

Ther varom mid j hia: Tracing linguistic diffusion in the history of Norwegian using kernel density estimation

Abstract

Tracing the diffusion of linguistic innovations in space from historical sources is challenging. The complexity of the datasets needed in combination with the noisy reality of historical language data mean that it has not been practical until recently. However, bigger historical corpora with richer spatial and temporal information allow us to attempt it. This paper presents an investigation into changes affecting first person non-singular pronouns in the history of Norwegian: first, individual changes affecting the dual (*vit* > *mit*) and plural (*vér* > *mér*), followed by loss of the dual-plural distinction by merger into either form or replacement of both by Danish-Swedish *vi*. To create dynamic spatial visualisations of these changes, the use of *kernel density estimation* is proposed. This term covers a range of statistical tools depending on the kernel function. The paper argues for a Gaussian kernel in time and an adaptive uniform (*k*-nearest neighbours) kernel in space, allowing uncertainty or multiple localisation to be incorporated into calculations. The results for this dataset allow us to make a link between Modern Norwegian dialectological patterns and language use in the Middle Ages; they also exemplify different types of diffusion process in the spread of linguistic innovations.

Keywords: historical dialectology, Norwegian, kernel density estimation, statistical methods, GIS

Summary

This paper proposes a new approach to a central methodological problem in historical dialectology: how to trace the changing distributions of linguistic features in space from noisy historical sources. Previous approaches to visualising historical dialectological data are surveyed, demonstrating that there is no consensus solution to this problem in the literature. For this paper, the problem and proposed solution are exemplified by an investigation into changes which affected first person dual and plural pronouns in the history of Norwegian. Old Norwegian dual *vit* and plural *vér* first become *mit* and *mér* before the dual-plural distinction is lost through three simultaneous and competing changes: replacement of *mit* by *mér*; replacement of *mér* by *mit*; and replacement of both by the borrowed form *vi*. A large dataset (comprising 16,701 tokens across 5939 texts) is collected from a corpus of dated and localised Middle Norwegian charters.

The method proposed for producing visualisations of the shifting spatial distributions behind this complex of changes is *kernel density estimation* (KDE). KDE is the name given to a family of nonparametric statistical methods for estimating some underlying function at a given point in space from sample data. It is calculated simply by taking a weighted average of all the data where the nearby (and accordingly most relevant) samples are given the highest weights and the distant (and accordingly least relevant) samples are given the lowest weights. In the case of historical dialectology, distance in both space and time must be taken into account. The paper discusses variations on this method which differ primarily by their *kernel function*, which specifies precisely how weights are calculated from distances. A specific form best suited for historical dialectology is proposed and argued for, as are ways by which to tune parameters for a given dataset.

The results of the method on the Middle Norwegian pronoun data are then presented. The early changes start in rural areas and show a counterhierarchical diffusion pattern, with cities lagging behind the countryside. The merged form *mit* then spreads within the south by contagion diffusion in an area clearly corresponding to its modern Norwegian dialectal reflex *mi*, whilst the borrowed form *vi* spreads from Sweden via a clear hierarchical pattern, with cities leading the change. Thus the method has produced a result which is both historically plausible and exemplifies all three hypothesised types of linguistic diffusion process.

1. Introduction

How to infer the changing distributions of linguistic features in space from their occurrence in texts is a longstanding methodological problem in historical dialectology. Although there are many languages with long-term traditions of writing for which we have diachronic data distributed in space, such data is typically unevenly distributed and noisy, making mapping difficult.

As a result, what we know about the dynamics of language changes in space is actually largely the result of research on synchronic data. The ‘wave model’ (German *Wellentheorie*, also called *contagion diffusion*), which posits that an innovative form spreads outwards continuously from a point of innovation like a ripple (Schmidt 1872), was inspired by examining the distributions of features in the synchronic records of Indo-European languages. The ‘gravity model’, which proposes hierarchical diffusion in which innovations instead spread discontinuously between denser population centres (Trudgill 1974), was inspired by distributions of vowels in synchronic Modern Norwegian data. Proposed ‘counterhierarchical diffusion’, in which features spread discontinuously between regions of low population density, was inspired by the synchronic distribution of dialectal phenomena in Oklahoma English (Bailey et al. 1993; Wikle and Bailey 1997).

Statistical modelling of the diffusion of language change does suggest that all three of these types should exist: language use happens in social networks which (particularly historically) are fundamentally local, implying continuous diffusion of new forms through space; centres of population density should be able to maintain different norms to the surrounding low-density areas, predicting both hierarchical diffusion and counterhierarchical diffusion depending on the point of innovation (see Burridge 2017; Burridge 2018). These theories should find their clearest (dis)confirmation in time-series maps of diachronic data, yet for purely methodological reasons such visualisations have been hard to achieve. This paper will investigate changes in the history of Norwegian first person plural and dual pronouns by using a statistical tool underutilised in historical linguistics to create diachronic, geographical visualisations: kernel density estimation (KDE).

1.1 First person non-singular pronouns in the history of Norwegian

Dual and plural pronouns have undergone a complex set of changes in the history of Norwegian. In Old Norwegian, we find *vér* for the first person nominative plural and *vit* for the first person nominative dual (Noreen 1970: 309). Reanalysis of the word-boundary between the pronoun the preceding verbal ending *-um* then gave pl. *mér* and du. *mit* in Middle Norwegian (Noreen 1970: 202–203; Indrebø 1951: 122; Seip 1955: 194–195, 317; Haugen 1976: 302–303). The form *vi* was borrowed from East Nordic, which had no dual-plural distinction. Modern Norwegian dialects, none of which preserve the dual-plural distinction, use a variety of forms for their merged plural: *vi*, *me* and *mi* are the most common (Jahr 1990). Dialect usage as described by articles in Jahr (1990) (background colour) and as demonstrated by speakers in the Nordic Dialect corpus (Johannessen et al. 2009) (point colour) are visualised in Figure 1.

Figure 1: Form of the 1.pl. subject pronoun in Modern Norwegian dialects from Jahr (1990) and NDC (Johannessen et al. 2009)

A series of research questions remain to be answered about this history:

1. precisely when did *mér* and *mit* arise, when did *vér* and *vit* disappear, when did *vi* diffuse into Norwegian, and when was the dual-plural distinction lost?
2. where were *mér* and *mit* innovated and by what pathway did they diffuse?
3. what was the route by which *vi* spread into Norwegian? did it spread from Swedish or Danish?
4. do *me* and *mi* reflect *mér* and *mit*, meaning that different dialects selected the plural (*mér*) and dual (*mit*) as the merged form?

1.2 Historical dialectology

Historical dialectology is the study of linguistic variation in space over time. The aim of a historical dialectological study might be to identify the position of isoglosses, the borders of domains within which there is consistent use and across which use differs; it might be to identify how the positions of isoglosses changed over time (i.e. how new forms diffuse through space); it might be to identify more complex distributions, with variable usage within a geographical domain. In a perfect world, we would have rich data on language use from many, evenly distributed localities across the geographical domain(s) of interest; if our study also examined change over time, these data would also be evenly distributed over the entire period of interest. In such a case, we could simply visualise the data on one or more maps and see the distribution of the linguistic features of interest.

In practice, our data are never like this. Working on any period before the twentieth century, our linguistic data are in the form of written texts, and text-production has never been geographically or temporally even. As a result of the uneven distribution of wealth, of centres of religious learning and of mercantile activities, we typically have clusters of texts from particular localities in certain periods alongside regions and times from there is little data. Text survival is also a non-random filter, offering a second explanation for the uneven geographical and temporal distribution of extant texts. These factors alone make identifying distributions in language use, which existed whether or not written texts were being produced, difficult.

The other major difficulty is statistical noise. Dating texts can be something of a dark art, with dates based on contemporary references, hand, or references in other sources. Given this, we can be sure that there are often errors in our proposed text dates. The same applies to localisation. Even where we are totally confident of where a text was *produced*, scribes are often anonymous. In such a case, it is always possible that the language of a text should properly be localised to the home-locality of an unknown scribe. Thus our localisations must also often be wrong. Finally, language itself is noisy. For the features in which we are interested, usage within a given community is rarely categorical. With only small samples per locality and time, we can never be totally confident that a given example really represents the 'norm' or majority variant of a given time and place. This is further complicated by writing, since conservative or external written norms might disrupt representation of a linguistic form in unpredictable and inconsistent ways.

For all these reasons, simply visualising raw data is often impossible or unhelpful. We might get an impression of a distribution but be unsure whether we are simply interpreting noise. Equally often, we might simply see so chaotic and complex a picture that we cannot get any subjective impression of a distribution. When we are dealing with diachronic data, these problems are multiplied since it is difficult to construct readable visualisations of point data across time, even if

the distributions are clear. In spite of all this, simply visualising raw data has often been the only tool used in historical dialectology. An example is Studer-Joho (2014), with visualisations of data from the Linguistic Atlas of Early Middle English. As can be seen in Figure 2, the need to use all of symbol position, shape and colour (to indicate spatial position, rate of use of the linguistic variant and date, respectively) results in a chaotic visualisation that is difficult to read with confidence. Studer-Joho concludes that the innovation ($\bar{a} > \bar{o}$) diffused south to north in the West Midlands, but this is far from unmistakable.

Figure 2: Distribution of <a> and <o> spellings for OE \bar{a} in LAEME, reproduced from (Studer-Joho 2014: 222)

There are historical dialectological studies which show near-categorical distributions of raw linguistic data in space, from which it is accordingly easy to read the position of isoglosses. However, this may be achieved by questionable methods. The Linguistic Atlas of Early Middle English (LAEME; Laing 2013) and the Linguistic Atlas of Late Medieval English (LALME; Benskin et al. 2013) provide some instructive examples. Figure 3 shows the distribution of *h*- forms of the 3.pl. pronoun in Middle English from LALME: we see two, clearly defined dialect regions in which *h*- forms are used, a larger region in which usage is variable, and a larger region still where *h*- forms do not occur. There are two reasons to be concerned about this reading of the data, however. Firstly, many of the individual texts in LALME cannot be localised on external grounds, and so have been placed on the map *on the basis of their linguistic features*: thus the method is circular and the very neat geospatial distribution we see may be an artefact. Secondly, the atlas covers a hundred-year period (1350-1450), but here there is no attempt to represent text dates. Thus we cannot tell whether non-categorical regions reflect synchronic variation or diachronic change.

Figure 3: Distribution of h- forms of the 3.pl. pronoun reproduced from LALME (Benskin et al. 2013)

An example of a more sophisticated approach to visualisation in historical dialectology is found in Versloot (2008)'s study of the history of Frisian. Versloot calculates trend surfaces for dialect features quantified from historical texts. This has the advantage of being accountable and specific, but has the disadvantage of being parametric (imposing an arbitrary limit on the complexity of the distributions that can be mapped) and synchronic (involving no representation of time, except by binning the data into periods and calculating separate surfaces for each period). Nevertheless, this approach is a great improvement on visualisation of raw data, and shares many commonalities with that advocated here including some of the specific kernel functions. An example is reproduced as Figure 4.

Figure 4: Distribution of vowels in Old Frisian 'setta', 'sella' and 'fenne', reproduced from Versloot (2008: 195)

Another recent and sophisticated approach is found in Willis's (2017) investigation of the diffusion of innovative second person pronoun *chdi* in Welsh. Willis uses geographically weighted regression (Fotheringham, Brunsdon and Charlton 2003) with synchronic, apparent-time data to estimate the year at which the rate of use of the innovative form passes 50% at each locality; an example is given in the context of the preposition *efo* 'with' in Figure 5. There is evidently a lot to be said for this approach, which offers clear visualisations from which the spatial dynamics of change can be read. Nevertheless, it is in fact a solution to a slightly different problem than that considered here: how to infer diachrony from synchronic, apparent-time data, rather than how to visualise the distribution of diachronic data. The type of model used, geographically weighted regression, is known to be particularly susceptible to issues with multicollinearity (Wheeler and Tiefelsdorf 2005), which might create problems for any case more complex than that examined by Willis.

Figure 5: Year at which innovative pronoun chdi passes 50% frequency in the context of the preposition efo in spoken Welsh by locality on the basis of GWR, taken from Willis (2017)

More numerous than any of these approaches are works which make no attempt to map the data at all, but merely describe impressions of geographical distributions in prose (potentially alongside non-geospatial visualisations of change over time). Examples which deal with the variable examined in this paper are found in Pettersen (1975; 1991). Such work suffers from a lack of accountability and accessibility in addition to being inevitably imprecise. Impressionistic accounts of distributions are hard to falsify, and, what is more, readers need a good knowledge of the geography of the regions in question in order to understand the patterns described.

1.3 About the data

The corpus used in this study is the *DN online*, a collection of charters from or relating to medieval Norway which have been localised for dialectological research (Blaxter 2017a). Restricting our view to original texts in a Nordic language, we have 10,791 texts containing 2,760,072 tokens. Almost all of these texts are specifically dated and localised. For more detail on corpus, localisation and other tagging, see Blaxter (2017b: 61–100). All first person plural and dual nominative pronouns in the corpus were identified; this involved checking each potential form to disambiguate homophones¹. Table 1 summarises the resulting dataset sorted into broad types. Each of these pronouns was then tagged for referent number on the basis of context; for our purposes, we can separate pronouns into those used to refer to two people (dual contexts) and those used for more

¹ The forms *me*, *mi*, *mid* and *mit* were disambiguated from forms of the preposition *með* 'with'; the form *vid* was disambiguated from the preposition *við(r)* 'at'; the forms *vir* and *vér* were disambiguated from forms of *vera* 'to be', *verja* 'to defend', *verr* 'worse' and *værr* 'comfortable'; the forms *mér* and *mir* were disambiguated from the 1.dat.sg. pronoun *mér* and from *meir* 'more'; the form *mit* was disambiguated from the nt.nom-acc.sg. of the first person possessive adjective; and the form *vi* was disambiguated from the Roman numerals VI, VII, VIII and VIII. Only one known form was excluded: *med* occurs extremely rarely as a form of the 1.pl. pronoun, but is a homograph of the preposition *með* 'with'; with nearly 40,000 occurrences in the corpus, the work of disambiguating these was not deemed worth the tiny number of additional tokens it would produce.

than two people (plural contexts). Pronouns used to refer to one person (generally uses of the ‘royal plural’) were excluded².

type	dual contexts	plural contexts	forms
<i>me</i>	3	7	me, mee
<i>mér</i>	460	3065	mæær, mæer, mæir, mær, meær, meer, meerr, meir, mer, merh, merr, meyr, mœr
<i>mi</i>	3	2	mi, mih, my
<i>mið</i>	174	87	mid, mið, midh, miid, mjd, myd, mydh, mydt
<i>mir</i>	3	4	mier, mir
<i>mit</i>	2944	405	miit, miith, miitt, mijth, mit, mith, mitt, mitth, mjjt, mjt, myt, myth, mytt, mytth
<i>miz</i>	11	12	midz, mitz
<i>ve</i>	5	3	ue, væ, ve, vee, wæ, we, wee
<i>vér</i>	61	918	uær, ueer, uer, vær, veer, ver, vér, vër, wær, weer, weir, wer, wuer
<i>vet</i>	6	0	uætt, væt, væth, vet, veth, vett, wætt, weet, wet, weth, wett
<i>vi</i>	2398	5970	ui, uii, uij, uy, vi, vii, viii, viij, vij, vj, vy, whij, wi, wii, wij, wj, wjj, wy
<i>við</i>	7	1	vid, við, vidd, vidh, vidt, vidtt, viid, viidh, viidt, vyd
<i>vir</i>	1	5	vier, vijr, vir
<i>vit</i>	146	0	uit, vit, vith, vitt, wit, witt

Table 1: Types, forms and token counts for first person nominative plural and dual pronouns in the *Diplomatarium Norvigicum*

This dataset features all of the challenges discussed in Section 1.2: we have intratextual variation, with variable text lengths and uneven distribution of texts in space and time; both localisation and dating are likely to be noisy. Figure 6 shows a small subset of the data to illustrate this: pronouns in dual contexts in texts from inland Telemark (constituting 188 tokens in 88 texts out of the 6172 tokens in dual contexts in 2420 texts). Point colour indicates the type; charter dates are given in text next to each set of points. It might be possible to draw some conclusions about distributions from raw data like these: *vi* is more common in later texts; *mér* is mostly found in the late fifteenth century; *vi* may be more common earlier further east. However, making stronger or more specific statements is difficult, even for this very small subset.

Figure 6: Pronouns in dual contexts by text in inner Telemark

² There is an interesting story to be told about the pronouns used in this function, but since they are only found in documents produced by a tiny number of individuals (primary kings, queens, bishops, abbots and abbesses) in a very limited number of localities, there is little of interest to historical *dialectology* in this story.

2. Methodology

2.1 Kernel density estimation

Kernel density estimation (KDE, also called kernel smoothing or the Parzen-Rosenblatt window method; Rosenblatt 1956; Parzen 1962; for an overview see Gramacki 2018) is a non-parametric method of estimating a variable at a point in n -dimensional space from a sample. It operates by taking a weighted average of all of the sample data in which data are weighted according to their distance from the point of interest: the further away the data, the less relevant they are for making an estimation about a given point, and so the lower they are weighted. By calculating these estimations at many points in space, we can produce an estimation of the overall distribution of the variable. In synchronic dialectology, our data is distributed in two-dimensional space; in the case of historical dialectology, the data is distributed three-dimensions: two-dimensional space and time. The function which derives a weight from the distance is called the *kernel function*, and a great many different such functions can be used.

So for bandwidth b with distance function D and kernel function K for a set of points x with a linguistic variable with set of variants y , the kernel density estimation at point x_j for a given variant y_v is:

$$\hat{f}(x_j, y_v) = \frac{\sum_{i=1}^n y_v K(bD(x_j, x_i))}{\sum_{z=1}^w \sum_{i=1}^n y_z K(bD(x_j, x_i))}$$

That is, for every sample data point, we calculate its distance from the point of interest. This distance and the bandwidth are then used with the kernel function to calculate a weight for each data point. We then sum these weights for all the data points of a specific variant, and divide this by the sum of the weights for all data points for *all* variants. This gives us an estimation of the rate of use of this variant as a proportion of all variants at the point of interest.

Conceptually, using KDE for historical dialectology relies on certain assumptions:

- that language use at a nearby locality is likely to be more similar than language use at a distant locality (*spatial autocorrelation*)³;
- that language use at a nearby point in time is likely to be more similar than language use at a more distant point in time (*temporal autocorrelation*)⁴;

3 This is one of the foundational findings of dialectometry (Séguy 1973; Nerbonne 2010) but also follows directly from a historical linguistic or sociolinguistic understanding of how language change operates geographically: whether a change spreads by wave/contagion diffusion (Schmidt 1872) or by gravity/hierarchy diffusion (Trudgill 1974), at any given point in time it is more likely to have reached a location near to its point of origin than a distant one, thus each successive change is most likely to create a difference between distant dialects.

4 This, too, follows straightforwardly from our understanding of how language change operates: each generation acquires language from the speech of its elders and this acquisition process is largely successful in the sense that the acquired grammars are the same as the targets in most respects. Change over time is thus gradual, with major change occurring only through the accrual of small changes over time. The notion of ‘catastrophic change’, in which minor, surface changes build up over time until a threshold is reached at which point they trigger a major change in underlying structure very abruptly (e.g. Lightfoot 1991; Lightfoot 1979), might be raised as a counterexample, but is irrelevant for two reasons. Firstly, the sense in which such change is discontinuous is purely a product of the theoretical analysis applied and is not a property of the linguistic data *per se*. Secondly, even a change that was perfectly abrupt—that is, occurred in a single generation of acquisition—would still be gradual from the point of view of the *population*, where older speakers without the change would remain even after the arrival of young speakers into the speech community, and in historical dialectology it is the language of the population which is our object of study.

- accordingly, that sample data from nearby localities and points in time are *more relevant* for establishing language use at a given place and time of interest than more distant sample data, and so should be weighted higher.

These assumptions are fundamental to the method and if they are violated, the method may generate misleading results. Giving higher weights to nearby data in space and time will tend to *generate* spatial and temporal autocorrelation (i.e. smooth curves) in the estimation, even if in reality there are sharp dividing lines in the data. Nevertheless, this should not be a great concern since these are extremely well-established assumptions in the field.

2.2 Kernel functions

A central question is what kernel function to use: we know that more distant samples should be weighted lower and nearby samples higher, but there are many different ways to achieve this. This can be decomposed into three questions:

1. Should the kernel be *static* (meaning that there is a consistent relationship between distance and weight) or *adaptive* (meaning that the relationship adapts to the local density of data)?
2. What is the shape of the function (that is, what is the shape of the curve by which weight decreases as distance increases)?
3. What is the *bandwidth* of the function (that is, on what scale does the decrease in weight take place)?

The simplest static kernel function is a *uniform kernel*, in which data within distance b (bandwidth) of the point of interest are weighted 1 and data outside distance b are weighted 0. This type of KDE could be described as simply taking an average of the points within distance b . This is visualised as the top left panel of Figure 7 for $b=1$: nearby samples (on the left) are weighted 1 until the distance b is reached after which all samples are weighted 0.

Other kernel functions allow weights to decrease gradually as distance increases. A *triangular kernel* means that the weight decreases on a straight line such that data at the point of interest are weighted 1 and data at distance b or beyond are weighted 0; this is visualised as the top right panel of Figure 7.

Alternatively, we can have weight decreasing on a curve as distance increases. For synchronic dialectology, Rumpf et al. (2009) suggest using a static Gaussian kernel (visualised as the bottom panel of Figure 7), meaning that data further from the point of interest are weighted progressively lower on a Gaussian curve. A Gaussian curve is the familiar ‘bell curve’, also called the Normal distribution. This type of kernel function is treated as the default in many fields of application. Conceptually, it is not dissimilar to a uniform kernel. Because weight decreases gradually for nearby points, then decreases rapidly before slowing again, data are roughly placed into two groups: nearby data which largely determine the estimate and distant data which have relatively little influence. Thus, the result will often be close to that of a uniform kernel. However, it has two advantages over a uniform kernel. Firstly, weight *does* fall off gradually within both groups, operationalising the intuition that the very nearest samples are marginally more relevant than those marginally further away. Secondly, because no distance would give a weight of 0, there is no possibility that all data will be weighted 0 and so there are no points at which the KDE cannot be calculated.⁵

5 A different function that might seem like a viable option in a linguistics context is simply $f(d) = \left(\frac{1}{d}\right)$, the curve found in Zipf law distributions. This would have the same advantages that a Gaussian kernel has: weight falls off

Figure 7: Weights from different static kernel functions

However, all of these static kernel functions rely on relatively evenly distributed data. To see that this is the case, consider the toy dataset visualised in Figure 8 and Figure 9; this is a relatively typical spatial distribution for medieval Norwegian data. Regardless of the shape of the kernel function selected, the problem comes when we try to set the bandwidth (b). We want to choose a bandwidth which is high enough that we are including a meaningful amount of data in each weighted average: with a very small bandwidth, our estimations will not be meaningfully smoothed and we will produce a noisy estimation that is very close to the raw data. However, if we set the bandwidth too high, we will smooth over potentially meaningful variation by including too much of the sample in every average. Given very unevenly distributed data, it is impossible to select a bandwidth that achieves both of these: if we select a bandwidth that suits densely sampled regions (as in Figure 8), we will leave data in sparsely distributed regions unsmoothed; if we select a bandwidth that suits the sparsely sampled regions (as in Figure 9), we will smooth over potentially meaningful variation in densely sampled regions.

Figure 8: Toy dataset with bandwidth 50km

Figure 9: Toy dataset with bandwidth 300km

The solution to this problem is to use an adaptive kernel. The simplest form of adaptive kernel, an *adaptive uniform kernel*, is simply to take an average of the b nearest neighbours: our function thus weights each of the b nearest neighbours 1 and all other sample data 0. This has been carried out for the toy dataset with $b=20$ in Figure 10. As with static kernels, more sophisticated options are possible; however, this simple option will suffice for our purposes.

Figure 10: KDE of toy dataset with adaptive uniform kernel with $b=20$

2.3 Incorporating time

In principle, we could simply treat the data of historical dialectology as distributed in a three-dimensional space and apply a single kernel function. This, however, would create two problems. Firstly, since we have different units in our different dimensions, it is not clear how we would scale them relative to one another: how many kilometres is equivalent to a year? Secondly, we would be forced to use the same shape of kernel function in both space and time, even though both the distributions of our sample in space and time and our assumptions about likely underlying distributions of the variable in space and time might be different. For example, we might have a sample that is relatively evenly distributed in time (at least for a central period of interest) but

gradually and nothing is weighted 0. However, because this type of curve has no ‘levelling off’ as the distance approaches 0, it would overweight nearby data and therefore give very noisy estimations wherever at least one sample fell very close to the point of interest. In the extreme case, if there was a sample actually *at* the point of interest, it would be weighted ∞ and so the estimate would simply reproduce this sample and ignore all others.

highly unevenly distributed in space. We also might have relatively specific assumptions about the shape of language change curves in time (viz., that change progresses by an *s*-curve) that do not necessarily hold in space.

For these reasons, we can apply two separate kernel functions. We can first apply a static kernel function in time, which will give us a weight for each sample datum on the basis of its distance in time from the date of interest. The obvious choice of temporal kernel is a Gaussian kernel, since all other things being equal this will tend give us an *s*-curve. We can then apply the uniform adaptive spatial kernel to these weights: instead of taking an average of the *b* nearest neighbours, we take an average of nearest neighbours with a sum weight *b*. This means that our adaptive window will expand relative not just to the local density of sample data in space, but the local density of sample data in space and time.

2.4 Incorporating uncertainty

It is sometimes possible to quantify some of the uncertainty in the localisation of historical texts. We may have texts for whose origin we have posited multiple possible points that we cannot reconcile on the basis of language-external evidence. In other cases, we may have texts which we can confidently localise to a large area, but cannot pin down to a specific locality within this area. Both of these can be seen as instances of the same type of uncertainty: localisation to an area can be understood as localisation to any one of the settlements (points) within that area.

Given that we are already taking weighted averages of our data, we can now easily incorporate this spatial uncertainty. If a text could potentially be localised to multiple points, it can be localised to all of these points; to avoid double-counting, we then divide the weight of the text at each of these localisations by the total number of localisations. An example of this is given in Figure 11 and Figure 12 for one text from the corpus, DN VI.624 (7th of October, 1498). This text:

- was published from a farm called Tveito (*thetta breff som giort war a Thwetom*, ‘this letter which was made at Tveito’);
- and was produced by a signatory named *Mattis Olaffson* who was the priest in a parish called Holla (*prest a Holla*)
- and provost in Numedal and Tinn (*prost offwer Nwmedal och Tin*).

This gives us four possible ‘true’ localisations (four guesses about where the person who shaped the language of the charter was from). We can thus localise the text to each of these places and divide the weight by 4 at each one (Figure 11). However, three of these are in fact not points but regions: Holla is a parish, Tinn a *herad* (‘hundred’) and the valley of Numedal a relatively large district. We can thus localise the text to the n_j known medieval farms in each of these regions and divide its weight at each by $4n_j$ (Figure 12).

Figure 11: Uncertain localisation for DN VI.624 (1)

Figure 12: Uncertain localisation for DN VI.624 (2)

2.5 Bandwidth setting

Optimal bandwidths were determined as follows. A series of simulated datasets were created with the same distribution of texts in space and time as the real data and a linguistic variable which was a

noisy function of time and distance from Oslo (i.e. datasets simulating a change diffusing from Oslo). KDEs were then calculated for each of these with a range of bandwidths and evaluated on their ability to place points on the correct side of the isogloss: points in the zone which the change had reached which got a KDE value of greater than or equal to 0.5 were considered correctly placed; points outside the zone which the change had reached which got a KDE value of less than 0.5 were considered correctly placed; other points were considered incorrectly placed. The optimal bandwidths are then those which correctly categorise the largest proportion of the data. This modelling procedure was done twice for each of the two datasets (dual contexts and plural contexts), once for a change starting in 1250 and expanding by 5.4km per year (thus taking around a century to reach the whole area) and once for a change starting in 1400 and expanding by 3.6km per year (thus taking around 150 years to reach the whole area); these timings and speeds were determined on the basis of the chronological overview of the data (Section 3.1, below). The optima are given in Table 2: the temporal bandwidth is expressed as the width of the Gaussian density function in years; the spatial bandwidth is a proportion of the whole weighted dataset.

		spatial bandwidth	temporal bandwidth
earlier change	dual contexts	6.15%	10.8 years
	plural contexts	6.18%	7.0 years
later change	dual contexts	4.11%	14.4 years
	plural contexts	4.21%	13.1 years

Table 2: Optimum bandwidths

3. Results

3.1 Chronology

We can get an initial view of the overall chronology of changes among forms by using temporal KDE alone. Figure 13 shows a temporal KDE for the different pronouns in dual contexts and Figure 14 for plural contexts; the kernel is a Gaussian kernel with bandwidth (standard deviation) of 10 years. We see a broadly similar picture for both contexts, with the exact forms mirrored. The Old Norwegian form (plural *vér* and dual *vit*) is close to categorical at the very beginning of the period covered by the corpus. The reanalysed forms (plural *mér* and dual *mit*) rapidly replace these at the beginning of the thirteenth century, around ten years earlier for the dual than the plural. The East Nordic form *vi* rises in frequency from the end of the fourteen century, becoming the majority form by the turn of the sixteenth century; this change is rather earlier and quicker in the plural. In addition to these broad trends, we see a rise of West Nordic plural forms (primarily *mér*) in dual contexts throughout the fifteenth century and of dual forms (primarily *mit* and *mið*) in plural contexts from the beginning of the fifteenth century through to the early sixteenth. Although this last trend attains a similar magnitude in both contexts (around 20% of all usage), a noteworthy difference between dual and plural contexts is that in plural contexts West Nordic dual forms (*mit*, *mið*) actually outstrip plural forms (*mér*, *vér*); the reverse does not happen in dual contexts.

Figure 13: Temporal KDE of pronoun types in dual contexts

Figure 14: Temporal KDE of pronoun types in plural contexts

3.2 Spatial distributions

Figure 15 and Figure 16 show individual years from the KDEs for the changes in dual contexts; these are frames taken from Video supplement 1 and Video supplement 2. In Figure 15 we can see that *mit* seems to replace *vit* in a counterhierarchical pattern, with Oslo distinguished from the surrounding countryside by its conservatism; this might, however, be an artefact of the spatial distribution of the data (almost all data from before 1300 is localised to the cities of Oslo, Bergen, Stavanger and Trondheim, meaning that their language use might appear to be more conservative than the surrounding countryside where all data is later). What we can say confidently, however, is that the change towards *mit* starts in the inland east, and that conservative *vit* is maintained longest in southern West Norway (Hordaland, Rogaland).

Figure 16 for the later change from *mit* to *vi* shows an unmistakable hierarchical diffusion pattern. The change begins in Jämtland, suggesting diffusion from Swedish; it then spreads into Trondheim from which it jumps to Bergen. At a still later stage, we can see the cities of Bergen, Hamar and Oslo standing out as innovative compared with the surrounding countryside. Finally, conservative *mit* is retained longest in the mountainous inland south: inner Telemark, Setesdalen, and inner Rogaland. Another interesting pattern we can see in this figure concerns the use of the historically plural form *mér* in dual contexts: this type, which we saw from Figure 13 reached 20% of usage around 1450, is largely restricted to a contiguous region of inland Buskerud and northern Telemark, with a particular hot-spot in an area corresponding to the modern-day Kongsberg municipality. This area is given in more detail (and with municipality names marked) in Figure 17.

Figure 18 and Figure 19 show years from the KDEs for the changes in plural contexts; these are individual frames from Video supplement 3 and Video supplement 4. In Figure 18 for the change from *vér* to *mér* we see a similar counterhierarchical pattern to that seen in Figure 15. Here, both because the dataset is a little larger and because the change is a little later (and so does not overlap so completely with the sparse early period where almost all data is from the cities), we can be more confident about this counterhierarchical pattern, particularly in the east. An examination of the individual examples suggests that *mér* arises first in Vestlandet (the first examples are all from Hordaland and Rogaland⁶); there are then examples in mountainous inner Norway⁷. Thus we seem to see an innovation which was relatively widespread in Fjell-Norge and rural Vestlandet at an early

6 The examples are:

- *Mér villium ollum monnum kunnikt gera at...* “We wish to make known to all men that...” and *þa gafom mer vart bref* “then we gave our charter” in DN IV.6, 26th of May 1293, from farms in Ullensvang;
- *Mer gærom ydr kunikt at...* “We make known to you that...” in DN I.98, around 1303, for which the only sure localisation is the farm Finne in Voss;
- *sættu mer vor jnsigli firir þætta bref* “we set our seals on this charter” in DN II.81, 20th of February 1306, from farms in Kvam and Ullensvang;
- *mætte mer þo muna þau rangynði sem oss varo gor j fyrri summer* “we can still remember the wrongs which were done to us last summer” in DN IV.72, 3rd of June 1307, from Stavanger and addressed to Klepp parish;
- *voro mer j hia j skruða huseno berfættobræðra j Berghwin* “we were present in the sacristy of the barefoot brethren in Bergen” in DN IV.82, 28th of November 1309, from Bergen, Stavanger, and other parts of Vestlandet.

7 The first is *sua warom mer ok j hia* “thus we were present” and *settum mer firir þætta bref war jnziglli* “we set our seals on this charter” in DN V.54, around 1310, from Gol in Hallingdal.

point, was slow to spread into the cities, and slowest of all to truly go to completion in the western cities of Bergen and Stavanger.

In Figure 19 we clearly see two competing changes. Plural *mér* is replaced by historically dual *mit* in parts of the south: this becomes a frequent pattern first in Rogaland before rising in frequency in Telemark and coastal Agder (presumably diffusing around the Agder coast); inland Setesdalen is a relatively late adopter, but is then the place where this usage is retained longest. At the same time, we see East Nordic *vi* replacing *mér*. Unsurprisingly, this is found first in Jämtland; here, it is difficult to know whether we are speaking of language change *per se* or of the shift from a Norwegian to a Swedish written standard, as Jämtland's political affiliation switches from west to east. Within Norway proper, the change arises first in Bergen, then Trondheim, Hamar, Oslo then Tønsberg; from each of these cities it proceeds to spread outwards into the surrounding countryside. This change seems to out-compete the replacement of *mér* by *mit*, so that later we see once *mit*-using areas such as Rogaland switching to *vi*.

Figure 15: KDEs for vit > mit in dual contexts, 1310-1355

Figure 16: KDEs for mit > vi in dual contexts, 1425-1500

Figure 17: KDEs for mit > vi and mit > mér in dual contexts for 1450, focusing on Telemark, Buskerud and Vestfold

Figure 18: KDEs for vér > mér in plural contexts, 1310-1355

Figure 19: KDEs for mér > vi and mér > mit in plural contexts, 1410-1470

4. Discussion

4.1 First person non-singular pronouns in the history of Norwegian

These visualisations of the diffusion of our six changes (in dual contexts: *vit* > *mit*, *mit* > *mér*, *mit* > *vi*; in plural contexts: *vér* > *mér*, *mér* > *mit*, *mér* > *vi*) enable us to answer a number of questions. As a reminder, the research questions we set up in Section 1.1 were as follows:

1. precisely when did *mér* and *mit* arise, when did *vér* and *vit* disappear, when did *vi* diffuse into Norwegian, and when was the dual-plural distinction lost?
2. where were *mér* and *mit* innovated and by what pathway did they diffuse?
3. what was the route by which *vi* spread into Norwegian? did it spread from Swedish or Danish?
4. do *me* and *mi* reflect *mér* and *mit*, meaning that different dialects selected the plural (*mér*) and dual (*mit*) as the merged form?

The first of these, on the specific timings, could largely have been answered with the chronological visualisations in Figure 13 and Figure 14 alone. The two reanalysed forms, *mér* and *mit*, probably arose in the mid-to-late thirteenth century; certainly they are already present as soon the volume of

data increases at the end of the thirteenth century. They both rise in frequency rapidly in the first half of the fourteenth century in characteristic *s*-curves. This change happens a little earlier in the dual than the plural, with conservative *vit* disappearing entirely shortly after 1350. The change from *vér* to *mér* is about 20 years later, and then stabilises with the conservative form at around 5% of usage rather than disappearing entirely (perhaps because it also exists as a royal/formal plural until a much later date). The East Nordic form *vi* starts to appear in Norwegian sources in the latter half of the fourteenth century; its period of rapid diffusion, however, is in the fifteenth. The dual-plural distinction is not totally lost everywhere by the latest period covered by the corpus: even in 1550, *mit* is marginally more common in dual contexts than it is in plural contexts.

For the other three questions, the spatial visualisations (Figure 15, Figure 16, Figure 18 and Figure 19) are needed. Both *mér* and *mit* seem to have arisen in inland Norway. In the case of *mér*, it seems clear that this was in Vestlandet and that it spread through Fjell-Norge before reaching the coasts and the cities; in the case of *mit* our evidence is less good, since the change is earlier. Both show a clear counterhierarchical pattern, particularly in the east, with Oslo standing out as a conservative island in the first couple of decades of the fourteenth century.

Our third research question concerns the East Nordic form *vi*. This diffuses by an unmistakeable hierarchical pattern in both dual and plural contexts: Bergen, Trondheim, Hamar, Oslo and Tønsberg are all leading areas compared with surrounding rural areas. The central role of Bergen in this change strongly suggests that the relevant contact is at least partly with Danish; if diffusion of *vi* was solely from Swedish, then surely only the eastern cities on the Swedish border would stand out.

An interesting problem is raised when we compare the diffusion of *vi* in the Middle Norwegian data (Figure 16 and Figure 19) to its distribution in Modern Norwegian (Figure 1): *vi* is common across a much larger region in late Middle Norwegian than in Modern Norwegian. Indeed, the change towards *vi* basically goes to completion everywhere by 1550, which does not correspond at all to the Modern Norwegian situation, where *vi* is an eastern, northern and urban variant only. The best explanation for this is that some proportion of the changes we see in the Middle Norwegian written sources are purely written phenomena: that what we are seeing here is not evidence of ongoing language change in fifteenth and sixteenth century spoken Norwegian, but the shift to Danish as the written medium. We might expect this shift to have operated hierarchically, since the cities were the centres of the Dano-Norwegian administrative state, of the church, and were the locations where the mobile aristocracy and merchant classes would have lived in the largest numbers. Nevertheless, it still seems likely that at least some of the change towards *vi* we see in the historical data does reflect spoken language change. Some of the leading areas in Middle Norwegian (Jämtland, Trondheim, Bergen, Oslo, north Norway, the rural south-east) are indeed the areas where *vi* is used today; Occam's Razor suggests we should assume that this written usage reflected spoken usage, rather than imagining two independent changes at totally different times, one in writing and one in speech.

Turning to our fourth question, the evidence of these data strongly support the supposition that Modern Norwegian *me* is the result of a merger of *mér* and *mit* into *mér* and *mi* the result of a merger of *mér* and *mit* into *mit*. The region in which *mit* is used in the plural in late Middle Norwegian corresponds very well to the area in which *mi* is used in Modern Norwegian. This correspondence is even closer if we look a little deeper: object forms which look like historical duals (*åkkō*, *åkkon*, *kånn*) are found in a slightly larger area in Modern Norwegian than subject

forms, and if in Middle Norwegian we look at *mér* and *mit* alone, ignoring the role of *vi*, we find that *mit* is the majority form in a very similar region (Figure 20). The merger into *mér* which must have given rise to *me* is harder to trace due to the interfering effects of the switch to Danish *vi*. Since the switch to written *vi* happens earlier in dual contexts, we cannot tell whether *mér* might have been diffusing in such contexts in the latter half of the fifteenth and first half of the sixteenth century but is obscured by *vi*, or whether it did not really spread until some time later. However, we can identify that this process had begun by the mid-fifteenth century in an area centred on Lower Buskerud (Figure 17), precisely corresponding to the *me* side of the *me:mi* border in Modern Norwegian.

Figure 20: Evidence for dual-descended forms in Modern Norwegian vs. pronouns in plural contexts in Middle Norwegian in 1500 excluding vi

4.2 *Forms of diffusion*

It was noted in Section 1 that the three proposed forms of linguistic diffusion (the wave model/contagion diffusion, the gravity model/hierarchical diffusion and counterhierarchical diffusion) are largely known by inference from synchronic data. Data of this sort clearly have the potential to confirm the existence of such patterns in a much more direct way, and indeed the data explored in this paper seem to show evidence for all three types. We see evidence for pure wave diffusion in the spread of *mit* in plural contexts, a change which arises first in the south-west before spreading around the coast and inland; crucially, this spread is continuous, with urban centres playing no particular role as resistors or leaders. We see evidence for hierarchical diffusion in the spread of *vi* in both contexts: cities are always more innovative than their surroundings, and the spread is discontinuous, jumping between urban areas and bypassing intervening countryside. Finally, we see evidence for counterhierarchical diffusion in the spread of the reanalysed forms *mér* and *mit*, which seem to arise in inland areas and spread throughout rural southern Norway fastest, with the cities, particularly Oslo in the east, being slow adopters.

4.3 *Kernel density estimation for historical dialectology*

The datasets of historical dialectology are typically very noisy and the stories of spatial diffusion we want to tell are complex, creating a problem in getting from raw data to justified narrative; this problem is only likely to become more acute in future, as our digitised data improve. This paper has presented one tool for mapping spatial diffusion patterns over time from such datasets, kernel density estimation (KDE). This tool is extremely well-studied and established in fields such as signal processing and pattern recognition; the contribution of this paper has been to propose the specific variant best suited to the problem of drawing shifting linguistic regions in historical dialectology, show how the parameters of the tool can then be tuned, and demonstrate its use with an example dataset from the history of Norwegian. It has clearly been demonstrated that the method allows us to tell clear and specific stories about diffusion processes, identifying points of origin and distinguishing types of diffusion, that would be nigh impossible to arrive at by impressionistic, qualitative methods.

The method is not infallible. Researchers using KDE on datasets like the ones demonstrated here should remain conscious of the number of data points behind their visualisations: the particular year an isogloss reaches somewhere or the exact course of a boundary relative to geographical features may be dependent on the placement of single data points in a sparsely distributed dataset, and no additional weight should be put on such findings just because they can be visualised in an appealing way. We should also be wary of artefacts of the method: any proximity-based smoothing method will have a tendency to draw boundaries through areas of low sample density, grouping areas of high sample density together. Thus, especially if a dataset is small, the positions of coastal features and other geographical barriers may warp the placement of isoglosses.

These warnings notwithstanding, temporal-spatial KDE of the type described here represents a major improvement on existing approaches to mapping distributions in historical dialectology.

5. References

- Bailey, Guy, Tom Wikle, Jan Tillery and Lori Sand. 1993. Some patterns of linguistic diffusion. *Language Variation and Change* 5. 359–390.
- Benskin, Michael, Vasilis Karaiskos, Keith Williamson and Margaret Laing. 2013. *An Electronic Version of A Linguistic Atlas of Late Medieval English*. Edinburgh: University of Edinburgh. <http://www.lel.ed.ac.uk/ihd/elalme/elalme.html>.
- Blaxter, Tam T. 2017a. *Annotated DN online*. Cambridge. http://www.icge.co.uk/dn_online/.
- Blaxter, Tam T. 2017b. *Speech in space and time: Contact, change and diffusion in medieval Norway*. Cambridge: University of Cambridge Ph.D. <https://doi.org/10.17863/CAM.15576>.
- Burridge, James. 2017. Spatial Evolution of Human Dialects. *Physical Review X* 7(3). doi:10.1103/PhysRevX.7.031008. <http://link.aps.org/doi/10.1103/PhysRevX.7.031008> (20 January, 2018).
- Burridge, James. 2018. Unifying models of dialect spread and extinction using surface tension dynamics. *Royal Society Open Science* 5(1). 171446. doi:10.1098/rsos.171446.
- Fotheringham, Stewart, Chris Brunsdon and Martin Charlton. 2003. *Geographically weighted regression*. John Wiley & Sons, Limited.
- Gramacki, Artur. 2018. *Nonparametric Kernel Density Estimation and Its Computational Aspects* (Studies in Big Data 37). Cham, Switzerland: Springer.
- Haugen, Einar. 1976. *The Scandinavian Languages: An Introduction to their History* (The Great Languages). London: Faber and Faber.
- Indrebø, Gustav. 1951. *Norsk målsoga*. Bergen: J. Griegs boktr.
- Jahr, Ernst Håkon (ed.). 1990. *Den Store dialektboka*. 1. oppl. Oslo: Novus.
- Johannessen, Janne Bondi, Joel Priestly, Kristin Hagen, Tor Anders Åfarli and Øystein Alexander Vangsnes. 2009. *The Nordic Dialect Corpus*. <http://tekstlab.uio.no/nota/scandiasyn/> (23 March, 2018).
- Laing, Margaret. 2013. *A Linguistic Atlas of Early Middle English, 1150–1325*. Edinburgh: University of Edinburgh. <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>.
- Lightfoot, David. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
- Lightfoot, David. 1991. *How to set parameters: Arguments from language change*. Cambridge, Mass.: MIT Press.
- Nerbonne, John. 2010. Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B* 365(1559). 3821–3828. doi:10.1098/rstb.2010.0048.
- Noreen, Adolf. 1970. *Altisländische und altnorwegische Grammatik: (Laut- und Flexionslehre) ; unter Berücksichtigung des Urnordischen* (Altnordische Grammatik von Adolf Noreen; 1). 5., unveränd. Aufl. Halle: Niemeyer.
- Parzen, Emanuel. 1962. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics* 33(3). 1065–1076. doi:10.1214/aoms/1177704472.

- Pettersen, Egil. 1975. *Språkbrytning i Vest-Norge 1450-1550*. Bergen.
- Pettersen, Egil. 1991. *Språkbrytning i Vest-Norge 1450-1550 II*. Bergen: Alma Mater.
- Rosenblatt, Murray. 1956. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* 27(3). 832–837. doi:10.1214/aoms/1177728190.
- Rumpf, Jonas, Simon Pickl, Stephan Elspass and Werner König. 2009. Structural analysis of dialect maps using methods from spatial statistics. *Zeitschrift für Dialektologie und Linguistik* 76(3). 280–308.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- Séguy, J. 1973. La relation entre la distance spatiale et la distance lexicale. *Rev. Linguist. Romane* 35. 335–357.
- Seip, Didrik Arup. 1955. *Norsk språkhistorie til omkring 1370*. 2nd edn. Oslo: H. Aschehoug & Co.
- Studer-Joho, Nicole. 2014. *Diffusion and change in early middle English: methodological and theoretical implications from the LAEME corpus of tagged texts* (Schweizer Anglistische Arbeiten 141). Tübingen: Francke.
- Trudgill, Peter. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3(2). 215–246. doi:10.1017/S0047404500004358.
- Versloot, Arjen P. 2008. *Mechanisms of language change: vowel reduction in 15th century West Frisian* (LOT International Series 195). Utrecht: Landelijke Onderzoekschool Taalwetenschap (LOT).
- Wheeler, David and Michael Tiefelsdorf. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7(2). 161–187. doi:10.1007/s10109-005-0155-6.
- Wikle, Thomas and Guy Bailey. 1997. The spatial diffusion of linguistic features in Oklahoma. *Proceedings of the Oklahoma Academy of Science* 77. 1–15.
- Willis, David. 2017. Investigating geospatial models of the diffusion of morphosyntactic innovations: The Welsh strong second-person singular pronoun *chdi*. *Journal of Linguistic Geography* 5(1).