GEOGRAPHICAL ORIGIN PREDICTION OF FOLK MUSIC RECORDINGS FROM THE UNITED KINGDOM

Vytaute Kedyte[1]   **Maria Panteli**[2]   **Tillman Weyde**[1]   **Simon Dixon**[2]

[1] Department of Computer Science, City University of London, United Kingdom

[2] Centre for Digital Music, Queen Mary University of London, United Kingdom

{Vytaute.Kedyte, T.E.Weyde}@city.ac.uk, {m.panteli, s.e.dixon}@qmul.ac.uk

ABSTRACT

Field recordings from ethnomusicological research since the beginning of the 20th century are available today in large digitised music archives. The application of music information retrieval and data mining technologies can aid large-scale data processing leading to a better understanding of the history of cultural exchange. In this paper we focus on folk and traditional music from the United Kingdom and study the correlation between spatial origins and musical characteristics. In particular, we investigate whether the geographical location of music recordings can be predicted solely from the content of the audio signal. We build a neural network that takes as input a feature vector capturing musical aspects of the audio signal and predicts the latitude and longitude of the origins of the music recording. We explore the performance of the model for different sets of features and compare the prediction accuracy between geographical regions of the UK. Our model predicts the geographical coordinates of music recordings with an average error of less than 120 km. The model can be used in a similar manner to identify the origins of recordings in large unlabelled music collections and reveal patterns of similarity in music from around the world.

1. INTRODUCTION

Since the beginning of the 20th century ethnomusicological research has contributed significantly to the collection of recorded music from around the world. Collections of field recordings are preserved today in digital archives such as the British Library Sound Archive. The advances of Music Information Retrieval (MIR) technologies make it possible to process large numbers of music recordings. We are interested in applying these computational tools to study a large collection of folk and traditional music from the United Kingdom (UK). We focus on exploring music attributes with respect to geographical regions of the UK and investigate patterns of music similarity.

The comparison of music from different geographical regions has been the topic of several studies from the field of ethnomusicology and in particular the branch of comparative musicology [13]. Savage et al. [17] studied stylistic similarity within music cultures of Taiwan. In particular, they formed music clusters for a collection of 259 traditional songs from twelve indigenous populations of Taiwan and studied the distribution of these clusters across geographical regions of Taiwan. They showed that songs of Taiwan can be grouped into 5 clusters correlated with geographical factors and repertoire diversity. Savage et al. [18] analysed 304 recordings contained in the 'Garland Encyclopedia of World Music' [14] and investigated the distribution of music attributes across music recordings from around the world. They proposed 18 music features that are shared amongst many music cultures of the world and a network of 10 features that often occur together.

The aforementioned studies incorporated knowledge from human experts in order to annotate music characteristics for each recording. While expert knowledge provides reliable and in-depth insights into the music, the amount of human labour involved in the process makes it impractical for large-scale music corpora. Computational tools on the other hand provide an efficient solution to processing large numbers of music recordings. In the field of MIR several studies have used computational tools to study large music corpora. For example, Mauch et al. [10] studied the evolution of popular music in the USA in a collection of approximately 17000 recordings. They concluded that popular music in the US evolved with particular rapidity during three stylistic revolutions, around 1964, 1983 and 1991. With respect to non-Western music repertoires Moelants et al. [12] studied pitch distributions in 901 recordings from Central Africa from the beginning until the end of the 20th century. They observed that recent recordings tend to use more equally-tempered scales than older recordings.

Computational studies have also focused on predicting the geographic location of recordings from their music content. Gomez et al. [3] approached prediction of musical cultures as a classification problem, and classified music tracks into Western and non-Western. They identified correlations between the latitude and tonal features, and the longitude and rhythmic descriptors. Their work illustrates the complexity of using regression to predict the geographical coordinates of music origin. Zhou et al. [23] also approached this as a regression problem, predicting latitudes

and longitudes of the capital city of the music's country of origin, for pieces of music from 73 countries. They used K-nearest neighbours and Random Forest regression techniques, and achieved a mean distance error between predicted and target coordinates of 3113 kilometres (km). The advantage of treating geographic origin prediction as a regression problem is that it allows the latitude and longitude correlations found by Gomez et al. [3] to be considered as well as the topology of the Earth. The disadvantage is not accounting for latitudes getting distorted towards the poles, and longitudes diverging at ± 180 degrees. Location is usually used as an input feature in regression models, however some studies have explored prediction of geographical origin in a continuous space in the domains of linguistics [2], criminology [22], and genetics [15, 21].

In this paper we study the correlation between spatial origins and musical characteristics of field recordings from the UK. We investigate whether the geographical location of a music recording can be predicted solely based on its audio content. We extract features capturing musical aspects of the audio signal and train a neural network to predict the latitude and longitude of the origins of the recording. We investigate the model's performance for different network architectures and learning parameters. We also compare the performance accuracy for several feature sets as well as the accuracy across different geographical regions of the UK.

Our developments contribute to the evaluation of existing audio features and their applicability to folk music analysis. Our results provide insights for music patterns across the UK, but the model can be expanded to process music recordings from all around the world. This could contribute to identifying the location of recordings in large unlabelled music collections as well as studying patterns of music similarity in world music.

This paper is organised as follows: Section 2 provides an overview of the music collection and Section 3 describes the different sets of audio features considered in this study. Section 4 provides a detailed description of the neural network architecture as well as the training and testing procedures. Section 5 presents the results of the model for different learning parameters, audio features, and geographical areas. We conclude with a discussion and directions for future work.

2. DATASET

Our music dataset is drawn from the World & Traditional music collection of the British Library Sound Archive[1] which includes thousands of music recordings collected over decades of ethnomusicological research. In particular, we use a subset of the World & Traditional music collection curated for the Digital Music Lab project [1]. This subset consists of more than 29000 audio recordings with a large representation (17000) from the UK. We focus solely on recordings from the UK and process information on the recording's location (if available) to extract the latitude and

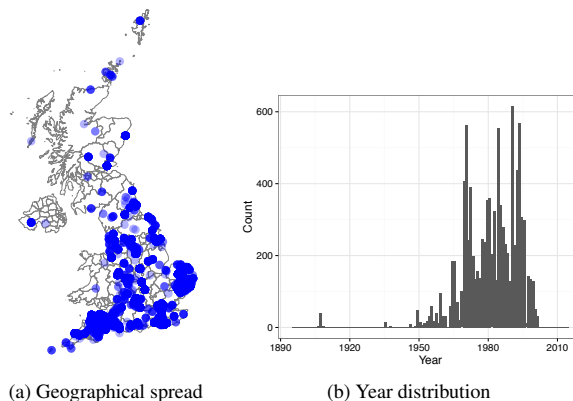(a) Geographical spread   (b) Year distribution

Figure 1: Geographical spread and year distribution in our dataset of 10055 traditional music recordings from the UK.

longitude coordinates. We keep only those tracks whose extracted coordinates lie within the spatial boundaries of the UK.

The final dataset consists of a total of 10055 recordings. The recordings span the years between 1904 and 2002 with median year 1983 and standard deviation 12.3 years. See Figure 1 for an overview of the geographical and temporal distribution of the dataset. The origins of the recordings span a range of maximum 1222 km. From the origins of all 10055 recordings we compute the average latitude and average longitude coordinates and estimate the distance between each recording's location and the average latitude, longitude. This results in a mean distance of 167 with standard deviation of 85 km. A similar estimate is computed from recordings in the training set and used as the random baseline for our regression predictions (Section 5).

3. AUDIO FEATURES

We aim to process music recordings to extract audio features that capture relevant music characteristics. We use a speech/music segmentation algorithm as a preprocessing step and extract features from the music segments using available VAMP plugins[2]. We post-process the output of the VAMP plugins to compute musical descriptors based on state of the art MIR research. Additional dimensionality reduction and scaling is considered as a final step. The methodology is summarised in Figure 2 and details are explained below.

Several recordings in our dataset consist of compilations of multiple songs or a mixture of speech and music segments. The first step in our methodology is to use a speech/music segmentation algorithm to extract relevant music segments from which the rest of the analysis is derived. We choose the best performing segmentation algorithm [9] based on the results of the Music/Speech Detection task of the MIREX 2015 evaluation[3]. We apply the segmentation algorithm to extract music segments from

---

[1] http://sounds.bl.uk/World-and-traditional-music

[2] http://www.vamp-plugins.org

[3] http://www.music-ir.org/mirex/wiki/2015:Music/Speech_Classification_and_Detection
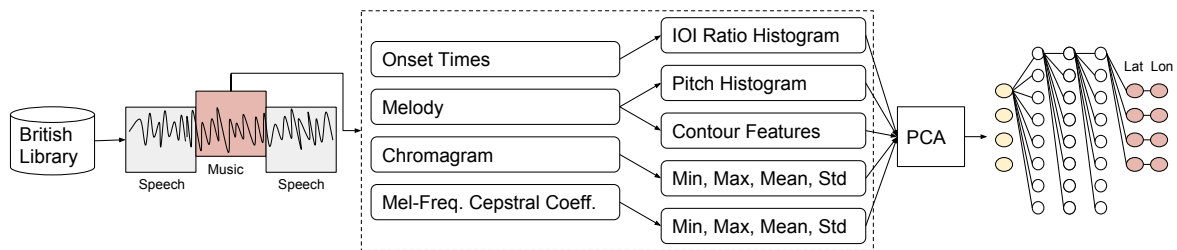
Figure 2: Summary of the methodology: UK folk music recordings are processed with a speech/music segmentation algorithm and VAMP plugins are applied to music segments. Audio features are derived from the output of the VAMP plugins, PCA is applied, and output is fed to a neural network that predicts the latitude and longitude of the recording.

each recording in our dataset. We require a minimum of 10 seconds of music for each recording and discard any recordings with total duration of music segments less than this threshold.

Our analysis aims to capture relevant musical characteristics which are informative for the spatial origins of the music. We focus on aspects of rhythm, melody, timbre, and harmony. We derive audio features from the following VAMP plugins: MELODIA - Melody Extraction [4], Queen Mary - Chromagram [5], Queen Mary - Mel-Frequency Cepstral Coefficients [6], and Queen Mary - Note Onset Detector [7]. We apply these plugins for each recording in our dataset and omit frames that correspond to non-music segments as annotated by the previous step of speech/music segmentation.

The raw output of the VAMP plugins cannot be directly incorporated in our regression model. We post-process the output to low-dimensional and musically meaningful descriptors as explained below.

Rhythm. We post-process the output of the Queen Mary - Note Onset Detector plugin to derive histograms of inter-onset interval (IOI) ratios [4]. Let $O = \{o_1, ..., o_n\}$ denote a sequence of n onset locations (in seconds) as output by the VAMP plugin. The IOIs are defined as $IOI = \{o_{i+1} - o_i\}$ for index $i = 1, ..., n-1$. The IOI ratios are defined as $IOIR = \{\frac{IOI_{j+1}}{IOI_j}\}$ for index $j = 1, ..., n-2$. The IOI ratios denote tempo-independent descriptors because the tempo information carried with the magnitude of IOIs vanishes with the ratio estimation. We compute a histogram for the $IOIR$ values with 100 bins uniformly distributed between $[0, 10)$.

Timbre. We extract summary statistics from the output of the Queen Mary - Mel-Frequency Cepstral Coefficients (MFCC) plugin [8] with the default values of frame and hop size. In particular, we remove the first coefficient (DC component) and extract the min, max, mean, and standard deviation of the remaining 19 MFCCs over time.

Melody. The output of the MELODIA - Melody Extraction plugin denotes the frequency estimates over time

of the lead melody. We extract a set of features capturing characteristics of the pitch contour shape and melodic embellishments [16]. In particular, we extract statistics of the pitch range and duration, fit a polynomial curve to model the overall shape and turning points of the contour, and estimate the vibrato range and extent of melodic embellishments. Each recording may consist of multiple shorter pitch contours. We keep the mean and standard deviation of features across all pitch contours extracted from the audio recording. We also post-process the output from MELODIA to compute an octave-wrapped pitch histogram [20] with 1200-cent resolution.

Harmony. The output of the Queen Mary - Chromagram plugin is an octave-wrapped chromagram with 100-cent resolution [5]. We use the default frame and hop size and extract summary statistics denoting the min, max, mean, and standard deviation of chroma vectors over time.

The above process results in a total of 1484 features per recording. Before further processing, the features were standardised with z-scores. Dimensionality reduction was also applied with Principal Component Analysis (PCA) including whitening and keeping enough components to represent 99% of the variance.

4. REGRESSION MODEL

The prediction of spatial coordinates from music data has been treated as a regression problem in previous research using K-nearest neighbours and Random Forest Regression methods [23]. We explore the application of a neural network method. Neural networks have been shown to outperform existing methods in supervised tasks of music similarity [7, 11, 19]. We evaluate the performance of a neural network under different parameters for the regression problem of predicting latitude and longitudes from music features.

A neural network with two continuous value outputs, latitude and longitude predictions, was built in Tensorflow. We used the Adaptive Moment Estimation (Adam) algorithm for optimisation, Rectified Linear Unit (ReLU) as activation function, and drop-out rate of 0.5 for regularisation. The evaluation of the model performance was based on the mean distance error in km, calculated using the Haversine formula [6]. The Haversine distance d between two points in km is given by

---

[4] http://mtg.upf.edu/technologies/melodia
[5] http://vamp-plugins.org/plugin-doc/qm-vamp-plugins.html#qm-chromagram
[6] http://vamp-plugins.org/plugin-doc/qm-vamp-plugins.html#qm-mfcc
[7] http://vamp-plugins.org/plugin-doc/qm-vamp-plugins.html#qm-onsetdetector

Parameters	Values
Target Scaling	True or False
Number of hidden layers	$\{3, 4\}$
Cost function	Haversine or MSE
Learning Rate	$\{0.005, 0.01, 0.05\}$
L1 regularisation	$\{0, 0.05, 0.5\}$
L2 regularisation	$\{0, 0.05, 0.5\}$

Table 1: The hyper-parameters and their range of values for optimisation.

$$d = 2r \arcsin([\sin^2(\frac{\phi_2 - \phi_1}{2}) +$$
$$\cos(\phi_1)\cos(\phi_2)\sin^2(\frac{\lambda_2 - \lambda_1}{2})]^{\frac{1}{2}}) \quad (1)$$

where ϕ represents the latitude, λ longitude, and r the radius of the sphere (with r fixed to 6367 km in this study). We further explored the performance of the model under architectures with different numbers of hidden layers, two different cost functions, and a range of regularisation parameters as explained below.

4.1 Parameter Optimisation

A grid-search of model hyper-parameters was performed to identify the combination that achieves best performance in cross-validation. The following hyper-parameters were considered for optimisation: whether or not to scale the targets (i.e., z-score standardisation of the ground truth latitude/longitude coordinates of each recording), the number of hidden layers, two possible cost functions, namely, the Haversine distance in km and the Mean Squared Error (MSE), and a range of values for learning rate, $L1$ and $L2$ regularisation parameters. The parameter optimisation is summarised in Table 1. We tested in total 216 combinations of hyper-parameters and selected the best performing combination to tune parameters and retrain the model for the final results.

4.2 Train-test splits

The training of the model was done in two phases. First the model was trained using the full set of features (Section 3) and the different hyper-parameters as defined in Table 1. The hyper-parameters were tuned based on the optimal performance obtained through cross-validation. In the second phase, the hyper-parameters were fixed to their optimal values and the model was retrained for different sets of features. Each new model's performance was assessed on a test set unique to that model.

In the first training phase, we sampled at random 70% from the total number of 10055 recordings for training. This resulted in a total of 7038 samples in the training set, of which 30% (2111) was set aside for validation. Following PCA, the feature dimensionality of the dataset was 368.

Target Scaling	Hidden Layers	Cost Function	Training Error (km)	Validation Error (km)
True	3	Haversine	**72.68**	**119.36**
True	3	MSE	166.21	166.27
True	4	Haversine	98.03	128.44
True	4	MSE	166.19	166.24
False	3	Haversine	165.34	166.79
False	3	MSE	169.91	169.30
False	4	Haversine	170.91	171.26
False	4	MSE	181.44	180.10

Table 2: Results for parameter optimisation. Learning rate, $L1$, and $L2$ regularisation parameters are fixed to $0.005, 0, 0.5$ respectively. Best performance is obtained when target scaling is combined with 3 hidden layers and Haversine distance as cost function.

We used cross-validation with $K = 5$ folds and tuned parameters based on the mean of the distance error on the validation set (Equation 1). In the second phase we retrained the model for different feature sets. For each feature set, the dataset was split into training (random 70%) and test (remaining 30%) and the performance of the model was assessed on the test set.

5. RESULTS

5.1 Parameter Optimisation

The model that produced the lowest mean error on the validation set (119 km) used the following hyper parameters: target scaling, 3 hidden layers, Haversine distance as cost function, learning rate of 0.005, and $L1$, $L2$ regularisation parameters of 0 and 0.5, respectively. The main hyper-parameters that determined the accuracy of the model were the use of Haversine distance as the cost function, and the application of target scaling. The performance of the model for different parameter values is shown in Table 2.

5.2 Results for different feature sets

The second set of experiments explored the performance of the model when trained for different sets of features. We estimated the random baseline from the origins of recordings in the training set. In particular, we computed the average latitude and average longitude coordinates of recordings and estimated the distance between each recording's location and the average latitude, longitude. Based on this estimate the mean distance error of the baseline approach was 167.4 km. Each model was compared to the baseline approach (i.e., the mean distance error of its test targets) with a Wilcoxon signed-rank test. The performances of the models trained on different sets of features and evaluated on separate test sets were compared with a pairwise Wilcoxon rank sum test (also known as Mann-Whitney) with Bonferroni correction for multiple comparisons. We consider a significance level of $\alpha = 0.05$ and denote the Bonferroni corrected level by $\hat{\alpha}$.

Model No.	Feature Set Name	Error (km)
1	All features	149.8
2	Rhythm: IOIR histogram	160.0
3	Harmony: Chromagram statistics	152.5
4	Timbre: MFCC statistics	129.0
5	Pitch histogram	160.1
6	Contour features mean	159.8
7	Contour features standard deviation	162.3
8	Melody: Pitch hist., contour features	152.6
9	Rhythm and Harmony	149.1
10	Rhythm and Timbre	120.1
11	Rhythm and Melody	150.5
12	Melody and Harmony	139.4
13	Melody and Timbre	117.1
14	Timbre and Harmony	**114.0**
15	Rhythm, Harmony, and Timbre	118.3
16	Rhythm, Harmony, and Melody	142.8
17	Rhythm, Timbre, and Melody	119.8
18	Harmony, Timbre, and Melody	140.3
–	Baseline	167.4

Table 3: The mean distance error (in km) of the test set for 18 models trained on different sets of features.
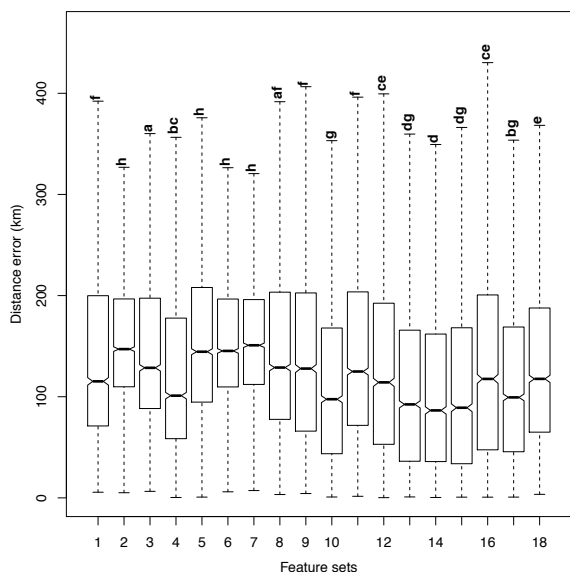
Figure 3: Distance error of predictions for different sets of features (see Table 3 for the feature set used to train each model). Labels $a-l$ in indicate features sets that have non-significantly different results ($p > \hat{\alpha}$) where they share the same letter. For example, feature set 3 shares the label a with feature set 8 but shares no label with any other feature set, indicating that results from model 3 are significantly different from all other models except for model 8.
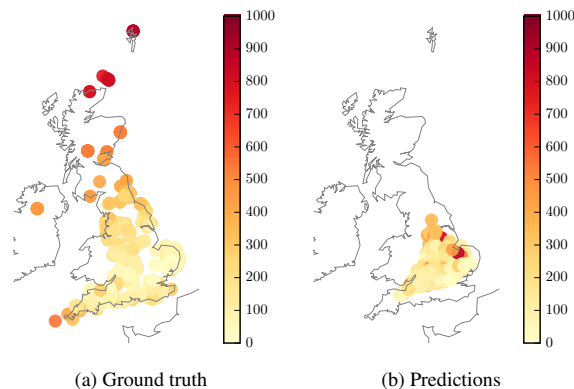
(a) Ground truth   (b) Predictions

Figure 4: (a) Ground truth and (b) predicted music recording origins, coloured by the distance error (in km) for the best performing model (no. 14).

All models achieved results significantly different from the baseline approach ($p < .0001$). The best performance (lowest error of 114.0 km) was achieved when combining the timbral and harmonic descriptors (model 14). This combines the summary statistics of the chromagram and the summary statistics of the MFCCs. The performance of this model was significantly different ($p < \hat{\alpha}$) from all other models except models 13 and 15 trained on melodic and timbral, and rhythmic, harmonic and timbral descriptors, respectively. The model achieved a mean error of 149.8 km on the test set when all features (Section 3) were used. The results from model 3 trained on harmonic descriptors were significantly different from all other models except model 8 trained on melodic features. The model trained on rhythmic descriptors (model 2) is amongst the weakest predictors. However, adding rhythmic features to any of melodic, harmonic, or timbral features, for example models $9, 10, 11$, significantly improves the performance of the model ($p < \hat{\alpha}$ for pairwise comparisons between models 3 and 9, 4 and 10, 8 and 11). Models $5, 6, 7$ trained on pitch histograms, contour features mean, and contour features standard deviation, respectively, are also amongst the weakest predictors but when all these features are combined together as in model 8, the performance is improved. See Table 3 for an overview of the prediction accuracy of models trained on different feature sets. Figure 3 provides a box-plot visualisation of the results from different feature sets and marks statistical significance between results.

5.3 Results for different regions

The last analyses aim to study the prediction accuracy with respect to the geographical origins of recordings. Figure 4 shows the ground truth and predicted coordinates for the best performing model (model no.14 as denoted in Table 3) coloured by the distance error in km. We observe that data points with the lowest predictive accuracy originate from the north-eastern and the south-western areas of the UK (Figure 4a). Predictions are mostly concentrated in the southern part of the UK. Data points predicted towards the
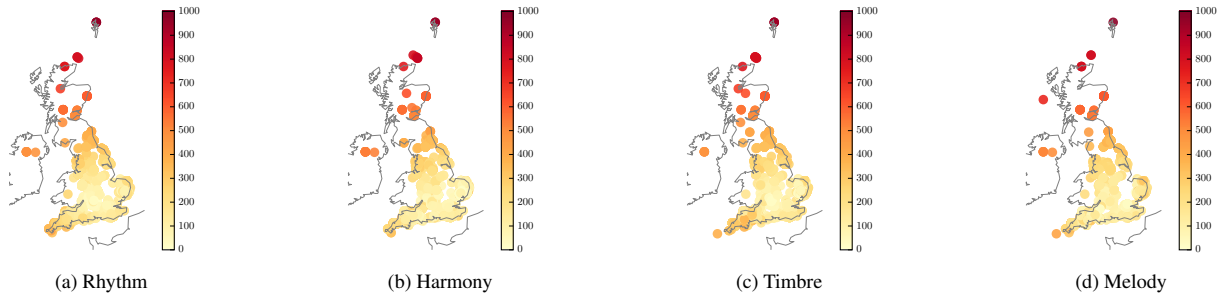
(a) Rhythm   (b) Harmony   (c) Timbre   (d) Melody

Figure 5: Music recording origins coloured by the distance error (in km) for models trained on (a) rhythmic, (b) harmonic, (c) timbral, and (d) melodic features (models no. $2, 3, 4, 8$ respectively as defined in Table 3).

eastern areas indicate a larger distance error (Figure 4b).

In Figure 5 we visualise the prediction accuracy of models trained on different feature sets with respect to geography. We observe that for all models the northern areas of the UK (i.e., in the region of Scotland) are predicted with a relatively large distance error (lowest accuracy). For the model trained on timbral features (Figure 5c) we also observe the south west of England predicted with lower accuracy than the models trained on harmonic and melodic features (Figures 5b and 5d).

6. DISCUSSION

Our results provide insights on the contribution of different feature sets and suggest patterns of music similarity across geographical regions. The methodology can be improved in various ways.

The initial corpus of folk and traditional music from the UK consisted of a total of 17000 of which only 10055 were processed in this study. The final dataset had a skewed geographical distribution with over-representation of the south-eastern and south-western UK regions, e.g., Devon and Suffolk, and under-representation of the North-Eastern, North-Western areas, e.g., Scotland and Northern Ireland. Effects from the skewness of the dataset could be observed in the distribution of predicted latitude and longitude coordinates (Figure 4b). A larger and more representative corpus can be used in future work.

We used features derived from the output of VAMP plugins to describe musical content of audio recordings. Some of these plugins were designed for different music styles and their application to folk music might not give robust results. A thorough evaluation of the suitability of the features can give valuable insights for improving their robustness to different corpora such as the one used in this study. We used feature representations averaged over time but in future work preserving temporal information in the features could provide better music content description.

We observed that results from models trained on individual features showed on average larger distance errors. When however combinations of features were considered, the model achieved on average higher accuracies. An exception is the case when all features were considered but the performance of the model had a relatively large distance error. This could be due to limitations of the model especially with regards to over-fitting or the lack of adequate music information captured by the features. Integrating additional audio features could help capture more of the variance of the data and improve the model.

The model was validated for a range of parameters and several approaches were considered to avoid over-fitting. However, evidence of over-fitting could still be observed in the final results. Training with more data could help make the model more generalisable in future work. What is more, oversampling techniques could be explored to overcome the problem of under-represented geographical regions in our dataset.

Neural networks in combination with audio features as proposed in this study, can provide good predictions of the origins of the music. This can aid musicological research as well as improve spatial metadata associated with large music collections.

7. CONCLUSION

We studied a collection of field recordings from the UK and investigated whether the geographical origins of recordings can be predicted from the music attributes of the audio signal. We treated this as a regression problem and trained a neural network to take as input audio features and predict the latitude and longitude of the music's origin. We trained the model under different hyperparameters and tested its performance for different feature sets. Highest accuracy was achieved for the model trained on timbral and harmonic features but no significant differences were found to the same model with rhythm features added or with melody replacing harmony. The southern regions of the UK were predicted with a relatively high accuracy whereas northern regions were predicted with low accuracy. Effects of the skewness of the dataset and the reliability of audio features were discussed. The corpus and methodology can be improved in future work and the applicability of the model could be extended to music from around the world.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] S. Abdallah, E. Benetos, N. Gold, S. Hargreaves, T. Weyde, and D. Wolff. The Digital Music Lab: A Big Data Infrastructure for Digital Musicology. *ACM Journal on Computing and Cultural Heritage*, 10(1), 2017.

[2] J. Eisenstein, B. O'Connor, N.A Smith, and E.P. Xing. A Latent Variable Model for Geographic Lexical Variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1277–1287, 2010.

[3] E. Gómez, M. Haro, and P. Herrera. Music and geography: Content description of musical audio from different parts of the world. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 753–758, 2009.

[4] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer. Evaluating rhythmic descriptors for musical genre classification. In *Proceedings of the AES 25th International Conference*, pages 196–204, 2004.

[5] C. Harte and M. Sandler. Automatic chord identifcation using a quantised chromagram. In *118th Audio Engineering Society Convention*, 2005.

[6] J. Inman. *Navigation and Nautical Astronomy: For the Use of British Seamen*. F. & J. Rivington, 1849.

[7] I. Karydis, K. Kermanidis, S. Sioutas, and L. Iliadis. Comparing content and context based similarity for musical data. *Neurocomputing*, 107:69–76, 2013.

[8] B. Logan. Mel-Frequency Cepstral Coefficients for Music Modeling. In *Proceedings of the International Symposium on Music Information Retrieval*, 2000.

[9] M. Marolt. Music/speech classification and detection submission for MIREX 2015. In *MIREX*, 2015.

[10] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi. The evolution of popular music: USA 1960-2010. *Royal Society Open Science*, 2(5):150081, 2015.

[11] C. McKay and I. Fujinaga. Automatic genre classification using large high-level musical feature sets. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 525–530, 2004.

[12] D. Moelants, O. Cornelis, and M. Leman. Exploring African Tone Scales. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 489–494, 2009.

[13] B. Nettl. *The Study of Ethnomusicology: Thirty-one Issues and Concepts*. University of Illinois Press, Urbana and Chicago, 2nd edition, 2005.

[14] B. Nettl, R. M. Stone, J. Porter, and T. Rice, editors. *The Garland Encyclopedia of World Music*. Garland Pub, New York, 1998-2002 edition, 1998.

[15] J. Novembre, K. Bryc, S. Bergmann, A.R. Boyko, C.D. Bustamante, A. Auton, M. Stephens, Z. Kutalik, A. Indap, T. Johnson, M.R. Nelson, and K.S. King. Genes mirror geography within Europe. *Nature*, 456(7218):98–101, 2008.

[16] M. Panteli, R. Bittner, J. P. Bello, and S. Dixon. Towards the characterization of singing styles in world music. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 636–640, 2017.

[17] P. E. Savage and S. Brown. Mapping Music: Cluster Analysis Of Song-Type Frequencies Within and Between Cultures. *Ethnomusicology*, 58(1):133–155, 2014.

[18] P. E. Savage, S. Brown, E. Sakai, and T. E. Currie. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, 112(29):8987–8992, 2015.

[19] D. Turnbull and C. Elkan. Fast recognition of musical genres using RBF networks. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):580–584, 2005.

[20] G. Tzanetakis, A. Ermolinskyi, and P. Cook. Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research*, 32(2):143–152, 2003.

[21] W. Yang, J. Novembre, E. Eskin, and E. Halperin. A model-based approach for analysis of spatial structure in genetic data. *Nature Genetics*, 44(6):725–731, 2012.

[22] J. M. Young, L. S. Weyrich, J. Breen, L. M. Macdonald, and A. Cooper. Predicting the origin of soil evidence: High throughput eukaryote sequencing and MIR spectroscopy applied to a crime scene scenario. *Forensic Science International*, 251:22–31, 2015.

[23] F. Zhou, Q. Claire, and R. D. King. Predicting the Geographical Origin of Music. In *IEEE International Conference on Data Mining*, pages 1115–1120, 2014.