon^i}}{h_s}$, and $\Lambda_{m^i} = \frac{c_m-c_{m^i}}{h_m}$. Here c_{lat}, c_{lon} and c_m are respectively the latitude, longitude and the month of cell c, c_{lat^i}, c_{lon^i} and c_{m^i} are respectively the latitude, longitude and the month of cell c_i. As the spatial kernel function K_s, we use a Gaussian distribution (Silverman, 1986) given by:

$$K_s(c_{lat}, c_{lon}) = \frac{1}{2\pi}\exp\left(-\frac{(c_{lat}-c_{lat^i})^2 + (c_{lon}-c_{lon^i})^2}{2h_s^2}\right)$$

As the temporal kernel K_m, we use a von Mises distribution (Taylor, 2008) which is a continuous probability distribution on the circle. The Von Mises distribution was chosen because of its wrap-around property (it is sometimes called circular Gaussian) which is well suited to the cyclic nature of the months of the year representation. Here, we first encode months using values in $\{0,...,11\}$, then the month value is mapped to its corresponding point on the circle by:

$$\theta(c_m) = \frac{2\pi c_m}{12}$$

Now 'January' is represented as $\frac{pi}{6}$, 'February' is represented as $\frac{pi}{3}$ and so on, as explained in Figure 2.
The von Mises distribution is computed by:

$$K_m(\theta(c_m)) = \frac{1}{2\pi I_0(h_m)}\exp\left(h_m\cos(\theta - \Theta)\right)$$

where I_0 is the modified Bessel function of order 0.

Finally, to model the spatiotemporal grid cell c using Flickr tags, we consider a vector $v_{KDE}(c)$ encoding the smoothed weight of all the tags $\{t_1...t_{nt}\} \in T_{KL}$ which is defined as $(KDE(t_1,c),...,KDE(t_{n_t},c))$. This vector $(v_{KDE}(c))$ will be used to train the proposed embeddings model in Section 5.

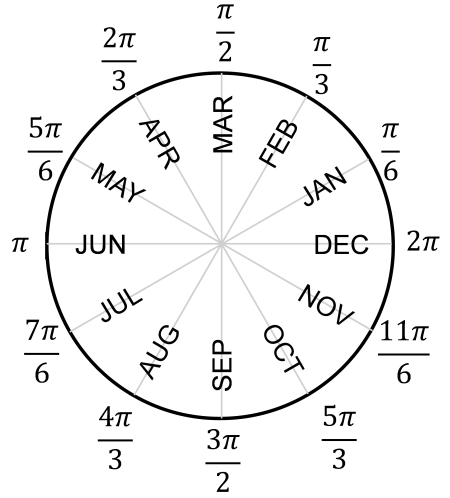


Figure 2: The representation of the months as circular data.

3.5. Bandwidth Selection

The critical parameter in any kernel based method is the selection of the optimal bandwidth. The variables h_s and h_m are of key importance and their values are generally considered to be more important than the type of the kernel itself. In general, large values lead to over-smoothing while small values lead to under-smoothing. Various methods have been developed for selecting the optimal kernel bandwidth. In this paper, we compare experimentally the performance of three of the most widely used methods.

1. The rule of thumb (Silverman, 1986) is a simple and fast method. It estimates a fixed kernel bandwidth based on the data driven scale of the distribution which is defined as:

$$h = \hat{\alpha}\left(\frac{n*(d+2)}{4}\right)^{-1/(d+4)} \tag{5}$$

   where n is the size of the data, d is the number of dimensions, and $\hat{\alpha}$ is the data standard deviation. Here we need to estimate two different bandwidths (the spatial and the temporal bandwidths). They are both estimated using Equation 5, however, for estimating the temporal bandwidth h_m, d is equal to 1 and $\hat{\alpha}$ is the circular standard deviation. For estimating the spatial bandwidth h_s, d is equal to 2 and $\hat{\alpha} = \frac{s_1+s_2}{2}$ where s_1 and s_2 are the standard deviations of the latitude and the longitude coordinates respectively.
2. The adaptive kernel bandwidth (Abramson, 1982; Brunsdon, 1995) is based on the idea of making the value of h vary between different regions according to the local density. In particular, a wider bandwidth is selected for regions with low density while a narrower bandwidth selected for regions with high density. It is usually achieved by the following steps. Firstly, compute a pilot estimate of $\hat{s}(t,c)$ (Equation 4) using the

fixed bandwidth as described above in Equation 5. This estimate is used to give an overall approximation of the smoothed value of the data. Secondly, compute a local bandwidth scalar, which is computed by:

$$b_c = \sqrt{\frac{g}{\hat{s}(t,c)}}$$

where g is the geometric mean of $\hat{s}(t,c_1), ..., \hat{s}(t,c_n)$, which is given by:

$$g = \left(\prod_{i=1}^{n} \hat{s}(t,c_i) \right)^{1/n}$$

Finally, the adaptive local bandwidths are given by $h_{s(c)} = h_s \cdot b_c$ and $h_{m(c)} = h_m \cdot b_c$ which can be used in Equation 4 to make the final estimation for tag t.

3. The leave one out kernel estimator (Bowman, 1984) is based on the idea of selecting the kernel bandwidth estimator that minimizes the mean integrated square error (MISE) (Seaman and Powell (1996)) given by:

$$MISE = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{s}(t,c_i) - p(t,c_i))^2}{p(t,c_i)} \tag{6}$$

where $\hat{s}(t,c_i)$ is the estimated density of tag t at the grid cell c_i after removing the cell c_i from the data. $p(t,c_i)$ is the probability of the $PPMI$ weight of tag t at the grid cell c_i (i.e. the true density), which is computed as:

$$p(t,c_i) = \frac{PPMI(t,c_i)}{\sum_{c' \in C} PPMI(t,c')}$$

And $\hat{s}_{-i}(t,c_i)$ is computed here as:

$$\frac{\sum_{j=1 j \neq i}^{n_t} PPMI(t,c_j) K_s(\Lambda_{lat^{ij}}, \Lambda_{lon^{ij}}) K_m(\Lambda_{m^{ij}})}{\sum_{j=1 j \neq i}^{n} K_s(\Lambda_{lat^{ij}}, \Lambda_{lon^{ij}}) K_m(\Lambda_{m^{ij}})} \tag{7}$$

The optimal bandwidths h_s and h_m that minimize Equation 6 can be used to smooth the tag t distribution over all the spatiotemporal grid cells in Equation 4.

4. Spatiotemporal Modelling Using Structured Environmental Data

There is a wide variety of structured scientific data that can be used for modelling the environment. In this section, we give an overview of the structured datasets that we will use in our experiments, and we explain how these datasets are used to generate a feature vector for each spatiotemporal cell c. We used the following external datasets as sources of numerical features:

- Monthly average of temperature, precipitation, solar radiation, wind speed and water vapour pressure, all of which are obtained from WorldClim[5].
- Elevation, obtained from the Digital Elevation Model over Europe (EU-DEM)[6].
- Population, obtained from the European Population Map 2006[7].

Several of the considered datasets have a resolution which is finer than our $10km \times 10km$ grid cells. To this end, we look up the feature values at 100 locations, distributed uniformly within the grid cell. To obtain a feature vector for the spatiotemporal grid cell c representing these numerical features, we first average these 100 values for each numerical feature across the grid cell. Then we normalise these features values using the standard z-score.

In addition, we used the following datasets as sources of categorical features:

- Land cover type, obtained from CORINE Land Cover 2006[8]. This dataset refers to land cover categories at three levels of granularity: a top level with 5 classes, an intermediate level with 15 classes and a detailed level with 44 classes.
- Soil type, obtained from SoilGrids[9], which classifies locations into 116 types of soil.

The categorical features are represented as a vector, encoding for each of the categories what percentage of the grid cell (i.e. the average of the 100 locations) belongs to that category.

Apart from the features from these external datasets, the geographic coordinates and time stamp of the cell c are clearly also important structured features, which should be included in the feature vector describing a spatiotemporal cell. For the spatial features, each grid cell c has been represented by the normalised coordinate values $norm(lat,c) = \frac{lat - min(latitude)}{max(latitude) - min(latitude)}$ and $norm(lon,c) = \frac{lon - min(longitude)}{max(longitude) - min(longitude)}$ where lat and lon are the latitude and longitude coordinates of the centre of the grid cell c respectively, and $max(latitude)$, $max(longitude)$, $min(latitude)$ and $min(longitude)$ are the maximum and minimum latitude and longitude over the study area. The reason for normalising these features is to ensure that they are within the same range as the other features. Note that we have also tried projecting the latitude and longitude coordinates into three-dimensional geographic coordinates, but that gave worse results. Finally, the month m corresponding to the cell c is represented as the coordinates $(cos(\theta(m)), sin(\theta(m)))$ of that month, as before (see Figure 2).

We will use the notation $v_s(c)$ for the feature vector representation of cell c encoding all the above mentioned structured features.

[5] http://worldclim.org

[6] http://www.eea.europa.eu/data-and-maps/data/eu-dem

[7] http://data.europa.eu/89h/jrc-luisa-europopmap06

[8] http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-2

[9] https://www.soilgrids.org

5. Spatiotemporal Embeddings

Our aim in this paper is to learn a low-dimensional vector space embedding of a set of spatiotemporal cells C. This representation will allow us to combine the textual information derived from Flickr with the numerical, categorical, spatial, and temporal information in an efficient way. Thus the ecological information can be effectively captured by the predictive model. The proposed embeddings model has the following objective function:

$$J = (1-2\alpha-2\beta)J_{tags}+\alpha(J_{nf}+J_{cat})+\beta(J_{spatial}+J_{temp}) \quad (8)$$

where $\alpha, \beta \in [0, 1]$ are parameters to control the importance of each component in the model with $2\alpha + 2\beta < 1$. The components J_{tags}, J_{nf}, J_{cat}, $J_{spatial}$ and J_{temp} intuitively encode the information we have about the spatiotemporal cells from the different sources. The objective function J thus encodes the available information in the form of an optimization problem. In particular, our goal is to learn vector representations for the spatiotemporal cells which minimize J.

Component J_{tags} will be used to constrain the representation of the cells based on their textual description (i.e. Flickr tags), J_{nf} will be used to constrain the representation of the cells based on their numerical features, J_{cat} will impose the constraint that cells belonging to the same category should be close together in the space, $J_{spatial}$ will be used to constrain the representation of the cells based on their spatial feature (i.e. the latitude and longitude coordinates), and J_{temp} will be used to constrain the representation of the cells based on their temporal feature (i.e. months of the year). The components J_{nf} and J_{cat} share the same weight (α) as they have the same key importance in our model and a relatively similar number of features (i.e. similar impact on the embeddings model). The components $J_{spatial}$ and J_{temp} share the same weight (β) for the same reasons. However, the component J_{tags} has a different weight as it involves a larger number of features, these features are of a different nature and their relative importance may also be quite different (e.g. the number of occurrences of a single tag is likely to be less important than the land cover class).

Tags based embedding. We now want to find a vector $v_{emb}(c) \in V$ for each spatiotemporal grid cell c. The component J_{tags} intuitively encodes the requirement that we want spatiotemporal cells whose associated Flickr tag distributions are similar to be represented by similar vectors. This is achieved by requiring that the scores $KDE(t_j, c)$ for each tag t_j can be predicted from the vector representation of the cell c. To this end, we use a close variant of the GloVe model, where tag occurrences are treated as context words of the spatiotemporal cell. In particular, with each cell c we associate a vector $v_{emb}(c)$ and with each tag t we associate a vector $\tilde{w_t}$ and a bias term $\tilde{b_t}$, and consider the following objective which is illustrated in Figure 3:

$$J_{tags} = \sum_{c \in C} \sum_{t_j \in T} (v_{emb}(c) \cdot \tilde{w_{t_j}} + \tilde{b_{t_j}} - KDE(t_j, c))^2$$

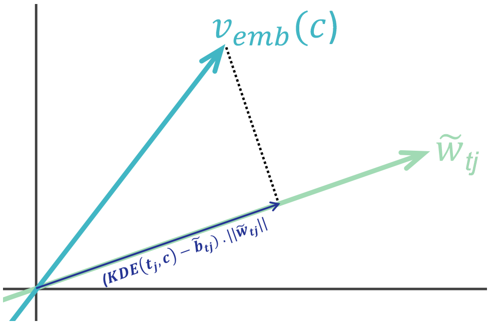

$v_{emb}(c)$
$\widetilde{w}_{tj}$
$(KDE(t_j, c) - \widetilde{b}_{tj}) \cdot ||\widetilde{w}_{tj}||$

Figure 3: The geometric intuition of tags based embeddings.

Note how tags play the role of the context words in the GloVe model, while instead of learning target word vectors we now learn vectors for spatiotemporal cells. In contrast to GloVe, our objective does not directly refer to co-occurrence statistics, but instead uses the $KDE(t_j, c)$ scores.

Numerical features based embedding. Numerical features can be treated similarly to the $KDE(t_j, c)$ scores. In particular, for each numerical feature f_k we consider a vector $\tilde{w_{f_k}}$ and a bias term $\tilde{b_{f_k}}$, and the following objective:

$$J_{nf} = \sum_{c \in C} \sum_{f_k \in NF} (v_{emb}(c) \cdot \tilde{w_{f_k}} + \tilde{b_{f_k}} - score(f_k, c))^2$$

where NF is the set of all numerical features and $score(f_k, c)$ is the value of feature f_k for cell c, after z-score normalization.

Categorical features based embedding. For the categorical features, we impose the constraint that cells belonging to the same category should be close together in the space. In particular, we represent each category type cat_l as a vector w_{cat_l}, and consider the following objective:

$$J_{cat} = \sum_{c \in C} \sum_{cat_l \in L} (v_{emb}(c) - w_{cat_l})^2$$

Spatial features based embedding. Latitude and longitude coordinates can be incorporated in the same way as the numerical features. However, we treat them as a separate constraint because this allows us to tune the importance of the geographic location of a grid cell c, relative to the numerical and categorical features, based on how we choose the parameters α and β. Therefore, for $s_c \in \{lat, lon\}$, we consider a vector $\tilde{w_{s_c}}$ and a bias term $\tilde{b_{s_c}}$, and the following objective to compute $J_{spatial}$:

$$\sum_{c \in C} \sum_{s_c \in lat, lon} (v_{emb}(c) \cdot \tilde{w_{s_c}} + \tilde{b_{s_c}} - norm(s_c, c))^2$$

Temporal features based embedding. We represent the temporal features, specifically the months of the year, as equidistant points on the unit circle (as shown in Figure 2). To encode temporal information in the embedding, we assume that there is a linear transformation that maps the

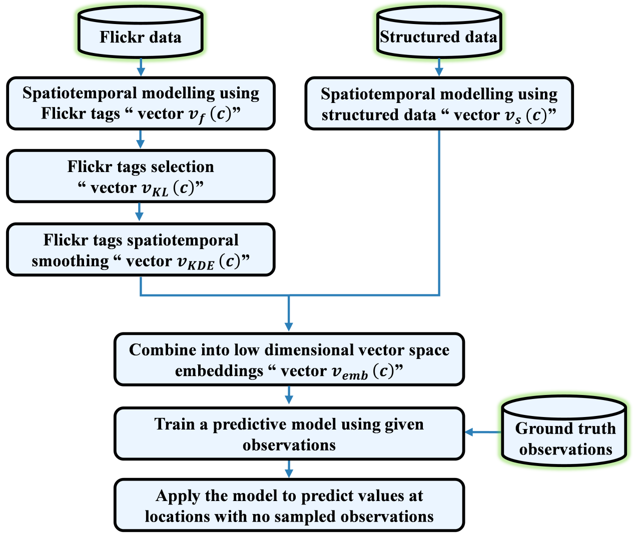

Flickr data
Structured data
Spatiotemporal modelling using Flickr tags " vector $v_f(c)$"
Spatiotemporal modelling using structured data " vector $v_s(c)$"
Flickr tags selection " vector $v_{KL}(c)$"
Flickr tags spatiotemporal smoothing " vector $v_{KDE}(c)$"
Combine into low dimensional vector space embeddings " vector $v_{emb}(c)$"
Train a predictive model using given observations
Ground truth observations
Apply the model to predict values at locations with no sampled observations

Figure 4: The spatiotemporal embeddings (SPATE) model.

vector representations of the spatiotemporal cells onto a 2-dimensional plane, such that all cells from a given month are (approximately) projected onto the vector representation of that month. This is similar to how we handle the spatial features, where the two linear lat/lon constraints could also be seen together as mapping the grid cells onto a 2-dimensional plane such that the projection reflects their geographic location. To formalize this constraint, we encode each month m_i by a 2-dimensional vector $\tilde{w_m}$ representing the coordinates $(cos(\theta(m)), sin(\theta(m)))$ of m_i on the temporal circle. We define a projection matrix P as a $2 \times n$ matrix that maps the spatiotemporal cell vector $v_{emb}(c)$ into 2-dimensional space and a 2-dimensional bias term $\tilde{b_m}$, and consider the following objective:

$$J_{temp} = \sum_{c \in C} ||v_{emb}(c).P + \tilde{b_m} - \tilde{w_m}||^2$$

6. Experimental Evaluation

In this section we will formally evaluate our proposed SPAtioTemporal Embeddings (SPATE)[10] model. The full model is illustrated in Figure 4. This figure shows how the Flickr tags representation from Section 3 are combined with the structured information from Section 4 to represent the spatiotemporal cells C that can be used to predict values at un-sampled locations.

We will start this section by evaluating the bandwidth selection methods that described in Section 3.5 and choose the best method for our problem. Then we will define our experimental setting and the proposed baseline methods. Subsequently, we will introduce our experiments and results with a detailed discussion. Finally, we will qualitatively evaluate our generated vectors.

6.1. Selecting the optimal bandwidth for each tag

We evaluate the performance of the considered bandwidth selection methods from Section 3.5 in term of MISE (see Equation 6) on a randomly selected sample of 100 cells for each tag in T_{KL}. For the leave-one-out kernel estimator method, we considered the range {2, 1, 0.5, 0.25, 0.125, 0.05, 0.025, 0.0125, 0} in latitude/longitude degrees for the spatial bandwidth h_s value and the range $\{2\pi, \pi, \pi/2, \pi/6, 0\}$ for the temporal bandwidth h_m value. The choice of these two ranges was found to be reasonable for most of the tags based on a small set of initial experiments. Note that a spatial bandwidth of value 0 would mean only temporal smoothing is applied, and vice versa if the temporal bandwidth is set to 0.

The results are summarized in Figure 5. We found that the fixed bandwidth selected by the rule-of-thumb method works reasonably well for tags with a uni-modal distribution (e.g. the name of a city). However, for tags with a multi-modal distribution (e.g. supermarket, beach and rain), it leads to a significant over-estimation of the bandwidth. The adaptive kernel bandwidth method performs better than the fixed bandwidth estimator in many cases, especially those with multi-modal distribution, but it is computationally expensive. However, we found that the leave-one-out kernel estimator method outperforms both of them. Therefore, in the remaining experiments, we will use the spatial and temporal bandwidths (h_s and h_m) estimated from the leave-one-out kernel estimator method as the optimal bandwidths. In particular, when applying KDE (see Equation 3), for each tag we use the specific bandwidth parameters that were selected with this method. In this way, we can choose the best spatial and temporal bandwidths for each tag. In particular, tags which refer to a very localised spatial area (e.g. tescoextra) will be assigned a small bandwidth, whereas tags that refer to a broader region (e.g. lakedistrictnationalpark) will receive a larger bandwidth. Note that tags that refer to a specific event, the optimal bandwidth that is selected will often be 0, reflecting the fact that smoothing in such cases would hurt the performance.

6.2. Experimental Settings

In all experiments, we use Support Vector Machines (SVMs) for classification problems and Support Vector Regression (SVR) for regression problems. In both cases, we used the SVM^{light} implementation[11] Joachims (1998). We also experimented with a multilayer perceptron (MLP), to check whether using a different predictive model might affect the results. The choice of SVM and MLP is motived by the fact that these models are widely used for environmental modelling, including for predicting species occurrence (Muñoz-Mas et al. (2017); Drake et al. (2006); Guo et al. (2005)) and predicting climate features (Aghelpour et al.

[10] The SPATE source code is available online at https://github.com/shsabah84/SPATE-model.git.

[11] http://www.cs.cornell.edu/people/tj/svm_light/

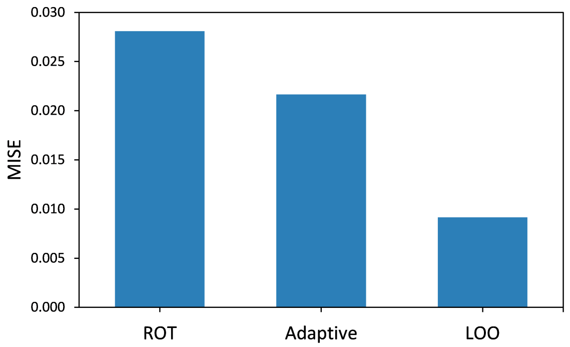


Figure 5: The average MISE of all the considered tags when using the rule of thumb (ROT), the adaptive kernel bandwidth (Adaptive), and the leave one out kernel estimator (LOO).

(2019); Salcedo-Sanz et al. (2016); Kashani and Dinpashoh (2012); Radhika and Shashi (2009)).

For all experiments, we randomly split the set of spatiotemporal grid cells C into one-third for testing and two-thirds for training and tuning. To evaluate the impact of the training data size on the model performance, we experimented with 1%, 10% and 100% of the training and tuning set. Each time we hold out 10% of the considered set for tuning the parameters and use the rest for training. In fact, the setting with a small amount of training data makes the problem more challenging and provides additional insight into the performance of our proposed model.

To compute KL divergence, the smoothing parameter δ was selected from $\{100, 1000, 10000\}$ based on the tuning data. Table 1 shows the 10 tags with highest KL divergence weight resulting from these smoothing values. Clearly, using $\delta = 100$ gives a set of tags that are specifically related to small geographic regions and/or particular times, while using $\delta = 1000$ gives a set of tags describing larger regions. Using $\delta = 10000$ gives a set of more general tags or even more general regions, as well as names of well known cities. We select the top 100 000 tags from the ranking with $\delta = 1000$ where it gave us the best results based on initial experiments. However, for a grid cell c, we only consider those tags t for which $KDE(t|c) > \frac{1}{3}$ for computational reasons.

All embedding models are learned with an Adagrad optimizer, which is used to minimize the objective function using 30 iterations and an initial learning rate of 0.5. The number of dimensions is chosen for each experiment from $\{10, 50, 300\}$ based on the tuning data. For the parameters of our model in Equation 8, we considered values of α from $\{0.01, 0.02, 0.04, 0.06, 0.08, 0.1\}$ and we considered values of β between 0 and 1 with an increment of 0.05. While we chose the best values of the parameters for each experiment separately, based on the tuning data, we noticed that consistently good results were obtained when using $\alpha = 0.04$ and $\beta = 0.45$. Note that we tune all parameters with respect to the F1 score for the classification tasks and Spearman ρ for the regression tasks.

$\delta = 100$	$\delta = 1000$	$\delta = 10000$
struy	islay	cambridge
may	gairloch	bournemouth
tiree	ashford	chester
strathglass	longleat	york
march	orkney	cornwall
waterfordhalf2009	sywell	cardiff
stmelliongolfclub	braintree	sheffield
bawdeswell	snetterton	lakedistrict
stkilda	popham	oxford
helmsdale	dungeness	norfolk

Table 1
Top 10 Flickr tags in terms of KL divergence.

6.3. Variants and Baseline Methods

For formal evaluation, we will compare our proposed SPATE model with the following main baseline representations:

- Structured uses the feature vector $v_s(c)$ modelling the structured information from Section 4.
- Flickr uses the KDE-based feature vector $v_{KDE}(c)$ modelling Flickr tags from Section 3.4.
- Structured + Flickr uses the combination of both structured data and Flickr data by concatenating the vectors $v_s(c)$ and $v_{KDE}(c)$.

To evaluate the impact of the spatiotemporal smoothing on Flickr tag representation, we will consider the following variants:

- Flickr-noKDE uses the PPMI-based feature vector $v_f(c)$ modelling Flickr tags from Section 3.2 (i.e. without including the tag selection and spatiotemporal smoothing steps).
- Flickr(1BW) uses the KDE-based feature vector modelling Flickr tags from Section 3.4. However, here we select the value of the bandwidths h_s and h_m that minimize the average MISE over all the considered tags when computing KDE weight (i.e. using the same bandwidths for all the tags). This variant will thus allow us to assess the effectiveness of using tag-specific bandwidth values.

6.4. Experimental Results

We consider two tasks to evaluate our proposed SPATE model: predicting species distribution and predicting climate related features.

6.4.1. *Predicting Species Distribution*

For this task, we use ground truth data from the National Biodiversity Network Atlas (NBN Atlas)[12]. The NBN is a collaborative project committed to making biodiversity information available via the NBN Atlas. This dataset covers

[12] NBN Atlas occurrence download at http://nbnatlas.org. Accessed 19 April 2018.

	1%			10 %			100 %		
	Prec	Recall	**F1**	Prec	Recall	**F1**	Prec	Recall	**F1**
Structured	0.424	0.246	0.311	0.501	0.345	0.409	0.525	0.422	0.468
Flickr-noKDE	0.091	0.005	0.010	0.141	0.022	0.038	0.388	0.034	0.063
Flickr-1BW	0.400	0.342	0.369	0.469	0.406	0.435	0.494	0.415	0.451
Flickr	0.436	0.373	0.402	0.529	0.454	0.489	0.631	0.466	0.536
Structured + Flickr	0.448	0.384	0.414	0.536	0.465	0.498	0.629	0.474	0.541
SPATE	0.485	0.423	**0.452**	0.540	0.476	**0.506**	0.610	0.487	**0.542**

Table 2
Results for predicting the monthly distribution of 50 species across the UK and Ireland using SVM. The percentages refer to the proportion of the training and tuning data set that was used.

	1%			10 %			100 %		
	Prec	Recall	**F1**	Prec	Recall	**F1**	Prec	Recall	**F1**
Structured	0.443	0.313	0.367	0.557	0.323	0.409	0.628	0.388	0.48
SPATE	0.461	0.382	**0.418**	0.508	0.458	**0.482**	0.585	0.512	**0.546**

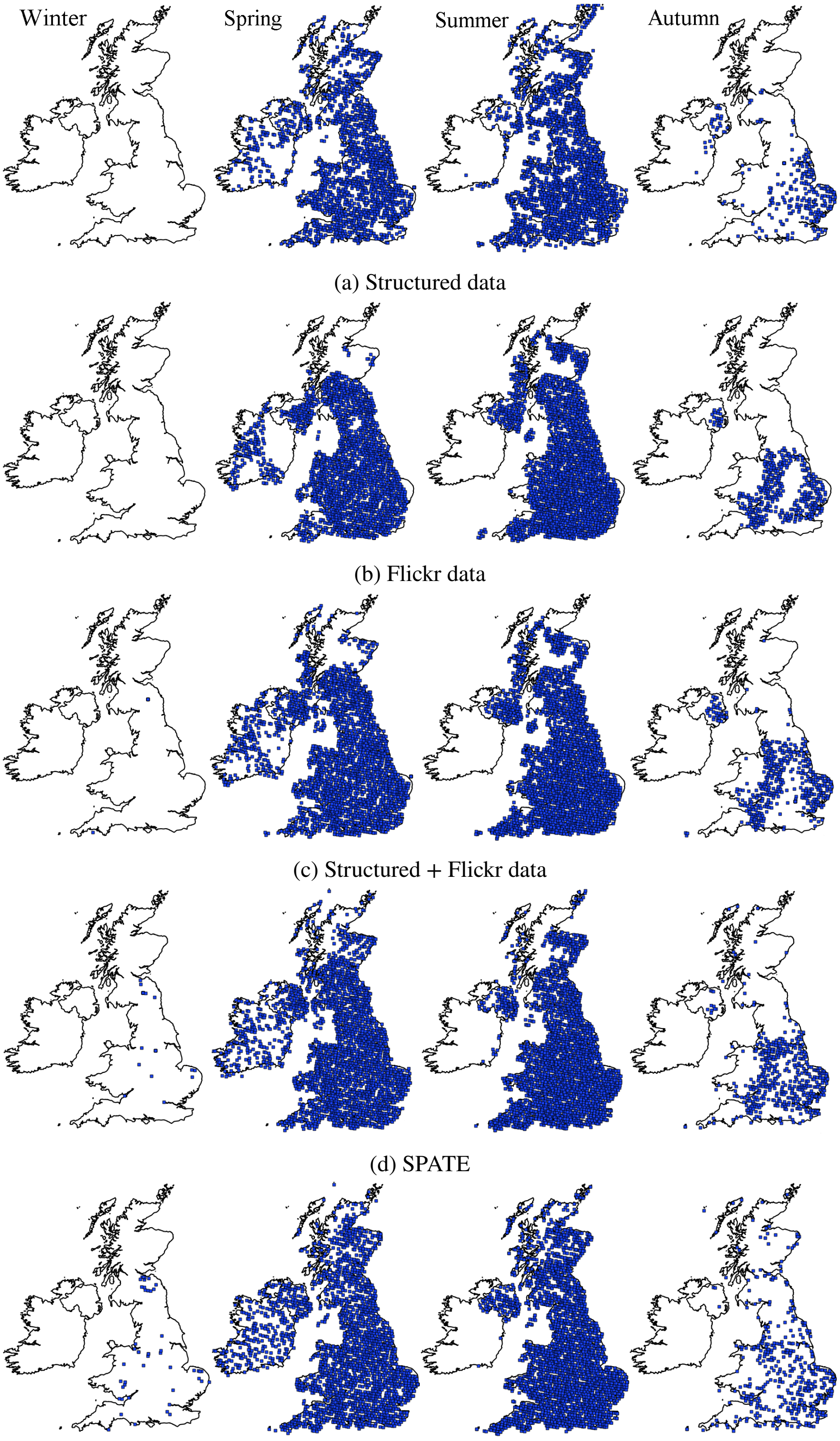
Table 3
Results for predicting the monthly distribution of 50 species across the UK and Ireland using MLP.

the UK and Ireland. We focused our evaluation on a random sample of 50 birds, each of which has at least 1000 observations in the NBN Atlas. This restriction to species with a sufficient number of observations is necessary to ensure that the ground truth is sufficiently reliable. Note that even species with a large number of observations may sometimes only occur in a few spatiotemporal cells. In NBN Atlas, each species record contains a set of meta-data including the observation's latitude, longitude and month, which is the information that we need in our experiments. For each of these 50 birds, we consider a binary classification problem, i.e. predicting whether or not the bird occurs in a particular cell (i.e. whether a grid cell contains at least one observation in the NBN Atlas data).

The results are reported in Table 2 and Table 3, for the SVM and MLP model respectively. The results are reported in terms of macro-average precision, recall, and F1 score over the 50 birds. However, note that all hyperparameters have been tuned with respect to the F1 metric. Training the MLP model is only feasible with relatively low-dimensional input representations. For this reason, we can only evaluate it on the Structured and SPATE representations, but not on the variants that include the bag-of-words representation. Indeed, the fact that neural network models are unsuitable for dealing with high-dimensional inputs is one of the main reasons why embedding models are used in practice.

Looking at the results from the SVM model, it can be clearly seen that combining Flickr tags with the available structured data leads to better results than using them separately. Moreover, combining them in our proposed spatiotemporal embeddings (SPATE) model leads to the best results. It significantly outperforms all the considered baselines, especially for the setting with the least amount of training data. Furthermore, note that the proposed KDE based spatiotemporal smoothing of Flickr tags leads to substantial improvements over the non-smoothed version in Flickr-noKDE and smoothing each tag with different bandwidths in Flickr consistently outperforms the method of smoothing all the tags with the same bandwidth in Flickr-1BW. We also found the normalization of the spatiotemporal KDE in Equation 3 to be critical to obtain good results. Based on the tuning data, for the SVM model, we found a linear kernel to be optimal when using Flickr data only and the combination of Structured + Flickr, and a Gaussian kernel to be optimal for the Structured, and SPATE models. For the embedding model, we found that the best results were obtained for 300 dimensions. We also found that the results of MLP in Table 3 are broadly in line with that of SVM.

As an example, Figure 6 visually compares the predictions that were made by the SVM classifier for the different representations. This example focuses on a particular bird, namely the Swift (Apus apus). The seasons in Figure 6 are defined as winter (December, January, February), spring (March, April, May), summer (June, July, August) and autumn (September, October, November). It can be clearly seen from Figure 6 that the predictions made by using Structured only, Flickr only, or Structured + Flickr are under-reported for winter and imbalanced (i.e. overestimated in some regions and underestimated in another) for the other seasons. However, SPATE leads to superior predictions over all the seasons. To get further insight into the performance of the considered representations, Figure 7 shows the monthly average F1 score for the predictions made for this particular species. Although using Flickr outperforms using Structured and Structured + Flickr improves the results, SPATE leads to the best results over all months. Interestingly, for the months with low numbers of occurrences, such as January and November, SPATE is the only model that made positive predictions while other models predicted all negatives. This example suggests that highly accurate distribution models can be learned using any of the considered models when we have sufficiently large numbers of occurrences as in the spring and summer months. However, our proposed SPATE model



(a) Structured data

(b) Flickr data

(c) Structured + Flickr data

(d) SPATE

(e) Ground truth data

Figure 6: Prediction of the seasonal distribution of Swift across the UK and Ireland using the SVM model with 1% of the data for training/tuning.

still performs better in the months with very low numbers of occurrences, as in the winter and autumn months. Additionally, when we look at the prediction confidence score of this species over the spatiotemporal grid cells, we found that our proposed SPATE model makes much higher confidence predictions than the other proposed baselines. As an example, Figure 8 shows the prediction confidence score obtained from different models for a particular location (lat-

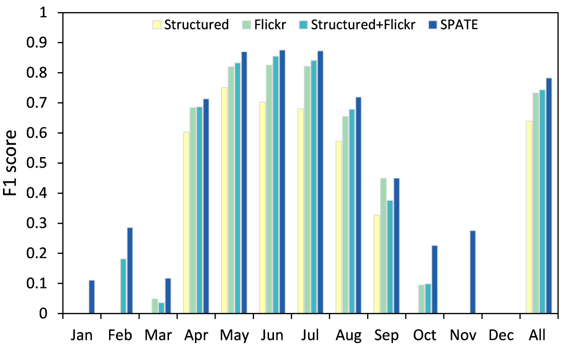


Figure 7: F1 score of predicting the monthly distribution of Swift using SVM

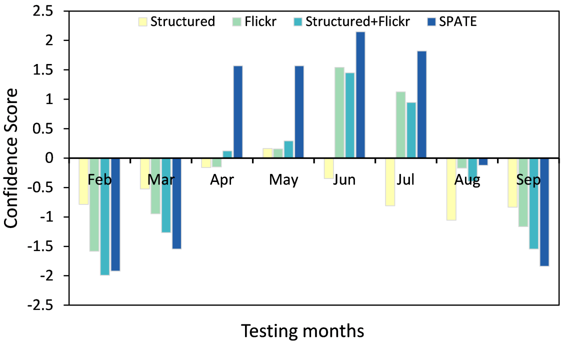


Figure 8: Prediction confidence score for location coordinate (54.815, -2.086) over the testing months using SVM. Jan, Oct, Nov and Dec are not shown in the figure because the corresponding cells are in the training set for that location. Note that Swift has positive ground truth observations in that location in April, May, June, July and August, and negative ground truth observations in February, March and September.

itude= 54.815 and longitude= -2.086) over all the testing months. Clearly, as can be seen in Figure 8 the predictions made by the SPATE model have very high confidence for the correct predictions and very low confidence for the incorrect predictions (see the incorrect prediction in August) which further illustrates the strong performance of our proposed model. Note that all the results reported in Figure 6, 7, 8 are for the setting where only 1% of the training/tuning data was used.

6.4.2. Predicting Climate Features

For this task, we consider five different regression problems: predicting the monthly average of precipitation, solar radiation, temperature, wind speed, and water vapour pressure. For these experiments, we do not include any of these climate features in the structured representations (and embeddings derived from them) as they serve here as ground truths. The results of these experiments are reported in Table 4 and Table 5 using SVR and MLP respectively. The results are presented in terms of mean absolute error (MAE) and Spearman ρ correlation between the predicted and actual values for all spatiotemporal cells in the testing set. Note that we tune all the parameters with respect to Spearman ρ. The mean and standard deviation of each of those features are shown in Table 6. Similar to the previous experiment, we were able to use MLP on the STRUCTURED and SPATE datasets only. Again, the results gained by both models are broadly in line for these datasets.

We can see from the results of SVR that combining structured and Flickr data outperforms using them separately. However, combining them using our proposed spatiotemporal embeddings (SPATE) leads to a substantial improvement over the baseline methods especially when we consider only 1% of the training/tuning data. Note that, for settings with more training data, it is unsurprising that all methods perform well as climate features are strongly autocorrelated in time and space.

In Figure 9, we visually illustrate the predictions made by using the different representations (for the case of the SVR model) for seasonal precipitation. The representations based on structured data perform worst while the SPATE representations are the best for all the seasons. While the overall differences between the results for precipitation (especially in term of Spearman ρ in Table 4) are small, clear differences between their performance are still noticeable in Figure 9. To get a clearer picture about the performance of each model, Figure 10 shows the monthly average MAE and Spearman ρ for predicting the precipitation. Although FLICKR performs better than STRUCTURED in terms of MAE, it performs worse in term of Spearman ρ. The combination of STRUCTURED + FLICKR performs in between them. Interestingly, our proposed SPATE model has the best performance in terms of MAE and Spearman ρ for all the months. Looking at the prediction of a particular location (latitude= 55.264 and longitude= -4.784) over all the testing months, we can see that the STRUCTURED model predictions do not deviate too far from the mean value, which have not affected the Spearman ρ score as much as MAE. The FLICKR model makes more varied predictions, although they are still far from the ground truth. The combination of STRUCTURED + FLICKR leads to more faithful predictions. However, the SPATE model performs significantly better. Again, all the results reported in Figure 9, 10, 11 are when using only 1% of the training/tuning data.

6.5. Location Similarity

In this section, we qualitatively evaluate the nature of the vectors generated by the SPATE model. Figure 12 and Figure 13 show the similarity maps of a number of selected locations in July and January respectively. The selected locations include cities of London, Dublin and Hull, the low populated but popular tourist areas of Snowdonia and Skye, which are mountainous, and the tourist area of Roseland Heritage Coast which is coastal and scenic, non-intensive agricultural land with small villages. The similarity has been measured according to the Euclidean distance between the vector representation of the cell which the considered location belongs

		1%		10 %		100 %	
		MAE	ρ	MAE	ρ	MAE	ρ
Precipitation	Structured	31.492	0.509	26.758	0.683	22.354	0.742
	Flickr-noKDE	32.214	0.125	31.724	0.202	30.808	0.268
	Flickr-1BW	28.750	0.538	23.492	0.697	22.865	0.725
	Flickr	28.240	0.549	23.601	0.698	22.562	0.741
	Structured + Flickr	27.385	0.562	22.999	0.711	20.780	**0.773**
	SPATE	24.509	**0.669**	22.971	**0.714**	21.402	0.767
Solar Radiation	Structured	4867.2	0.776	2476.0	0.895	1083.1	0.947
	Flickr-noKDE	5266.1	0.333	4603.9	0.386	4440.6	0.419
	Flickr-1BW	2434.5	0.829	1621.6	0.895	1534.4	0.914
	Flickr	2359.4	0.841	1575.9	0.901	1480.3	0.928
	Structured + Flickr	2045.2	0.884	1076.4	0.950	936.5	**0.973**
	SPATE	1415.3	**0.907**	1041.4	**0.955**	1030.6	0.960
Wind Speed	Structured	1.072	0.246	0.956	0.429	0.901	0.492
	Flickr-noKDE	1.081	0.082	1.070	0.130	1.063	0.170
	Flickr-1BW	1.099	0.217	0.963	0.418	0.897	0.493
	Flickr	1.084	0.251	0.959	0.421	0.874	0.512
	Structured + Flickr	1.001	0.347	0.938	0.456	0.873	0.522
	SPATE	0.953	**0.442**	0.930	**0.467**	0.848	**0.523**
Water Vap Press.	Structured	0.193	0.586	0.154	0.699	0.126	0.760
	Flickr-noKDE	0.234	0.110	0.226	0.250	0.225	0.279
	Flickr-1BW	0.187	0.607	0.155	0.698	0.136	0.748
	Flickr	0.186	0.612	0.152	0.707	0.134	0.752
	Structured + Flickr	0.176	0.661	0.143	0.738	0.126	0.777
	SPATE	0.135	**0.752**	0.122	**0.771**	0.119	**0.779**
Temperature	Structured	2.060	0.826	1.063	0.929	0.837	0.953
	Flickr-noKDE	3.415	0.228	3.142	0.350	2.979	0.397
	Flickr-1BW	1.653	0.849	1.372	0.888	1.074	0.919
	Flickr	1.636	0.845	1.306	0.891	1.034	0.931
	Structured + Flickr	1.302	0.907	1.054	0.932	0.823	**0.961**
	SPATE	1.164	**0.920**	1.010	**0.939**	0.935	0.946

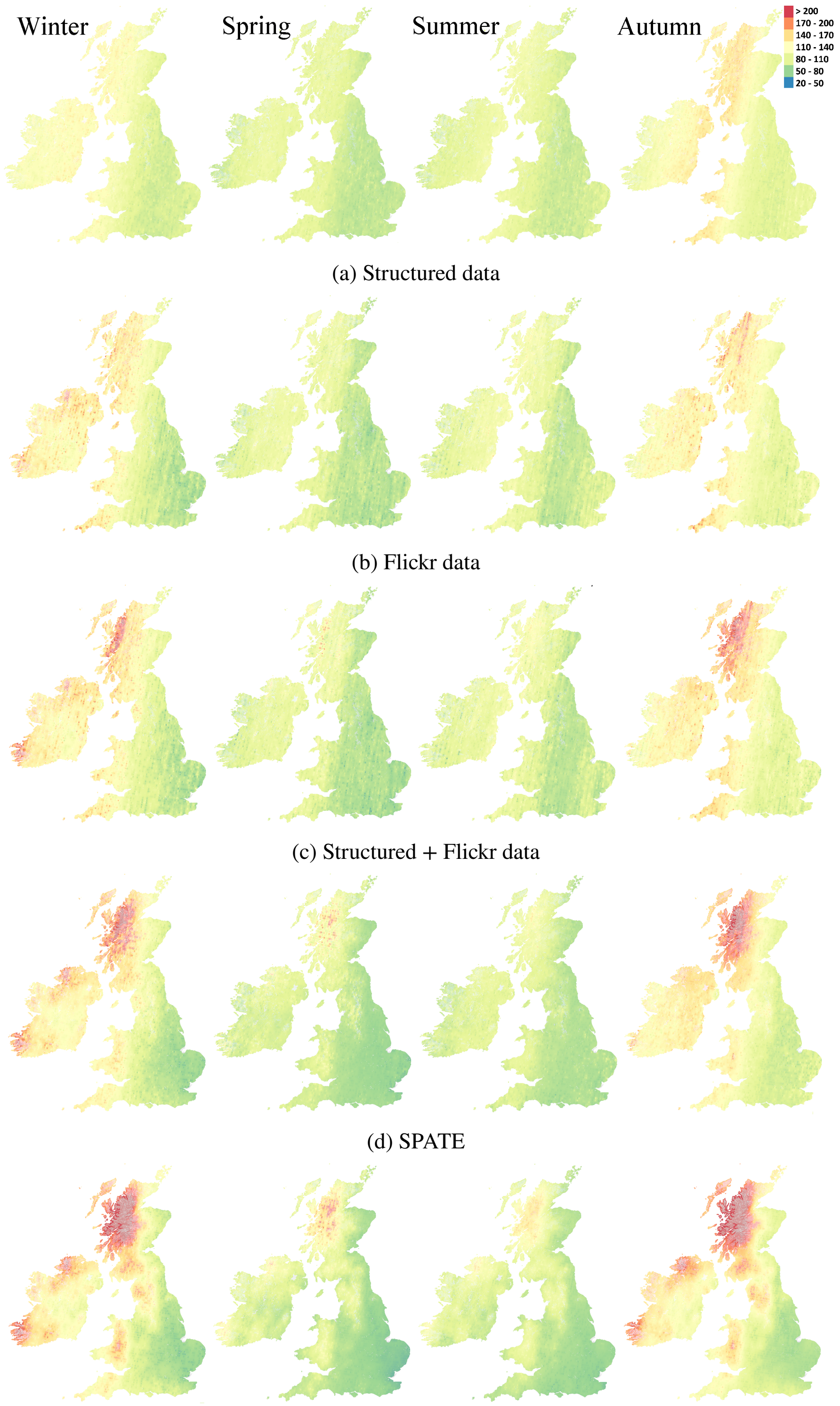
Table 4
Results for predicting the monthly average climate features using SVR.

		1%		10 %		100 %	
		MAE	ρ	MAE	ρ	MAE	ρ
Precip	Structured	23.713	0.682	19.904	0.776	17.137	0.809
	SPATE	20.882	**0.824**	16.043	**0.855**	12.367	**0.876**
Solar Rad.	Structured	3957.6	0.678	2303.7	0.762	1379.7	0.832
	SPATE	3866.1	**0.742**	1264.6	**0.817**	867.0	**0.865**
Wind Speed	Structured	0.829	0.557	0.737	0.652	0.689	0.687
	SPATE	0.889	**0.818**	0.706	**0.854**	0.545	**0.879**
Water Vap	Structured	0.143	0.735	0.115	0.798	0.093	0.830
	SPATE	0.128	**0.858**	0.108	**0.878**	0.086	**0.892**
Temp	Structured	1.667	0.832	1.353	0.858	1.063	0.875
	SPATE	1.040	**0.860**	0.940	**0.871**	0.796	**0.881**

Table 5
Results for predicting the monthly average climate features using MLP.

to and the other cells using 300 vector dimensions.

As a general observation, in all cases, the maps do succeed in highlighting regions that are very similar in several respects to the respective selected location. Thus London and Dublin are both similar to other major urban conurbations such as Birmingham, Manchester, Glasgow, Newcastle upon Tyne, Bristol, Cardiff and Belfast. They are least similar to low populated, mountainous rural areas such as the Highlands of Scotland and, in the case of London, the west of Ireland. Note that Dublin, the capital of Ireland is



(a) Structured data

(b) Flickr data

(c) Structured + Flickr data

(d) SPATE

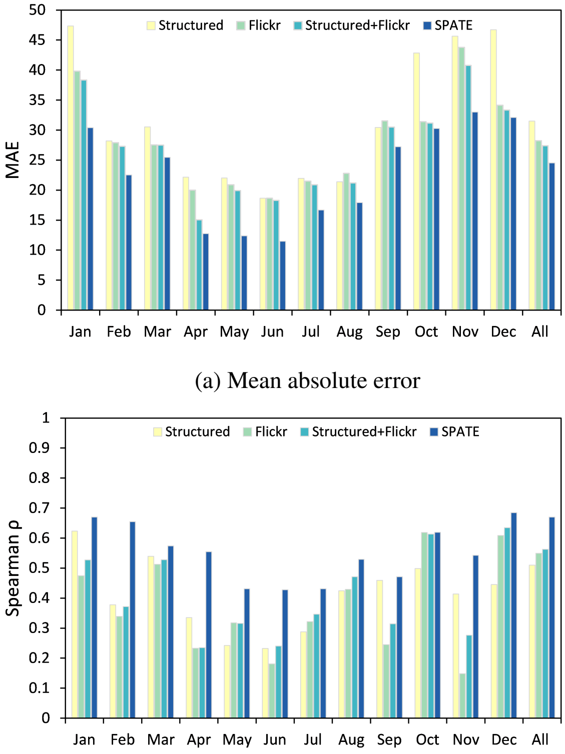
(e) Ground truth data

Figure 9: Prediction of the seasonal precipitation across the UK and Ireland using the SVR model with 1% of the data for training/tuning.

more similar to the rural west of Ireland, to which it is culturally related, than is London, just as London, the capital of England, is more similar, than is Dublin, to the geographically much closer rural areas of East Anglia in England. This latter distinction can be attributed to the general vocabulary of Flickr which is more similar, in references to places and activities, between regionally adjacent places. Hull is an industrial seaport and city. Similar locations in summer and

	Mean	STDEV
Precipitation (mm)	94.750	44.037
Solar Radiation (kJ $m^{-2}day^{-1}$)	9243.9	5847.7
Wind Speed (m s^{-1})	4.750	1.454
Water Vapor Press (kPa)	0.897	0.302
Temperature (°C)	9.021	3.970

Table 6
Mean and Standard deviation of the climate data.



(a) Mean absolute error

(b) Spearman ρ

Figure 10: The monthly prediction results of precipitation using SVR.

winter are other commercial and industrial coastal locations such as Liverpool, Newcastle upon Tyne, Bristol, Cardiff and Southampton, along with other relatively highly populated industrial inland regions such as Birmingham, Leeds and Manchester. It is most different from the west of Ireland and the highlands of Scotland which are mountainous regions with low population and pastoral agriculture.

Differences between summer and winter are much less marked than the differences between regions at the same times of the year, particularly for the cities. However, an example of a seasonal city difference can be observed for London, which is more different in the summer (July) from relatively remote rural areas such as parts of Wales and Cornwall. In the latter regions (Wales and Cornwall) there might be higher levels of observations in summer of the natural environment and of outdoor leisure activities when there are more tourists than in winter. The nature of different types of tourist activity might also explain the pronounced differences in summer between the mountainous but popular tourist area of Snowdonia and the also popular coastal tourism areas of south-west Ireland and south-east England. The Isle of Skye, while generally similar in summer and winter to other relatively low populated rural areas, has a greater difference from the south-east of England in winter than in summer. Speculatively, this might reflect the fact that, in winter, Skye with its low indigenous population and very much lower levels of tourism (in winter) will have relatively low levels of contribution to social media than the more populated areas of south-east England.

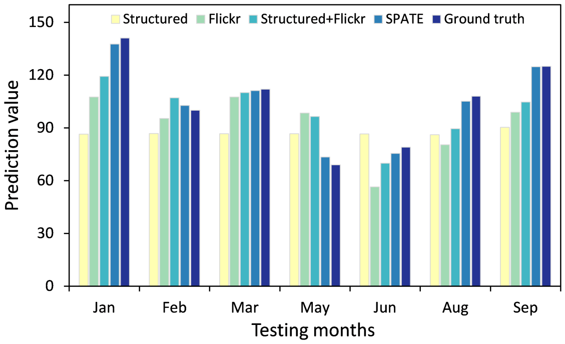
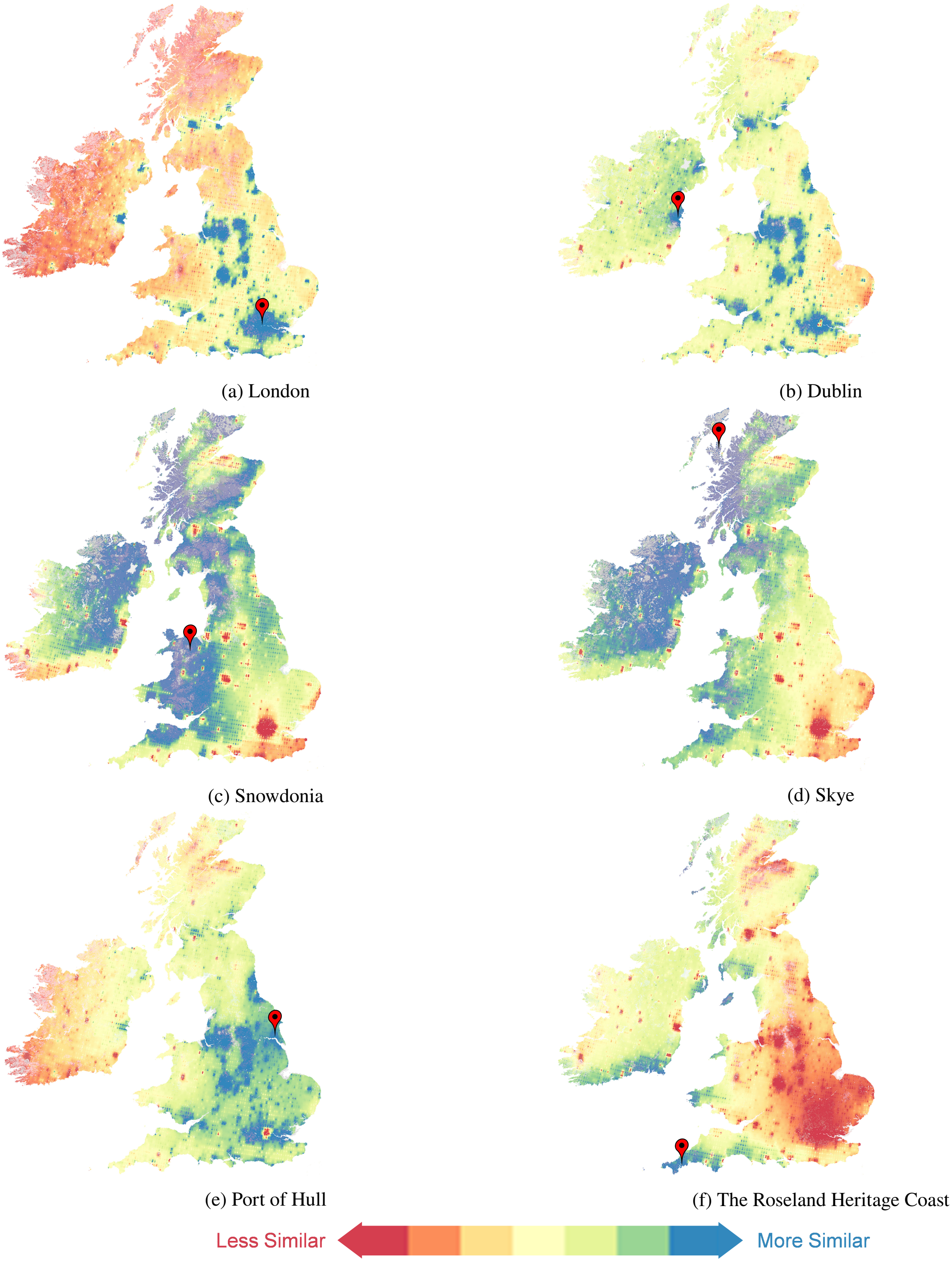


Figure 11: Monthly average value of predicting the amount of precipitation for location coordinate (55.264, -4.784) over the testing months using SVR. April, July, October, November, and December are not shown in the figure because the corresponding cells are in the training set for that location.

7. Conclusions and Future Work

In this paper, we have proposed a novel model for learning vector space embeddings of spatiotemporal entities which is able to integrate structured environmental information and textual information from Flickr tags. Furthermore, to handle the problem of Flickr data sparsity, we present a method based on kernel density estimation to smooth the distribution of Flickr tags over space and time. For evaluation, we have considered two experimental tasks. The first experiment aimed to predict the monthly distribution of species across the UK and Ireland, using observations from the National Biodiversity Network Atlas as ground truth. In the second experiment, we looked at predicting five climate related features.

The experimental results show that smoothing the distribution of Flickr tags leads to substantial improvements in comparison with the non-smoothed version. Moreover, combining Flickr tags with structured data consistently outperformed using them separately. This strongly suggests that Flickr can be a valuable supplement to more traditional datasets. Notably, our proposed spatiotemporal embeddings (SPATE) model provides an efficient integration of Flickr tags with structured information that outperforms all the considered baselines, especially when we considered very



(a) London

(b) Dublin

(c) Snowdonia

(d) Skye

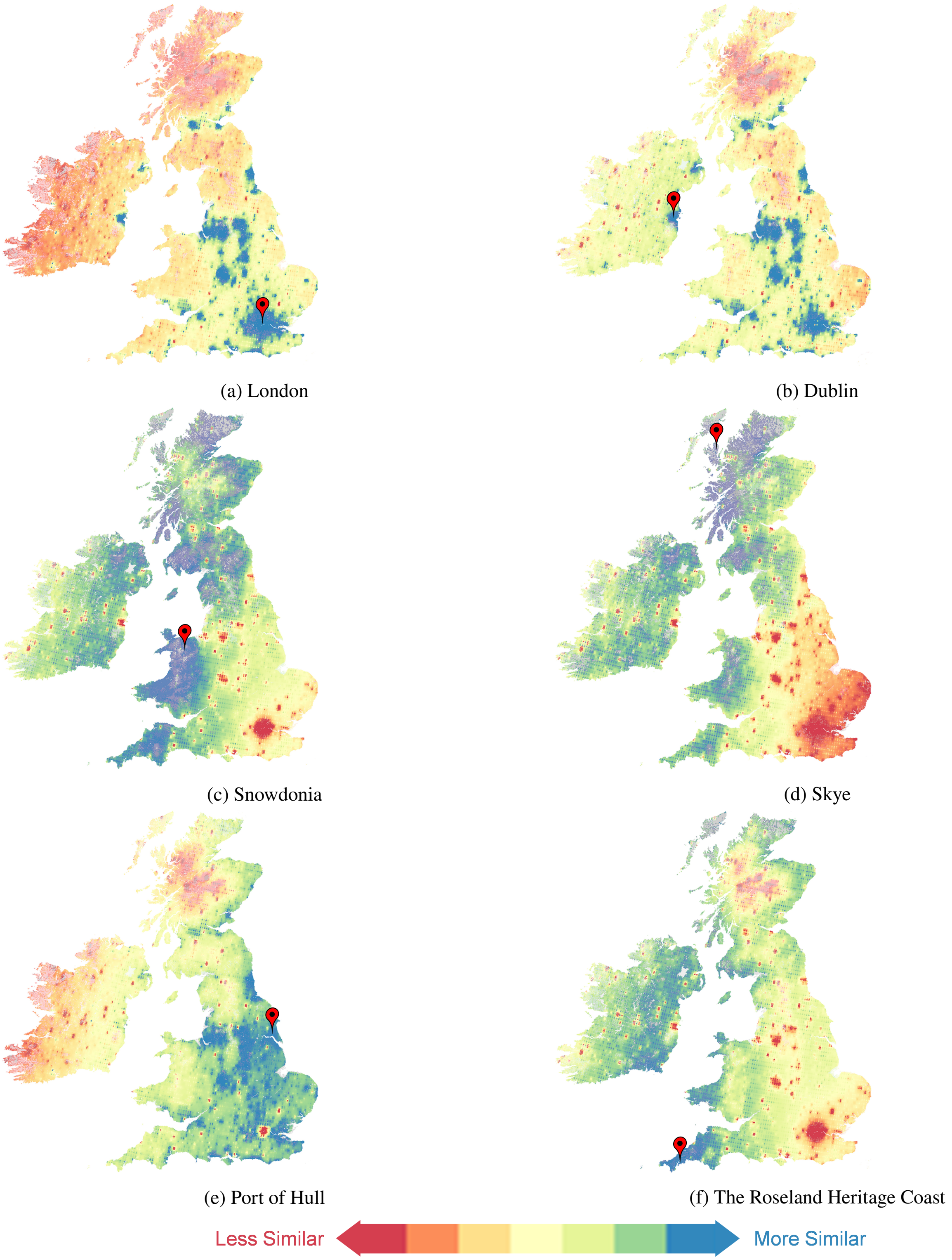
(e) Port of Hull

(f) The Roseland Heritage Coast

Figure 12: Location's similarity maps in July

small training datasets.

There are a number of directions for future work. First, we could learn a low dimensional vector space embedding for each species. This can be done by encoding the available ecological and habitat information about the considered species as well as all Flickr tags that occur in photographs tagged by the species name. We can also consider the textual and structured data about the considered species from other resources such as the Encyclopedia of Life. All these features would be integrated into a low di-



(a) London

(b) Dublin

(c) Snowdonia

(d) Skye

(e) Port of Hull

(f) The Roseland Heritage Coast

Figure 13: Location's similarity maps in January

mensional vector space embedding representing this species which could then be used to predict or confirm species observation. Second, extending the same analysis to data collected from other social media platforms such as Twitter, Instagram, and Wikipedia may alleviate the problem of data sparsity and improve the quality of the prediction. We could also consider additional scientific data sources, for example, remote sensing and earth observation data. Any new dataset can be added as an additional constraint in our embedding model.

Acknowledgments

Shelan Jeawak has been sponsored by HCED Iraq; Steven Schockaert has been supported by ERC Starting Grant 637277.

References

Abramson, I. S., 1982. On bandwidth variation in kernel estimates-a square root law. The Annals of Statistics, 1217–1223.

Adams, B., Janowicz, K., 2012. On the geo-indicativeness of non-georeferenced text. In: Sixth International AAAI Conference on Weblogs and Social Media.

Aghelpour, P., Mohammadi, B., Biazar, S. M., 2019. Long-term monthly average temperature forecasting in some climate types of iran, using the models sarima, svr, and svr-fa. Theoretical and Applied Climatology, 1–10.

Bamman, D., Dyer, C., Smith, N. A., 2014. Distributed representations of geographically situated language. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Vol. 2. pp. 828–834.

Barve, V. V., 2015. Discovering and developing primary biodiversity data from social networking sites. Ph.D. thesis, University of Kansas.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O., 2013. Translating embeddings for modeling multi-relational data. In: Advances in neural information processing systems. pp. 2787–2795.

Bowman, A. W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71 (2), 353–360.

Brunsdon, C., 1995. Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. Computers and Geosciences 21 (7), 877–894.

Brunsdon, C., Corcoran, J., Higgs, G., 2007. Visualising space and time in crime patterns: A comparison of methods. Computers, Environment and Urban Systems 31 (1), 52–75.

Cocos, A., Callison-Burch, C., 2017. The language of place: Semantic value from geospatial context. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Vol. 2. pp. 99–104.

Cunha, E., Martins, B., 2014. Using one-class classifiers and multiple kernel learning for defining imprecise geographic regions. International Journal of Geographical Information Science 28 (11), 2220–2241.

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., Yu, C., 2010. Constructing travel itineraries from tagged geo-temporal breadcrumbs. In: Proceedings of the 19th International Conference on World Wide Web. pp. 1083–1084.

Delmelle, E., Dony, C., Casas, I., Jia, M., Tang, W., 2014. Visualizing the impact of space-time uncertainties on dengue fever patterns. International Journal of Geographical Information Science 28 (5), 1107–1127.

Drake, J. M., Randin, C., Guisan, A., 2006. Modelling ecological niches with support vector machines. Journal of applied ecology 43 (3), 424–432.

Eisenstein, J., O'Connor, B., Smith, N. A., Xing, E. P., 2010. A latent variable model for geographic lexical variation. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. pp. 1277–1287.

ElQadi, M. M., Dorin, A., Dyer, A., Burd, M., Bukovac, Z., Shrestha, M., 2017. Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in Australia. Ecological Informatics 39, 23–31.

Estima, J., Fonte, C. C., Painho, M., 2014. Comparative study of land use/cover classification using Flickr photos, satellite imagery and corine land cover database. In: Proceedings of the 17th AGILE International Conference on Geographic Information Science, Castellon, Spain. pp. 1–6.

Estima, J., Painho, M., 2014. Photo based volunteered geographic information initiatives: A comparative study of their suitability for helping quality control of corine land cover. International Journal of Agricultural and Environmental Information Systems (IJAEIS) 5 (3), 73–89.

Feng, S., Cong, G., An, B., Chee, Y. M., 2017. Poi2vec: Geographical latent representation for predicting future visitors. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 102–108.

Fotheringham, A. S., Crespo, R., Yao, J., 2015a. Exploring, modelling and predicting spatiotemporal variations in house prices. The Annals of Regional Science 54 (2), 417–436.

Fotheringham, A. S., Crespo, R., Yao, J., 2015b. Geographical and temporal weighted regression (GTWR). Geographical Analysis 47 (4), 431–452.

Ge, L., Moh, T.-S., 2017. Improving text classification with word embedding. In: IEEE International Conference on Big Data. pp. 1796–1805.

Ghermandi, A., Sinclair, M., 2019. Passive crowdsourcing of social media in environmental research: A systematic map. Global Environmental Change 55, 36–47.

Grave, E., Mikolov, T., Joulin, A., Bojanowski, P., 2017. Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. pp. 427–431.

Grothe, C., Schaab, J., 2009. Automated footprint generation from geotags with kernel density estimation and support vector machines. Spatial Cognition & Computation 9 (3), 195–211.

Grover, A., Leskovec, J., 2016. node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 855–864.

Guo, Q., Kelly, M., Graham, C. H., 2005. Support vector machines for predicting distribution of sudden oak death in california. Ecological modelling 182 (1), 75–90.

Hasegawa, M., Kobayashi, T., Hayashi, Y., 2018. Social image tags as a source of word embeddings: A task-oriented evaluation. In: LREC. pp. 969–973.

Hovy, D., Purschke, C., 2018. Capturing regional variation with distributed place representations and geographic retrofitting. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 4383–4394.

Hu, Y., Wang, F., Guin, C., Zhu, H., 2018. A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. Applied Geography 99, 89–97.

Hulden, M., Silfverberg, M., Francom, J., 2015. Kernel density estimation for text-based geolocation. In: Twenty-Ninth AAAI Conference on Artificial Intelligence.

Jeawak, S., Jones, C., Schockaert, S., 2018. Mapping wildlife species distribution with social media: Augmenting text classification with species names. Liebniz International Proceedings in Informatics.

Jeawak, S. S., Jones, C. B., Schockaert, S., 2017. Using Flickr for characterizing the environment: an exploratory analysis. In: 13th International Conference on Spatial Information Theory. Vol. 86. pp. 21:1–21:13.

Jeawak, S. S., Jones, C. B., Schockaert, S., 2019. Embedding geographic locations for modelling the natural environment using Flickr tags and structured data. In: European Conference on Information Retrieval (ECIR 2019). pp. 51–66.

Joachims, T., 1998. Making large-scale svm learning practical. Tech. rep., SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund.

Kashani, M. H., Dinpashoh, Y., 2012. Evaluation of efficiency of different estimation methods for missing climatological data. Stochastic environmental research and risk assessment 26 (1), 59–71.

Kim, Y., Chiu, Y.-I., Hanaki, K., Hegde, D., Petrov, S., 2014. Temporal analysis of language through neural language models. arXiv preprint arXiv:1405.3515.

Kulkarni, V., Perozzi, B., Skiena, S., 2016. Freshman or fresher? quantifying the geographic variation of language in online social media. In: Tenth International AAAI Conference on Web and Social Media.

Le, Q., Mikolov, T., 2014. Distributed representations of sentences and documents. In: International conference on machine learning. pp. 1188–1196.

Leung, D., Newsam, S., 2012. Exploring geotagged images for land-use classification. In: Proceedings of the ACM multimedia 2012 workshop on Geotagging and its applications in multimedia. pp. 3–8.

Lilleberg, J., Zhu, Y., Zhang, Y., 2015. Support vector machines and word2vec for text classification with semantic features. In: Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference on. pp. 136–140.

Liu, Q., Jiang, H., Wei, S., Ling, Z.-H., Hu, Y., 2015. Learning semantic word embeddings based on ordinal knowledge constraints. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. pp. 1501–1511.

Liu, Q., Ling, Z.-H., Jiang, H., Hu, Y., 2016a. Part-of-speech relevance weights for learning word embeddings. arXiv preprint arXiv:1603.07695.

Liu, X., Liu, Y., Li, X., 2016b. Exploring the context of locations for personalized location recommendations. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. pp. 1188–1194.

McLean, M. I., 2018. Spatio-temporal models for the analysis and optimisation of groundwater quality monitoring networks. Ph.D. thesis, University of Glasgow.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119.

Muñoz-Mas, R., Martínez-Capel, F., Alcaraz-Hernández, J. D., Mouton, A. M., 2017. On species distribution modelling, spatial scales and environmental flow assessment with multi–layer perceptron ensembles: A case study on the redfin barbel (barbus haasi; mertens, 1925). Limnologica 62, 161–172.

Nickel, M., Kiela, D., 2017. Poincaré embeddings for learning hierarchical representations. In: Advances in Neural Information Processing Systems. pp. 6341–6350.

Ono, M., Miwa, M., Sasaki, Y., 2015. Word embedding-based antonym detection using thesauri and distributional information. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 984–989.

Pennington, J., Socher, R., Manning, C., 2014. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing. pp. 1532–1543.

Phillips, L., Shaffer, K., Arendt, D., Hodas, N., Volkova, S., 2017. Intrinsic and extrinsic evaluation of spatiotemporal text representations in Twitter streams. In: Proceedings of the 2nd Workshop on Representation Learning for NLP. pp. 201–210.

Qiu, L., Cao, Y., Nie, Z., Yu, Y., Rui, Y., 2014. Learning word representation considering proximity and ambiguity. In: AAAI. pp. 1572–1578.

Quercia, D., Schifanella, R., Aiello, L. M., 2014. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In: Proceedings of the 25th ACM conference on Hypertext and social media. pp. 116–125.

Radhika, Y., Shashi, M., 2009. Atmospheric temperature prediction using support vector machines. International journal of computer theory and engineering 1 (1), 55.

Richards, D. R., Friess, D. A., 2015. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: content analysis of social media photographs. Ecological Indicators 53, 187–195.

Saeidi, M., Riedel, S., Capra, L., 2015. Lower dimensional representations of city neighbourhoods. In: AAAI Workshop: AI for Cities.

Salcedo-Sanz, S., Deo, R., Carro-Calvo, L., Saavedra-Moreno, B., 2016. Monthly prediction of air temperature in australia and new zealand with machine learning algorithms. Theoretical and applied climatology 125 (1-2), 13–25.

Seaman, D. E., Powell, R. A., 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. Ecology 77 (7), 2075–2085.

Shaddick, G., Zidek, J. V., 2015. Spatio-temporal methods in environmental epidemiology. Chapman and Hall/CRC.

Silverman, B. W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall.

Speer, R., Chin, J., Havasi, C., 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 4444–4451.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B., 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vol. 1. pp. 1555–1565.

Taylor, C. C., 2008. Automatic bandwidth selection for circular density estimation. Computational Statistics & Data Analysis 52 (7), 3493–3500.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G., 2016. Complex embeddings for simple link prediction. In: International Conference on Machine Learning. pp. 2071–2080.

Van Canneyt, S., Schockaert, S., Dhoedt, B., 2013a. Discovering and characterizing places of interest using flickr and twitter. International Journal on Semantic Web and Information Systems (IJSWIS) 9 (3), 77–104.

Van Canneyt, S., Schockaert, S., Dhoedt, B., 2013b. Discovering and characterizing places of interest using Flickr and Twitter. International Journal on Semantic Web and Information Systems (IJSWIS) 9 (3), 77–104.

Van Laere, O., Quinn, J. A., Schockaert, S., Dhoedt, B., 2014. Spatially aware term selection for geotagging. IEEE transactions on Knowledge and Data Engineering 26, 221–234.

Vendrov, I., Kiros, R., Fidler, S., Urtasun, R., 2015. Order-embeddings of images and language. arXiv preprint arXiv:1511.06361.

Wang, J., Korayem, M., Crandall, D., 2013. Observing the natural world with Flickr. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 452–459.

Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W., Yang, S., 2017. Community preserving network embedding. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 203–209.

Weston, J., Bengio, S., Usunier, N., Oct. 2010. Large scale image annotation: Learning to rank with joint word-image embeddings. Machine. Learning 81 (1), 21–35.

Xu, C., Bai, Y., Bian, J., Gao, B., Wang, G., Liu, X., Liu, T., 2014. Rc-net: A general framework for incorporating knowledge into word representations. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 1219–1228.

Yan, B., Janowicz, K., Mai, G., Gao, S., 2017. From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 35:1–35:10.

Yang, B., Yih, W., He, X., Gao, J., Deng, L., 2015. Embedding entities and relations for learning and inference in knowledge bases. In: Proc. of ICLR-15.

Yang, J., Eickhoff, C., 2018. Unsupervised learning of parsimonious general-purpose embeddings for user and location modeling. ACM Transactions on Information Systems (TOIS) 36 (3), 32.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model. International Journal of Geographical Information Science 31 (4), 825–848.

Zhang, C., Zhang, K., Yuan, Q., Peng, H., Zheng, Y., Hanratty, T., Wang, S., Han, J., 2017a. Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning. In: Proceedings of the 26th International Conference on World Wide Web. pp. 361–370.

Zhang, C., Zhang, K., Yuan, Q., Tao, F., Zhang, L., Hanratty, T., Han, J., 2017b. React: Online multimodal embedding for recency-aware spatiotemporal activity modeling. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 245–254.

Zhang, H., Korayem, M., Crandall, D. J., LeBuhn, G., 2012. Mining photo-sharing websites to study ecological phenomena. In: Proceedings of the 21st international conference on World Wide Web. pp. 749–758.

Zhao, S., Zhao, T., King, I., Lyu, M. R., 2017. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In: Proceedings of the 26th International Conference on World Wide Web Companion. pp. 153–162.