

Visualization of Joint Spatio-temporal Models via Feature Recognition with an Application to Wildland Fires

Devan G. Becker¹ ^a, Douglas G. Woolford¹ and Charmaine B. Dean²

¹Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada

²Statistics and Actuarial Science, Waterloo University, Waterloo, Ontario, Canada

Keywords: Image Recognition, Non-negative Matrix Factorization, Log-Gaussian Cox Processes, Dimension Reduction.

Abstract: Many spatial statistics applications result in a collection of spatial estimates, especially if a different (but possibly correlated) estimate is produced for a sequence of time epochs. For a small collection of epochs, the connections or trends between estimates and the prominent or common features can be found via inspection of the spatial estimates. As the number of spatial estimates grows, this task becomes much more difficult. We present a method of summarizing a sequence of estimates using an image recognition technique called Non-Negative Matrix Factorization which results in a meaningful decomposition of the source images into basis functions and coefficients. This visualization technique allows for investigation of trends over time as well as common spatial features of the estimates without needing to fit a temporal model or use pre-specified spatial regions. We apply this technique to a sequence of models that jointly model the spatial location of wildland fires with the total burn area of each of the fires. We discuss the extensions of the visualization technique to the joint modelling framework and are able to draw new insights about the connection between the location and size of the fires.

1 INTRODUCTION

Sequences of continuous spatial fields are becoming more common with more data collection. These can come in the form of spatial fields measured or estimated at discrete time points (e.g. sea surface temperature measured daily, yearly estimates of flood risk, etc.), multiple variables measured or estimated over the same spatial regions (e.g. distribution of different species across the same habitat), or some combination of the two (e.g. species distribution measured monthly). Much work has been done to estimate multivariate spatio-temporal models, but these models are easy to misspecify and difficult to estimate. This is especially true in the presence of a large number of spatial fields.

Some recent examples of analyses that resulted in a collection of joint spatial and spatio-temporal models are as follows. A joint spatial model for predator-prey relationships of marine species was fit for each year in Sadykova et al. (2017). They found that most of the covariates which were significant for habitat usage were likely to change with the changing climate.

Jones-Todd et al. (2018) fit a joint spatio-temporal model to determine predator-prey relationships in avian species. Their model accounted for spatio-temporal variation, but the results of the analysis still included a large number of spatial plots to be interpreted. Finally, Python et al. (2016) fit a joint spatial model to yearly terrorist attacks around the globe.

In previous work, we performed a spatial analysis of wildland fires (Becker et al. 2020). This involved jointly modelling the location of fires along with the size of those fires. Locations of fires were modelled using a Log-Gaussian Cox Process (LGCP) framework and sizes were modelled using a Log-Normal survival model with assumed interval censoring; a shared random effect was used to jointly model these two outcomes. Due to computational complexity as well as the winter creating a discontinuity between fire seasons, we restricted our data to one year at a time. This model setup resulted in two spatial estimates per year for 47 years worth of data. While not excessive, it was still difficult to see broad trends with so many spatial estimates.

Here, we employ feature recognition algorithms to summarise the joint spatial fields into basis functions. Such techniques have been used for analysis of shot

^a  <https://orcid.org/0000-0003-3796-3946>

location in both basketball (Miller et al. 2014; Franks et al. 2015) and hockey (Becker, Woolford, and Dean 2020). Those papers estimated a spatial point process estimate for each player, then treated these estimates as images to determine how players utilized regions of the basketball court or hockey rink. The results illuminated similarities and differences between players that would have been very difficult to discern with the naked eye. The authors all used this analysis to create some measure for shot optimality. In what follows we describe an image recognition technique and demonstrate that applying it to estimates of a sequence of spatial models is a useful visualization technique.

2 PRELIMINARIES

2.1 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF) is a dimension reduction technique that decomposes a data matrix into a matrix of basis functions and a matrix of coefficients for those basis functions. This was popularized as an image recognition technique by Lee and Seung (1999), and has been used in a wide variety of applications since (see, e.g., Gillis 2014).

For our purposes, NMF has the attractive feature of being purely additive. That is, the estimated basis functions as well as their coefficients are both non-negative, so a linear combination of the bases can only add to the estimate. In PCA, the bases and coefficients and bases are allowed to be negative, so one basis may be allowed to counteract the effect of another. With purely additive bases, the basis functions all represent a single feature. This restriction makes the estimated basis functions directly interpretable.

The NMF algorithm works by factorizing an $n \times m$ matrix V with non-negative entries into an $n \times r$ matrix W and an $r \times m$ matrix H , where r is the number of basis functions and must be specified prior to estimation. The columns of W represent the basis functions. Each row in the matrix H represents the coefficients for the corresponding basis in W . With these matrices, we can approximate the i th column of V , which we will denote $V_{:,i}$, as the matrix product $WH_{i,:}$.

Under certain constraints, NMF is particularly useful for feature recognition in images (Gillis (2014)). Assuming all images have the same pixel dimensions and the colour of each pixel is a single, non-negative number (e.g., grey scale), a matrix of images can be constructed such that each column represents a single image. For instance, suppose we have

a collection of grey scale images of faces, all of which have the same pixel dimensions and the faces are all aligned in the same way (e.g. all eyes and mouths are at the same location of the image). Each image can be represented by a vector of non-negative numbers. The NMF algorithm will estimate basis functions that correspond to facial features. The coefficient matrix will determine how much of each basis function is required to construct a face.

To make this process more clear, consider the following example. Suppose we have a matrix V with $n = 7$ rows and $m = 2$ columns, resulting in 14 entries total. The first column is $[1, 2, 3, 4, 3, 2, 1]^T$ and the second column is $[2, 3, 5, 7, 5, 3, 2]^T$. Clearly, both columns have a similar pattern (or feature). If we want to characterize this feature, we could use an NMF decomposition, where we choose $r = 1$ since we know there is a single feature. Doing so results in $W = [1.259, 2.099, 3.359, 4.618, 3.359, 2.099, 1.259]^T$, which is a matrix with seven rows (n) and one column (r). The coefficient matrix is $H = [0.886, 1.496]$. From this, the approximation for the first column of V is $W \times 0.886 = [1.116, 1.86, 2.977, 4.093, 2.977, 1.86, 1.116]^T$, which is quite close to the original first column (but with some approximation error). Note that W and H together have 9 entries compared to the original 14 and that inspection of W tells us about both columns of the original matrix simultaneously.

The choice of r , the number of bases, also known as the rank, is non-trivial. Too many basis functions and the algorithm will simply be modelling the noise. Too few and the approximations will not be accurate. In some contexts, prior knowledge will be sufficient for choosing the number of bases. In other contexts there are numerous heuristic approaches. Techniques have been proposed by Brunet et al. (2004), Hutchins et al. (2008), and Frigyesi and Höglund (2008). A properly motivated choice of r is imperative whenever NMF is being used for analysis. As a visualization technique, however, the choice of rank is dependent on the usefulness of the visualizations.

Estimation of NMF models has been shown to be NP-hard (Vavasis 2007). There have been many algorithms developed to estimate the matrices (Wang and Zhang 2013), and these methods have been implemented in multiple software packages. We use the NMF package in the R Statistical Software language, and details can be found in Gaujoux and Seoighe (2010).

2.2 NMF for Spatial Fields

To apply the NMF algorithm to estimates of a spatial field, we will follow the algorithm of Miller et al. (2014) in their analysis of shots in professional basketball. Their algorithm broadly follows four steps:

1. Set up a matrix to represent the pixel locations, then convert this matrix to a vector. For an $q \times s$ pixel resolution, the first row of this matrix can be labelled $\underline{p}_1 = p_{11}, p_{12}, \dots, p_{1s}$ and the k th row can be labelled $\underline{p}_k = p_{k1}, p_{k2}, \dots, p_{ks}$. The vector of pixels will then have the form $\underline{p} = (\underline{p}_1, \underline{p}_2, \dots, \underline{p}_q)$.
2. Estimate each of the N models at the locations \underline{p} . The j th model is labelled $\underline{\lambda}_j(\underline{p})$, which we will shorten to $\underline{\lambda}_j$ for convenience. Note that these estimates must be non-negative.
3. Combine the vectors of spatial models as column vectors in a matrix, $\mathbf{\Lambda} = [\underline{\lambda}_1, \underline{\lambda}_2, \dots, \underline{\lambda}_N]$, where each $\underline{\lambda}_j$ is a column.
4. Use an NMF algorithm to approximate matrices W and H such that $\mathbf{\Lambda} \approx WH$.

Due to the setup of this algorithm, the columns of W will be spatial basis functions at pixels \underline{p} and the j th column of H will be a vector with r elements which represent the contribution of each basis to the approximation of $\underline{\lambda}_j$.

2.3 Shared Spatial Effects Models

The methods developed in this paper are applicable to any joint modelling framework where a spatial field is estimated, but are especially useful for models with a spatial effect that is split into two.

Suppose we have a joint model of the form $f_{X,Y}(x(s), y(s)|Z(s), \Lambda(s), \phi)$, where $x(s)$ and $y(s)$ are our spatially referenced response variables, $Z(s)$ and $\Lambda(s)$ are two estimated spatial fields, and $\phi = (\phi_X, \phi_Y)$ where ϕ_X and ϕ_Y are the vectors of remaining model parameters for X and Y , respectively, which may or may not include further spatial terms and may have some identical elements. Suppose further that our model can be factored as:

$$\begin{aligned} f_{X,Y}(x(s), y(s)|Z(s), \Lambda(s), \phi) \\ = f_X(x(s)|Z(s), \Lambda(s), \phi_X) f_Y(y(s)|Z(s), \phi_Y) \end{aligned} \quad (1)$$

That is, conditional on the random field $Z(s)$, $X(s)$ and $Y(s)$ are independent. In this formulation, it is entirely possible that one of the elements of ϕ_Y is another spatial field. In fact, the visualization technique that we will develop later is immediately extensible

to such a situation. For now, we are primarily interested in a model that contains a model-specific spatial component as well as a shared spatial component.

For the purposes of this study, we do not need to specify an estimation procedure. It is imperative that the estimates from this technique are reasonable for the visualization technique to be useful, but the visualizations that we will present are agnostic to the particulars of the estimation procedure. In fact, the general framework of the visualization techniques does not require an estimate at all; it will suffice to have any combination of estimates and/or spatially and/or temporally referenced data.

3 NMF FOR JOINT SPATIAL MODELS

Suppose we have a sequence of N non-negative estimates of spatial fields $Z(s)$ and $\Lambda(s)$ estimated at pixel locations \underline{p} . We will define $\hat{\mathbf{Z}} = [\hat{Z}_1(\underline{p}), \hat{Z}_2(\underline{p}), \dots, \hat{Z}_N(\underline{p})]$ as the matrix of spatial estimates defined in Section 2.2, with $\hat{\mathbf{\Lambda}}$ defined similarly.

There are several potential ways to apply the NMF algorithm to these joint spatial models. The naive approach would be to approximate $\hat{\mathbf{Z}} \approx W^{(Z)}H^{(Z)}$ and $\hat{\mathbf{\Lambda}} \approx W^{(\Lambda)}H^{(\Lambda)}$. This would result in a set of basis functions for $\hat{\mathbf{Z}}$ denoted $W^{(Z)}$, and a set of basis functions for $\hat{\mathbf{\Lambda}}$, $W^{(\Lambda)}$. These basis functions would be entirely separate and would miss pertinent shared features.

Given the joint spatial modelling approach, we desire a method for summarising the two fields that retains any joint features. For instance, if large values in $Z(s)$ tend to correspond to large values in $\Lambda(s)$, we would like our visualizations to reflect this.

To achieve this goal, we can stack the $\hat{\mathbf{Z}}$ and $\hat{\mathbf{\Lambda}}$ matrices as follows:

$$\mathbf{V} = \begin{bmatrix} \hat{Z}_1(\underline{p}) & \hat{Z}_2(\underline{p}) & \dots & \hat{Z}_N(\underline{p}) \\ \hat{\Lambda}_1(\underline{p}) & \hat{\Lambda}_2(\underline{p}) & \dots & \hat{\Lambda}_N(\underline{p}) \end{bmatrix} \quad (2)$$

In this construction, each column represents both spatial fields present in the joint modelling approach. From this construction, any given basis function contains information about both spatial fields. It is trivial to extend this to any number of spatial fields (assuming the fields are well estimated).

Alternatively, one could combine the spatial fields side-by-side:

$$\mathbf{V}^{(alt)} = \begin{bmatrix} \hat{Z}_1(\underline{p}) & \hat{Z}_2(\underline{p}) & \dots & \hat{Z}_N(\underline{p}) & \hat{\Lambda}_1(\underline{p}) \\ & \hat{\Lambda}_2(\underline{p}) & \dots & \hat{\Lambda}_N(\underline{p}) & \end{bmatrix} \quad (3)$$

In this construction, \hat{Z} and $\hat{\Lambda}$ are still estimated separately but will rely on the same basis functions. To visualize the relationship between these two fields, one could inspect the coefficients.

To visualize common similarities between spatial field, we believe that the “stacked” approach in Equation (2) is more useful than the side-by-side approach in Equation (3). The approximation to the estimates will incorporate both spatial fields rather than drawing from each field separately. The interpretation of the basis functions will make it immediately clear how the spatial fields are related.

4 APPLICATION TO WILDLAND FIRES

Our data consist of the locations and total burn area of fires from 1953 to 2000 in the province of British Columbia, Canada. The model that we will be summarising is given in full detail in Becker et al. (2020). A brief description follows below.

The total burn area (in hectares) of a fire is modelled as a log-normal random variable such that the mean depends on an intercept, fire weather covariates, and a spatial component. We assume that the spatial component of this distribution can be modelled by a Gaussian field $\Lambda(s)$.

The location of lightning-caused wildland fires is modelled by a log-Gaussian Cox process. This model assumes that there is an underlying Gaussian field $S(s)$, and conditional on this field the number of points in a region B is Poisson with rate parameter $\int_B \exp(\beta_0 + S(s) + C(s)) ds$, where β_0 is an intercept and $C(s)$ is the collection of spatial covariates (including distance to the nearest highway or roadway and elevation).

To link these two models, we separate $S(s)$ into two independent Gaussian fields $S_1(s)$ and $S_2(s)$ such that $\exp(S(s)) = \exp(S_1(s) + \gamma S_2(s)) = Z(s)\Lambda(s)^\gamma$. We chose this notation to make it explicit that $Z(s)$ and $\Lambda(s)$ are both non-negative. The linking parameter γ exists so that the estimation procedure can cause the joint component to vanish from the LGCP model while retaining a spatial field in the size model.

To demonstrate the results of this model, a comparison of the model output versus separate non-parametric estimates is shown in Figure 1. The LGCP random effect is clearly estimating the spatial distribution of fires, and the joint component is estimating the spatial variation in size. Note that the colours are normalized such that the maximum value is the brightest spot and should not be used for comparison. The plot on the left (including both the LGCP and

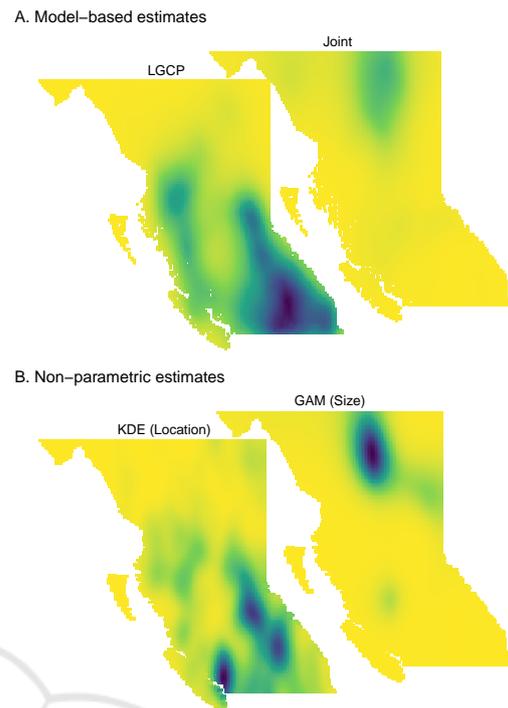


Figure 1: A. Spatial field estimates from the joint spatial model for wildland fires in 1990. B. Spatial non-parametric estimates of the locations and sizes for the same data. The colour scale is chosen for so that the largest values of the given field are the brightest colour and are not meaningful for comparison.

the Joint components) was converted into a vector of pixels and combined with all of the other estimates from this model. The NMF algorithm was run 100 times with different initial values at $r = 3$ to $r = 20$. Based on the peaks in the cophenetic, dispersion, and silhouette plots, and the “elbow” in the residual sum of squares plot (as described in Chalise and Fridley 2017), we chose 8 basis functions (either 6 or 9 would have also been supported by the diagnostic plots; they were not as definitive as this makes it sound). Upon visual inspection, this appears to retain interpretability while avoiding “modelling noise”.

The resultant basis functions and coefficients are shown in Figure 2. The colours will be described in the next section. Recall that the bases are additive, so all of the original estimates can be approximated by adding together the non-negative bases.

The NMF algorithm allows the basis functions to overlap, so there are multiple bases that cover the same regions. In particular, the mountainous area near the south east of British Columbia is partially covered by bases 1, 2, 5, and part of 8. In contrast, it appears that the diagonal line down the south east of BC in basis 1 is cut out of basis 6, which has a conspicuously low value in the same area.

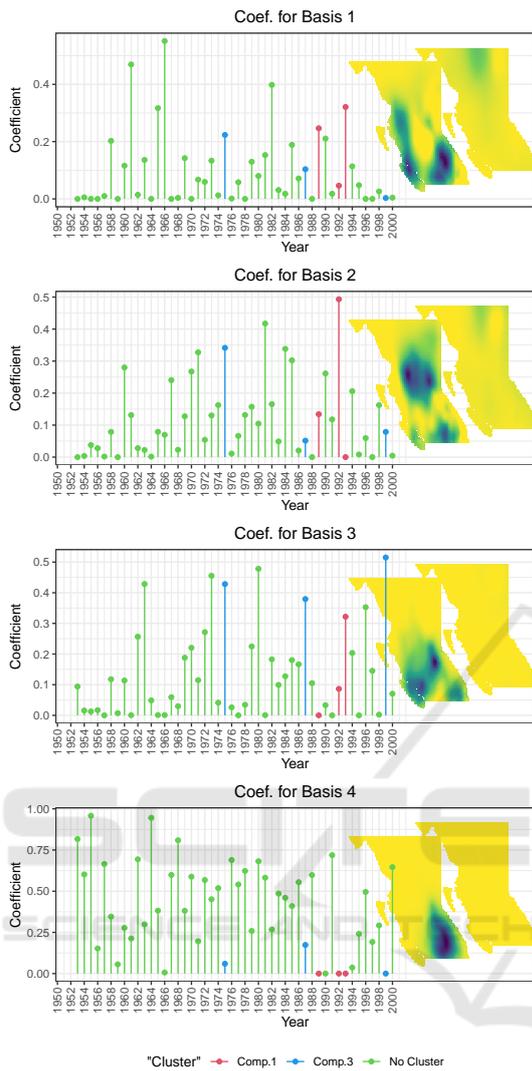


Figure 2: Bases and Coefficients for bases 1 to 4. Within each basis plot (the maps), the lower-left maps are the LGCP-only effect and the upper-right maps are the shared effects.

Basis 7 demonstrates that the estimated basis functions are not required to be contiguous. The multiple regions in basis 7 represent places where fires tend to ignite (or not ignite) in the same year. Much like how bases 1 and 6 fit together, basis 7 also appears to be fit with basis 8 like a puzzle piece.

There appears to be a stark contrast between years where fires occur in the mountainous region in the east-by-south-east region, as described by basis 1, and years where there are no fires in this region. The coefficients indicate that basis 1 is either a large component of the LGCP estimate or it is not part of it at all.

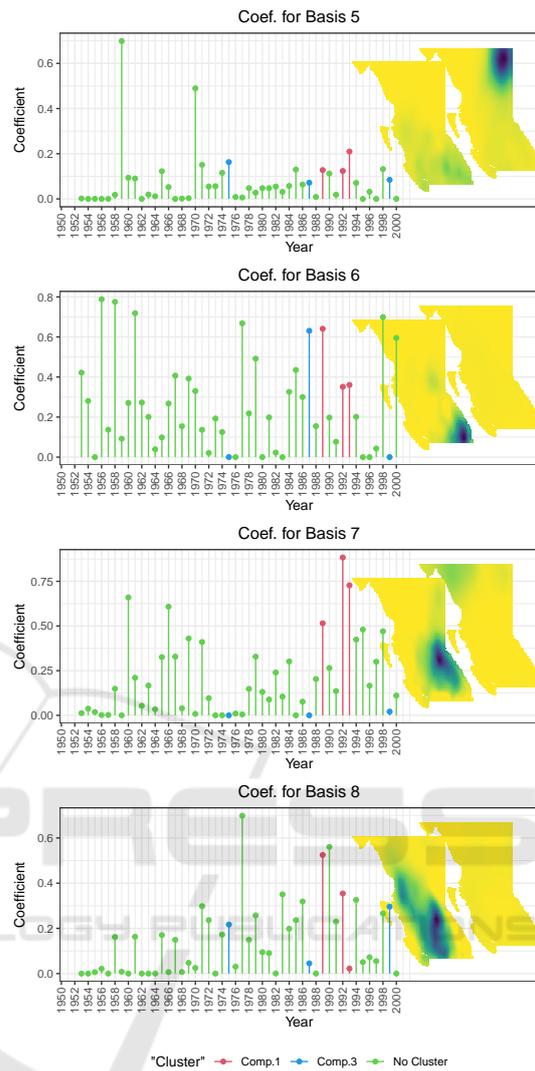


Figure 3: Bases and Coefficients for bases 5 to 8.

The difference in coefficients for different years reveals some potential patterns over time. This is most evident in the increase in the coefficient for basis 2 over time. This increase is mirrored by the decrease in the coefficient for basis 5 over time. In particular, the coefficient for basis 2 is much higher after 1985 than before, whereas the coefficient for basis 5 is lower after 1985. Judging by the basis functions, this corresponds with more fires in the middle of the province rather than the south east.

For the joint component, we see that the algorithm has split the northern area into west (basis 1), centre (basis 7), and east (basis 3). No other bases have positive values, indicating that the joint effect only appears to exist in the northern regions of BC. It is also noteworthy that the joint effects are associated strongly with the LGCP-only effects.

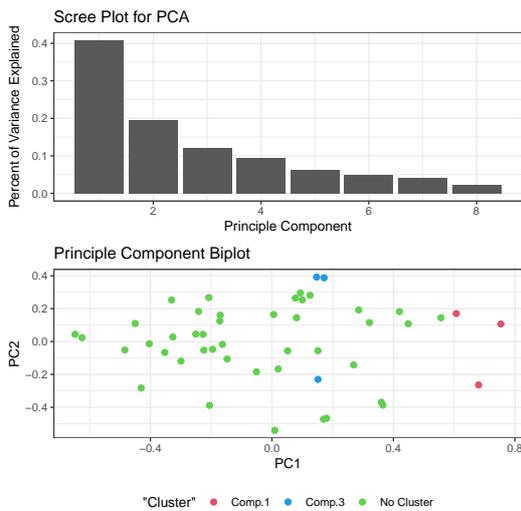


Figure 4: Scree plot and biplot, showing the colours that were used in Figure 2. The points are coloured based on how “abnormal” they are in comparison to the other points. “Abnormality” was determined by visual inspection. The green points are abnormal in terms of principle component 3, which is not shown.

The colours in Figure 2 are based on a principle components analysis of the estimated coefficients. The scree plot and the biplot are shown in Figure 4. The “clusters” are based on simply checking whether a given point has a principle component larger than a specified value, where the specified value was chosen by hand. This is not a rigorous outlier detection method, but rather another visualization technique. The green points, labelled “Comp.3”, are from the same sort of heuristic in the direction of the third principle component.

5 CONCLUSION

We have shown that stacking multiple spatial estimates in Non-negative Matrix Factorization is an effective technique for summarising a sequence of joint spatial estimates. The algorithm setup that we have described can easily be extended to numerous applications where a large number of spatial estimates must be interpreted by eye.

For our particular application, the NMF technique quickly revealed that the shared random field primarily acts in the northern regions of BC. This provides insight into the relationship between fire ignition, spread, weather, and suppression efforts. Intuitively, weather that is conducive to large fires would also be conducive to multiple ignitions. This relationship is characterized by the joint modelling approach. However, fires are only allowed to grow un-

abated in the northern regions of BC. The fires in the more populated southern regions are more likely to be suppressed, which confounds the relationship between count and size.

The analysis also revealed years where the fire behaviour was different from other years. The increase in the value of the coefficient for basis 2 is especially interesting. In Becker et al. (2020), we found a significant effect for distance to highway or railway, with this effect approaching 0 over time. We believe that this may be due to imperfect detection of fires that are far from human populations. The increase in coefficient for basis 2 appears to support this conclusion; the region of British Columbia covered by the basis saw an increase in population over time. The increased importance of this basis may indicate better detection of lightning-caused fires, rather than a change in fire behaviour.

The technique that we described is not exclusive to joint models in the framework of our research. The techniques would be valid if one or both of the spatial fields were observed perfectly. The technique also works for spatial models that are estimated separately, in which case the basis functions and coefficients would help deduce spatial correlation between the variables of interest. There is no theoretical constraint on N , the number of observations of spatial fields, nor is there a constraint on the number of types of spatial fields that are stacked on top of each other.

As with any other application of NMF without pre-defined bases, this technique is limited by the need to investigate the rank. As a visualization technique it is worthwhile to investigate multiple values of r , but this may be tedious and/or time consuming depending on the size of the data.

NMF is frequently used as a clustering technique, which is often interpreted as meaning that there is the intention of prediction of new observations or inference for population parameters. As described in this paper, neither of these interpretations are valid. We are applying NMF to point estimates from another model while ignoring the rest of the variance. The conclusions from this paper are broad trends, and further investigation into individual model outputs is required before and predictions or inferences can be made. This visualization technique is not intended to replace careful inspection of model output and interpretation of model parameters. Instead, it is a single, useful step in the long process of spatio-temporal model estimation.

ACKNOWLEDGEMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference numbers RGPIN-2015-04221 and RGPIN-2014-06187]. Additional support was provided by a CANSSI Collaborative Research Team grant and the Institute for Catastrophic Loss Reduction. We would also like to thank Michael Schuckers and Nathan Sandholtz for conversations regarding the work on NMF for hockey data and Steve Taylor for his insight into the work on wildland fire modelling.

REFERENCES

- Becker, Devan G., Douglas G. Woolford, and Charmaine B. Dean. 2020. "Algorithmically Deconstructing Shot Locations as a Method for Shot Quality in Hockey." *Journal of Quantitative Analysis in Sports* Ahead of Print.
- Becker, Devan G., Douglas G. Woolford, Charmaine B. Dean, and W. John Braun. 2020. "Visualization and Joint Analysis of Monitored Multivariate Spatio-Temporal Data with Applications to Forest Fire Modelling and Sports Analytics." *Western University Electronic Thesis and Dissertation Repository*, March.
- Brunet, J. P., P. Tamayo, T. R. Golub, and J. P. Mesirov. 2004. "Metagenes and Molecular Pattern Discovery Using Matrix Factorization." *Proceedings of the National Academy of Sciences* 101 (12): 4164–9. <https://doi.org/10.1073/pnas.0308531101>.
- Chalise, Prabhakar, and Brooke L. Fridley. 2017. "Integrative Clustering of Multi-Level 'Omic Data Based on Non-Negative Matrix Factorization Algorithm." Edited by Shyamal D. Poddada. *PLOS ONE* 12 (5): e0176278. <https://doi.org/10.1371/journal.pone.0176278>.
- Franks, Alexander, Andrew Miller, Luke Bornn, and Kirk Goldsberry. 2015. "Characterizing the Spatial Structure of Defensive Skill in Professional Basketball." *The Annals of Applied Statistics* 9 (1): 94–121. <https://doi.org/10.1214/14-AOAS799>.
- Frigyesi, Attila, and Mattias Höglund. 2008. "Non-Negative Matrix Factorization for the Analysis of Complex Gene Expression Data: Identification of Clinically Relevant Tumor Subtypes." *Cancer Informatics* 6 (January): CIN.S606. <https://doi.org/10.4137/CIN.S606>.
- Gaujoux, Renaud, and Cathal Seoighe. 2010. "A Flexible Software Package for Nonnegative Matrix Factorization," 9.
- , Nicolas. 2014. "The Why and How of Nonnegative Matrix Factorization." *arXiv:1401.5226 [Cs, Math, Stat]*, March. <http://arxiv.org/abs/1401.5226>.
- Hutchins, Lucie N., Sean M. Murphy, Priyam Singh, and Joel H. Graber. 2008. "Position-Dependent Motif Characterization Using Non-Negative Matrix Factorization." *Bioinformatics* 24 (23): 2684–90. <https://doi.org/10.1093/bioinformatics/btn526>.
- Jones-Todd, Charlotte M., Ben Swallow, Janine B. Illian, and Mike Toms. 2018. "A Spatiotemporal Multi-species Model of a Semicontinuous Response." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67 (3): 705–22. <https://doi.org/10.1111/rssc.12250>.
- Lee, Daniel D., and H. Sebastian Seung. 1999. "Learning the Parts of Objects by Non-Negative Matrix Factorization." *Nature* 401 (6755): 788–91. <https://doi.org/10.1038/44565>.
- Miller, Andrew, Luke Bornn, Ryan Adams, and Kirk Goldsberry. 2014. "Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball." *arXiv:1401.0942 [Stat]*, January. <http://arxiv.org/abs/1401.0942>.
- Python, André, Janine Illian, Charlotte Jones-Todd, and Marta Blangiardo. 2016. "A Bayesian Approach to Modelling Fine-Scale Spatial Dynamics of Non-State Terrorism: World Study, 2002-2013." *arXiv:1610.01215 [Stat]*, October. <http://arxiv.org/abs/1610.01215>.
- Sadykova, Dinara, Beth E. Scott, Michela De Dominicis, Sarah L. Wakelin, Alexander Sadykov, and Judith Wolf. 2017. "Bayesian Joint Models with INLA Exploring Marine Mobile Predator-Prey and Competitor Species Habitat Overlap." *Ecology and Evolution* 7 (14): 5212–26. <https://doi.org/10.1002/ece3.3081>.
- Vavasis, Stephen A. 2007. "On the Complexity of Nonnegative Matrix Factorization." *arXiv:0708.4149 [Cs]*, September. <http://arxiv.org/abs/0708.4149>.
- Wang, Yu-Xiong, and Yu-Jin Zhang. 2013. "Nonnegative Matrix Factorization: A Comprehensive Review." *IEEE Transactions on Knowledge and Data Engineering* 25 (6): 1336–53. <https://doi.org/10.1109/TKDE.2012.51>.