

IN SEARCH OF A CLIMATE CHANGE SIGNAL IN NOVA SCOTIA: THE ALEXANDER MACKAY DATA, 1901-1923

MADISON CULBERTSON-PAOLI¹, LAURA FARRO^{1,2},
ANDREW LONG^{1,*}, AND STEVEN WILKINSON¹

¹*Department of Mathematics and Statistics, Northern Kentucky University,
Highland Heights, Kentucky, USA 41099*

²*Department of Physics, Northern Kentucky University,
Highland Heights, Kentucky, USA 41099*

ABSTRACT

Our objective in this work is to model First Appearance Time (FAT) of flowering in five species of plants in Nova Scotia, Canada, as a function of climatic variables (such as temperatures) and geographical factors (such as latitude). Dr. Alexander H. MacKay was the superintendent of public schools in Nova Scotia from 1891-1926. Beginning in 1896 MacKay instructed all the school teachers of Nova Scotia to have their students collect data on the first appearances of numerous plants, animals, and seasonal events, and then summarized the data himself. The summaries of the phenological data collected in this massive citizen science project were then published in the *Proceedings of the Nova Scotian Institute of Science* in a consistent fashion over the period 1901-1923. We analyze five species from the summary MacKay data for Nova Scotia, producing a model for First Appearance Time of flowering for each, as a function of latitude, longitude, mean monthly temperatures for many months, and sea ice off the coast of Newfoundland in winter months. Our model produces good agreement between predicted FATs and those FATs we find in the literature.

Key Words: 1901-1923, Alexander MacKay, First Appearance Time (FAT), flowering, latitude, longitude, mean monthly temperatures, Nova Scotia, sea ice.

INTRODUCTION

“Seasonal timing of biological events, phenology, is one of the strongest bio-indicators of climate change.” (CaraDonna *et al.* 2014) Our purpose is to use phenological data collected and summarized by Dr. Alexander MacKay to model First Arrival Time (FAT) of flowering of several species of plants in the early 1900s in Nova

* Author to whom correspondence should be addressed: longa@nku.edu

Scotia. Having obtained a model, we use it to predict a change in FAT based on climatic variables (e.g. temperatures, sea ice), suggesting that a consequence of climate change is phenological change, and allowing us to guess how further changes in climate will affect the phenology of these flowering species of plants. We also use our model to compare our predictions to data and models on FATs from the late 20th century and early 21st century. Studies showing signs of changes in the four seasons in response to climatic change are becoming quite common (Garbary 2018; Santor, et al. 2018); this is our contribution.

This paper is certainly an homage to Dr. Alexander MacKay, to whom we are indebted for our data. He was a beloved member of the Nova Scotian Academy of Sciences and a brilliant scientist in his own right. Appointed Superintendent of Education for Nova Scotia on November 4th, 1891, he retired nearly 35 years later, on July 31st, 1926. He left an indelible mark on Nova Scotia's public educational system. "Under his able administration not only the [Pictou] Academy, but the whole province, was coordinated into one efficient system of education, in imitation of Ontario and American models." (Wood 1994) "Under Principal MacKay's strong administration the Academy made rapid strides. It became celebrated throughout the province and far beyond its limits. Students flocked in from all quarters until there was not room enough to receive." (MacPhie 1914) Perhaps MacKay's disposition had something to do with his popularity: "Only once did I hear him speak with unwonted warmth or perhaps anger, and that was when an educated man seemingly tried to demonstrate to a large audience that the world might be flat!" (Piers 1930) We include a photograph of MacKay (Fig 1), a photo favored by Harry Piers, author of MacKay's "scientific obituary" (ibid).

MacKay himself described the history of the phenological work in Canada, and presented the state of affairs in 1897, when the contributions of his students began to appear in the summaries:

In 1890, Section IV of the Royal Society of Canada passed the following resolution: 'That the various Natural History and Scientific Societies affiliated with the Royal Society be requested by it to obtain accurate records in their individual localities of meteorological phenomena, dates of the first appearance of birds, of the leafing and flowering of certain plants, and of any events of scientific interest for collation and publication in the Transactions of the Society.'... By 1897 the idea was extensively

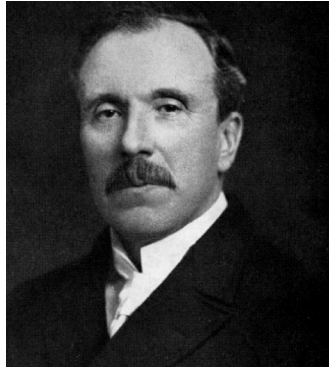


Fig 1 This photograph, chosen by the author of MacKay's "scientific obituary" (Henry Piers), is one personally sent to him with MacKay's handwritten inscription ("Yours very truly").

taken up in the public Schools of the province of Nova Scotia, the pupils of a whole school section or district being the observers, under the direction and criticism of the teacher. The observations were a part of the prescribed 'Nature Studies' in all schools, although the recording of them in the Phenological Records of the locality and the sending of a duplicate of the observations to the Inspector to be transmitted to the Education Office, were entirely voluntary. While the schedule of the Botanical Club had a list of about sixty objects for the observation of their first occurrence each season, the schedule of the public schools had over one hundred objects on its list, with instructions and a column for the observation of (1) the first occurrence and (2) when each began to be common. Over two hundred fairly well filled schedules were sent in from as many localities throughout the province. (MacKay 1902a)

MacKay gave us our marching orders from across the years:

...we may consider a phenological date to be a sort of mathematical function of variables, several of which are already being very systematically and accurately observed and recorded by the meteorological departments of most countries, such as the variations of temperature, of atmospheric pressure, sunshine, precipitation. Then there are local constants, such as latitude, elevation, slope, proximity of bodies of water, and character of the soil. All of these influences affect the phenological date,

and conversely the date may be considered as a summation or integration of all these and other more or less unknown elements. (ibid, p. 76).

Note that MacKay mentions only latitude: we discover here that longitude is nearly as important. Otherwise, his suggestions are spot-on: if we had the information he suggests at our fingertips, we would include it. As it was, we used what was available.

His observations of discrepancies in phenochrons can be considered a second charge to us:

An interesting irregularity in the phenochrons of the different counties is shown in nearly every part of this table. Their order is not parallel in the different counties. Very often it is reversed. As the phenochrons are averages of ten observations, it cannot be laid altogether to the charge of defective observation. The rarity of certain species in certain counties, or in the districts in which the observations were made, tends to make the phenochron later, for the plants may be in flower several days before they may be met with. But the character of the soil, the elevation, the slope, etc., must have had some influence. And, then, may it not be possible that the same species may develop a tendency to an earlier or later maturing in different regions? These are questions that careful future observations may help to answer. (ibid, p. 81).

It is fascinating to hear his mature scientific voice in his concern for using averages to arrive at stable estimates; in his theorizing about the influence of low sampling, and environmental influences; and in proposing that he and his students would be laying the groundwork for future scientists.

MacKay presented his phenological study strategy in his 1901 report on the phenology of Nova Scotia (MacKay 1903):

I present herewith a summary of the phenological observations made in about 450 of the public schools of the Province of Nova Scotia, each county being represented by a greater or less proportion of observers.... The observers are specially directed to the determination of two dates (phenochrons) – one for the first appearance of the event (leafing, flowering, ripening of fruit, etc.), the other for the date when it may be said to be ‘becoming common.’

For the years 1901 to 1923 (with the exception of 1910 for which we were unable to find the data), we found and digitized published

summaries of the observations made by the public school children of Nova Scotia from the *Proceedings of the Nova Scotian Institute of Science*. The printed form the teachers and students used in individual schools was published in 1902 (MacKay 1902b), along with a report (ibid, pp. 58-63) on difficulties associated with compiling those regional reports of the raw data into Nova Scotian summaries in the case of 1901.

As for MacKay's scientific legacy, those who indexed the *Proceedings of the Nova Scotian Institute of Science* report that "[t]he phenological data, collected over 31 years by Dr. A.H. MacKay, are a major contribution to Canadian science..." (Nova Scotian Institute of Science 1992, p. 150). MacKay collected data Canada-wide in the early reports, from a handful of observers across the Commonwealth. The value of his work was recognized in his own time: The President of the Nova Scotian Institute of Science said in his address of March 14, 1898 that MacKay's phenological observations "...may lead to some important generalizations regarding the relation of organized life to latitude and other climatic conditions." (MacKay 1899, p. ii) Note again the emphasis on latitude only – we will find that longitude is perhaps equally important.

Our primary objective here, however, is to prove the President correct in his assessment: MacKay's observations do lead to important generalizations regarding the relation of life to climate.

MATERIALS AND METHODS

Data collection and processing

The data were obtained from summaries of MacKay's phenological data published in the *Proceedings of the Nova Scotian Institute of Science* (MacKay 1903), and every year thereafter until 1923, with the exception of 1910. We analyze five species from the summary data for Nova Scotia from the period 1901-1923, species chosen for special consideration by MacKay himself. After an initial cleaning we imputed missing values (using a method we describe), after which some regions were combined (described below). Once the data were deemed clean, we proceeded to the modeling step.

The original raw data were collected by school teachers of the more than 1400 schools which MacKay oversaw as superintendent of public schools in Nova Scotia. This was one of the earliest and largest

citizen science projects ever created. Not every school participated every year, but MacKay strongly encouraged it, and for more than 25 years MacKay oversaw not just the collection, but also the analysis of the data. His summaries were published every year from 1901 to 1923 (with the exception of 1910); his summaries were based on the summaries of colleagues from each of the 9-11 regions of Nova Scotia (the number of these districts varied over time).

Descriptions of the difficulties encountered in the regional summarization and the process of data collection itself (including the forms the teachers used) were published in the *Journal of Education* (MacKay 1902c). MacKay and his collaborators were very concerned about accuracy, stating:

[C]are must be exercised in selecting schedules, the observations of which appear to have been carefully made, neglecting any which give reason for doubt, when selecting for summation on the form within. Great care must also be exercised in copying the figures and entering them, so that no slip may occur. Every entry must be checked. One slip may spoil the effect of all the accurate numbers entering into the summation. In like manner, great care has to be taken in adding and averaging the figures; and for this purpose every sum should be done twice in reverse order, so as to give absolute confidence in the accuracy of the work. (MacKay 1903, p. 490).

While we had access to the school-level raw data (digitized by Fenech, et al. 2005) from the data publicly archived in ledgers in the Nova Scotia Museum of Natural History, Nova Scotia (Austen 2000), we chose to limit ourselves to the summary data published by MacKay. Clearly there were issues with the raw data which were thoroughly and carefully addressed by MacKay's regional compilers; and while we had access to over 100 variables featured on the summary reports, we chose a small subset to work with. MacKay and his collaborators also carefully chose a subset of the data, that being of good quality, and we know that they made a conscious culling of the data to avoid including data poorly or improperly collected, fabricated, or otherwise suspect:

The various points for consideration in choosing Schedules are a fair distribution of the Stations over the Belt, the number and accuracy of the observations, the sex and temperament of the observer, the neatness of the work, the method of stating dates

and in some cases the Compiler's personal knowledge of the observer. (MacKay 1902c, p. 59).

The compilers of the data from each of the regions often gave informative (and sometimes amusing) examples of the trials and tribulations of data collection and reporting (ibid, pp. 60-63). By using the summary data, carefully scrutinized and cleaned by reliable colleagues of MacKay, we avoid some of the noise associated with the raw data. Nonetheless it is certain that we inherit some of the data problems of citizen science. Some interesting and/or amusing examples from individual compilers follow (many more can be found in that reference):

It would appear in many cases that the observer has given the date of his first seeing a plant instead of the actual date of its first appearances in a locality. Frequently the date 'when becoming common' is given the same as 'when first seen'. In such cases the former is probably the correct one.

- *One or two observations are obviously guesses. I strongly suspect that they were all filled out about the close of the term, possibly from memory or aided by the pupils.*
- *G. is very early with 55, 57, 60, in Musquodoboit Harbor, but as Rev. Mr. Rosborough is there and instructs the teachers often in Botany, I accept them.*
- *Unusually early dates are often given, based upon plants growing in exceptionally favourable situation.*

We ourselves found instances for which the “when common” date was earlier than the “first seen” date. We found other values which appeared to be outliers, perhaps of the sort described above (values occurring early for the wrong reasons – some teacher wanting to be first, or basing the date on a plant growing in exceptional circumstances). We managed to handle these problems with what we felt were reasonable strategies. For example, sometimes an outlying data value appeared suspiciously similar to a value in a column or row nearby (perhaps an error at the time of printing). In that case, we could use the averages (which were also frequently given) to check: if the data didn't give the proper average, we could assume that an error had been made in that suspicious position, and fix it.

Data Subsetting

In his 1901 report (MacKay 1903), MacKay presents a graphic (we'll call it “the 1901 graphic”, Fig 2) of the first appearances of

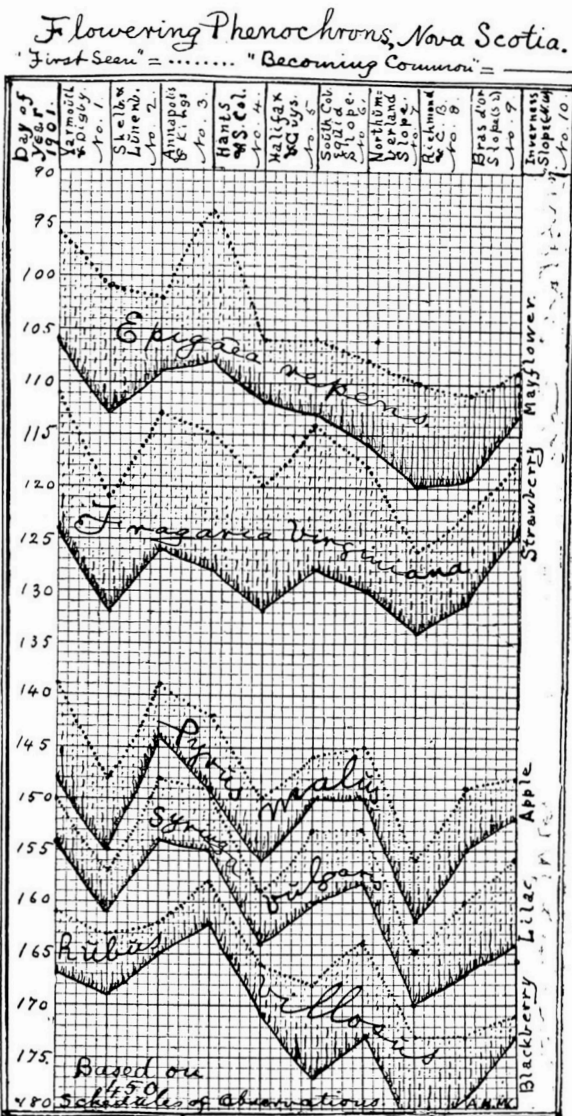


Fig 2 In his 1901 report (MacKay 1903), MacKay presents this graphic of the first appearances of mayflower, strawberry, apple, lilac, and blackberry (summary data rows 3, 13, 51, 57, and 30, as noted in the report (p. 495).

mayflower, strawberry, apple, lilac, and blackberry (summary data rows 3, 13, 51, 57, and 30, as noted in the report, p. 495). The report includes it along with this remark:

A plate of graphs showing the relation between the flowering phenochrons in each region of the province of Nova Scotia for the dates 'when first seen' and 'when becoming common' is given on page 496. 'When becoming common' must always be a matter of personal judgement; so that the general conformity of the five pairs of curves for the flowering of the Mayflower, Strawberry, Apple, Lilac, and Blackberry, on the said plate is very interesting.

In the Report from 1900, MacKay noted:

...the curves for the three years are to a great extent conformable, which demonstrates the important effect of the position of each county. The variations from conformability, are probably due to the differences in the winds and sunshine. (MacKay 1901).

This graphic was obtained from the summary data, a subset of 450 of the more than 500 data sets submitted by his many teachers that year. The regions are sorted by latitude: MacKay's evident supposition was that first appearances would come later the more northerly a locality. We wonder if he chose those variables because they spanned the season, and don't generally overlap – so that the distinction between each of the species will be (generally) clear. We haven't been able to find an explanation for his choice. However, given that MacKay felt that these five species are important enough to focus on, we too decided to focus on them. They are well-represented across the years (with minor exceptions). It may be that he suspected that they would show the most dramatic results, or that they were easiest for his students to identify (so were likely to be the most accurately represented). At any rate, the analyses in this paper pertain only to five species that MacKay chose for the 1901 graphic: mayflower, strawberry, apple, lilac, and blackberry.

Combining Regions

Another important problem that we confronted was that, over the years, some regions were split up, and others were joined together: MacKay's regions ("phenochrons" – Fig 3) varied over time. For example, some years combined regions 9 and 10 to create one data point. In other years, region 6 was split into two separate parts, 6A and 6B. These inconsistencies lead to problems when it comes to

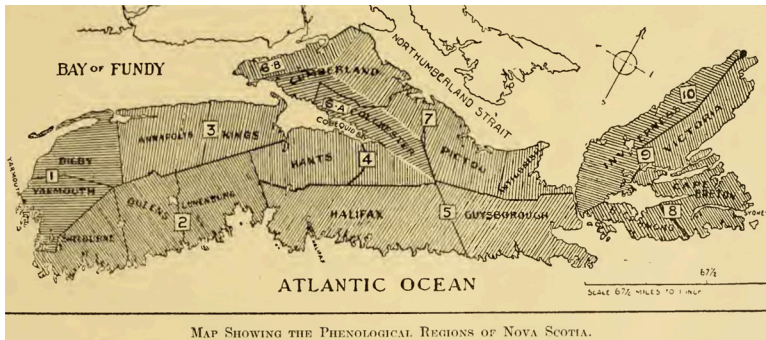


Fig 3 From the Phenology Report of 1909: “Each province may be divided into its main climatic slopes or regions which may often not be coterminous with the boundaries of counties. Slopes, especially those to the coast, should be subdivided into belts such as (a) the coast belt, (b) the low inland belt, and (c) the high inland belt.”

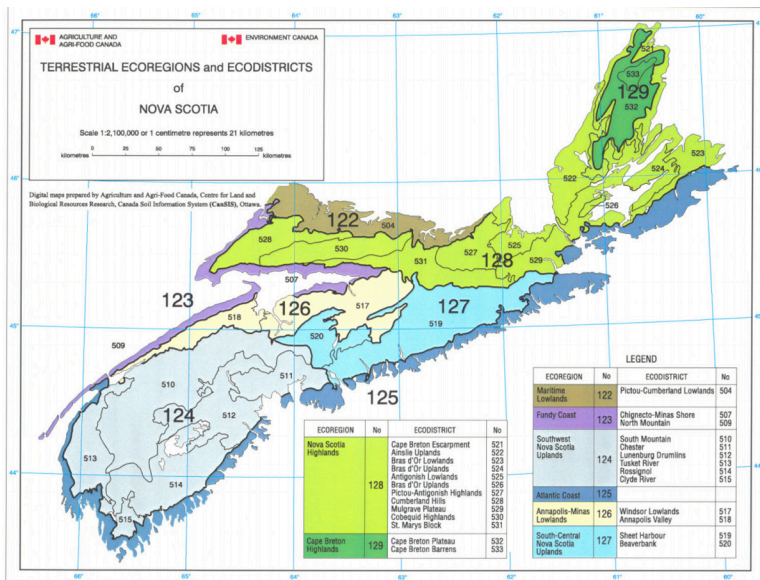


Fig 4 Map of the Terrestrial Ecosystems and Ecodistricts of Nova Scotia (Webb and Marshall 1999), obtained from <http://sis.agr.gc.ca/cansis/publications/surveys/ns/nsee/index.html>.

data analysis. We elected to create a standard set of nine regions to utilize for analysis. Regions 1-5 and 7-8 are identical to those created by MacKay. Region “6” is identical to MacKay’s region 6 in most

cases. Where there is a 6A and 6B in MacKay's original data set, our region "6" is a combination of 6A and 6B. In all cases, regions 9 and 10 were combined to create one region. In the cases where MacKay combined regions 9 and 10, this last region is identical to the original. The configuration of the final data set used is shown in Table 1, and we describe in that table how the regions were averaged together.

Modern day analysis (Webb and Marshall 1999) suggests that the ecosystems of Nova Scotia should look more like the subdivisions of Fig 4.

Data Imputation Using the Singular Value Decomposition (SVD)

The summary data we study consists of five species, reported in nine regions, over 23 years. We can think of each year's data as a matrix: five rows for the five species, and nine columns for the nine regions. Data from 1910 was "accidentally misplaced" (MacKay 1911), and MacKay's report from that time period is discouraging (MacKay 1909). It appeared for a time that the phenological reporting might cease; however, it recommenced in 1911, and continued unabated from then until 1923.

Missing or Corrupt Data in a Matrix

One small challenge was a dozen or so missing data values: several years were missing values for one or more species. For that, we required a method of imputation (estimation), and we chose a method based on the Singular Value Decomposition (SVD). A glance at the 1901 graphic strongly suggests that the five species' FATs are synchronized; they are in lockstep, to a certain degree. Clearly there is some underlying similarity in the graphics: as MacKay remarked, "...the general conformity of the five pairs of curves for the flowering of the Mayflower, Strawberry, Apple, Lilac, and Blackberry, on the said plate is very interesting." (MacKay 1903).

We suspect that the similarity is a consequence of species' relatively similar response to their geographical location (longitude and latitude). There is a certain one-dimensional commonality to their behavior, and the SVD is a good tool for identifying this one-dimensional commonality. In what follows we turn this idea on its head, however, and use the requirement of commonality to estimate the missing values using the SVD.

Table 1 This table lays out how regions were combined: 1) “Original” means that the new region is identical to the one created by MacKay. For 6A and 6B, it means that the regions were not originally split into two parts. For regions 9 and 10, this means that the regions were combined by MacKay, and were therefore not changed for the final data set. 2) In cases where we had to combine regions, the weighted ratios are given. 2a) If the ratio is “given,” we were able to find the number of reports for each region and use that ratio to weight the data points. For example, in 1917, there were six reports from 6A and fifteen reports from 6B. Therefore, to create a region “6,” A was weighted by six and B was weighted by fifteen. 2b) If the ratio is “estimated,” there was no information given on the number of reports from each region. In order to create a ratio to use, we used the total ratio of reports from the years in which a ratio was given. For regions 6A and 6B, the ratio is combination of the years 1913-1919. For regions 9 and 10, the ratio is a combination of the years 1903, 1906, and 1918. In some cases, the data points given for regions were the same, so the regions ended up being given equal weight.

Year	Regions 6A and 6B	Regions 9 and 10
1901	Original	Weighted 14/27—estimated
1902	Original	Weighted 14/27—estimated
1903	Original	Weighted 4/12—given
1904	Original	Weighted 14/27—estimated
1905	Original	Weighted 14/27—estimated
1906	Original	Weighted 1/6—given
1907	Original	Original
1908	Original	Original
1909	Original	Original
1910		
1911	Original	Weighted 14/27—estimated
1912	Original	Original
1913	Original	Original
1914	Original	All same—weighted equally
1915	Original	Same except strawberry—weighted 14/27
1916	Original	Original
1917	Weighted 6A/15B--given	Original
1918	Weighted 5A/13B--given	Original
1919	Original	Original
1920	Weighted 3A/7B--estimated	Original
1921	Weighted 3A/7B--estimated	Original
1922	All same—equal weight	Original
1923	Weighted 3A/7B--estimated	Original

The SVD Theorem: Decomposing Matrices

Our strategy in this paper owes its starting point to a powerful tool in linear algebra – the Singular Value Decomposition (SVD) – and to strategies based on it and explored by one of the authors in his dissertation (Long, 1994) and in a subsequent publication (Long and Long, 2001). The latter publication introduced a strategy for

interpolating (or estimating) a matrix that was described as “skinning the matrix, even though we are putting a skin on rather than taking one off.” The authors emphasized the flexibility of their skinning scheme, which allowed for an unlimited number of different skins. Each choice of one-dimensional interpolators gives rise to a different skin, with characteristics derived from the one-dimensional interpolators. Long in his dissertation explored the generalization of the SVD to tensors. In this paper, we explore the generalization of the SVD to tensors, but “skin the tensor” as well. In particular, we take advantage of the flexibility of the interpolation scheme to explore one-dimensional linear regressors to create the skin for the tensor. One could then use the error bounds of the regressors to deduce error bounds for the tensor.

The Singular Value Decomposition theorem has been called “The Fundamental Theorem of Linear Algebra” by one of the masters of the subject (Strang 1993). It states that a real matrix can be decomposed into a product of orthogonal matrices and a diagonal matrix, $A = U\Lambda V^T$ where Λ is diagonal with positive entries $\lambda_1, \dots, \lambda_n$; these are called the singular values, ordered from largest to smallest (with possible ties and zero values), and U and V are orthonormal (have columns mutually orthogonal and of unit length).

Another way of representing the “decomposition” is as a sum of outer products: this is the important representation of that we will focus on. The SVD says that

$$A = \lambda_1 \underline{u}_1 \underline{v}_1^T + \dots + \lambda_m \underline{u}_m \underline{v}_m^T$$

where \underline{u}_i and \underline{v}_i the columns of U and V . The terms $\lambda_k \underline{u}_k \underline{v}_k^T$ are the skeletons of the “skins” mentioned above. As before, we may “throw away” some of the outer-products (generally those corresponding to the smallest singular values) and write

$$A \approx A^* \equiv \lambda_1 \underline{u}_1 \underline{v}_1^T + \dots + \lambda_k \underline{u}_k \underline{v}_k^T,$$

where $k < m$. We then model A with A^* (having removed what we construe as noise). A number of imputation methods are based on this SVD technique of ignoring the “noise” of all but the most important singular values (Arciniegas-Alarcon, et al. 2014). We use an imputation technique that fits into this class of algorithms. We choose missing values so that the first singular value λ_1 is as

large as possible relative to the other singular values. In particular, we maximize .

$$\frac{\lambda_1}{\sqrt{\lambda_1^2 + \dots + \lambda_n^2}} .$$

MODEL BUILDING

MacKay gave us some help in our modeling process. He was a thorough scientist, whose preliminary analysis and speculation provided us some direction (MacKay 1901). He speaks to the importance of altitude, the proximity to the Atlantic, to the general SW to NE trend, to differences in sampling patterns; concluding:

[t]he manner in which the other curves intersect each other have also their explanations. But we are not yet in a position to be able to state them.

The general trend is seen in the later flowering as the counties lie north and east. There is a general conformity in this trend between the eight plants which fall into four groups, the Mayflower averaging 113.01 (24th April), the Dandelion and Strawberry 137 + (18th May), the Wild Cherry, Blueberry, Buttercup and Apple 154 + (4th June), and the Lilac 165 + (15th June).

A general trend is also seen in passing from Guysboro in the east back to Cumberland in the west; although moving on the whole northward, the flowering becomes earlier. A similar change takes place in passing from Richmond to Cape Breton. This latter is more remarkable, for Cape Breton is not only north but also east of Richmond. This seems to suggest that the observers in Guysboro and particularly in Richmond, might not have been so keen in the search for the first flowering as those in Cumberland and Cape Breton. The small number of observers in these counties also suggests such a possibility. But by reference to the table, it will be seen, that as a rule, in counties where the observation stations are so numerous that ten could be selected from the coast, ten from the low inlands, and ten from the high inlands, the earliest flowering is on the low inlands, then on the coast, and latest on the highlands. It must be remembered, that there is a very great difference in the altitudes of what are called the low and high inlands in the different counties.

From such considerations, it is proposed in future to divide the Province into meteorological districts and subdistricts, instead of counties – the subdistricts being the coast belt, low inland belt and highland belt of each district; each district including a simple meteorological region or geographic slope.

Among the peculiarities shown by these curves are, for instance, the lateness of the Strawberry as compared with the Dandelion in Shelburne, Queens and Guysboro; and its advanced appearance in Kings, Cumberland, Inverness and Victoria. Does the breath of the Atlantic retard the flowering of the Strawberry as compared with the Dandelion?

It also appears that the southern and sea surrounded Yarmouth is favorable to the early flowering of the Mayflower, but comparatively not so favorable to the Lilac. The manner in which the other curves intersect each other have also their explanations. But we are not yet in a position to be able to state them.

The stations of observations are, necessarily, not the same in each county each year. It is therefore possible that the phenochrons might be affected by a change in the relative number of coastal, inland and highland stations.

As all these observations are bound carefully into a large volume for each year, anyone having the time can use the facts recorded in any combination promising the most useful results. The present selection of ten plants, and the comparison of their flowering phenochrons in each county is merely a sort of preliminary or provisional testing of the possibilities and probably value of such observations – sufficient to interest the observers while they are developing accuracy – and a record of facts for future generalization.

....

In the meantime we can make no mistake in recording and preserving as many accurate local phenological facts as possible. In a few years we shall be better able to estimate their value for many purposes.

THE SINGULAR VALUE DECOMPOSITION (REPRISE)

Tilted Geographical Coordinates

The SVD is useful for separating out the three dimensions of our data tensor: the species, the region, and the year. At the outset we did an analysis of FAT of each species by region, year after year, as a function of latitude and longitude. As mentioned previously, we expected the latitudinal component to be significant, representing as it does the northerliness of a region. We were surprised to discover, however that as often as not longitude played a role. A glance at Fig 4 shows a couple of reasons for why this might be:

1. Longitude is a proxy for latitude: the further east one goes, the further north one goes. So the two measures are confounded.
2. Longitude represents how far the province juts out into the Atlantic Ocean. Ocean currents vary about Nova Scotia (Wu and Tang 2011), and there are several different major bodies of water that impact the province. Hence, longitude may play a more complex role.

However, the first of the two issues impelled us to try rotating the coordinates to align one along the major axis of Nova Scotia (thinking of it as an ellipsoid) and the other along the minor axis. These tilted longitude/latitude coordinates were then used for further modeling, and we found that we were far more likely to achieve significant results using them. Table 2 illustrates a case for which neither geographical coordinate was significant on its own, but for which the tilted coordinates produced a significant and useful model.

A Tensor Version of the SVD

A precise analogue for the SVD does not exist for the tensor case (Kolda and Bader 2009). However, we can directly extend one method for constructing the SVD to tensors, leading to the method we call the TSVD – Tensor SVD.

One method for constructing the SVD of $m \times n$ matrix A of rank r entails maximizing the scalar product

$$\underline{u}^T A \underline{v}$$

over all unit vectors \underline{u} and \underline{v} . This will succeed because we are maximizing this product over a closed and bounded set (the cross-product of spheres in m and n dimensions, respectively).

Table 2 This example regression illustrates a case for which neither geographical coordinate was significant on its own, but for which the tilted coordinates produced a significant and useful model. Tilted longitude is significant, as is the intercept term. The issue is that latitude, which we along with MacKay suspected would be the primary determinant of climate, is confounded by longitude. More northerly parts of Nova Scotia are more easterly, and vice versa.

Parameter	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	55.568	291.197	0.191	0.855
Latitude	2.978	4.244	0.702	0.509
Longitude	1.365	1.699	0.804	0.452
Tilted Parameter	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	97.1914	2.4968	38.926	1.92e-08
Tilted Latitude	2.3314	4.5091	0.517	0.6236
Tilted Longitude	2.3017	0.7507	3.066	0.0221

Once we compute \underline{u}_1 and \underline{v}_1 for the maximum value λ_1 of $\underline{u}^T A \underline{v}$,

$$\lambda_1 = \underline{u}_1^T A \underline{v}_1$$

we remove that weighted outer-product

$$A_1 = A - \lambda_1 \underline{u}_1 \underline{v}_1^T$$

and then repeat the process on matrix A_1 . We continue recursively removing these weighted outer-products where λ_k is the maximum value of $\underline{u}^T A_{k-1} \underline{v}$ over all possible vectors \underline{u} and \underline{v} . Unit vectors \underline{u}_k , \underline{v}_k are the ones that realize this maximum value

$$\lambda_k = \underline{u}_k^T A_{k-1} \underline{v}_k$$

Once we have found λ_k , \underline{u}_k , \underline{v}_k , we use them to construct a lower rank matrix.

$$A_k = A_{k-1} - \lambda_k \underline{u}_k \underline{v}_k^T$$

Because each subtraction reduces the rank by one, eventually the resulting matrix is the zero matrix, so

$$A = \sum_{k=1}^r \lambda_k \underline{u}_k \underline{v}_k^T$$

This is just a different representation of the singular value decomposition theorem, in the form of a method.

We can carry out this same method in the more general tensor case. In our case we have a 3-tensor given by the three-dimensional data set of (*species, region, year*).

$$T = \sum_{k=1}^{16} \lambda_k \underline{s}_k \otimes \underline{r}_k \otimes \underline{y}_k$$

where the tensor product \otimes is this same outer product. In the tensor case the value of N is not as well defined as in the matrix case, and there are other important distinctions between the decompositions in the tensor and matrix cases.

THE FINAL TENSOR MODEL

Having written our data tensor T as a sum of 16 such terms,

$$T = \sum_{k=1}^{16} \lambda_k \underline{s}_k \otimes \underline{r}_k \otimes \underline{y}_k$$

we proceeded to model the different dimensions (species, region, year) in three different ways.

The species vectors were left as is: we didn't do any modelling for those. Our efforts were focused on the other two dimensions, region and time. We used linear regression to model the regions vectors as functions of tilted longitude and latitude (*tlong* and *tlat*), whereas the time series (year) vectors were modelled as functions of climate variables (notably average monthly temperatures up to a year prior to flowering, and sea ice off the coast of Newfoundland).

Whereas the climate variables were included only linearly, we allowed quadratic terms for the region model (*tlat* and *tlong* variables). We assumed that we might have coastal effects: and since the *tlat* and *tlong* coordinates run along the "major axis" and "minor axis" of Nova Scotia, we felt that a linear model would not be appropriate – a planar model of FATs across Nova Scotia – but a parabolic model would allow for one value along the coasts on opposite sides of the "heights of Nova Scotia", and a different value for the heights themselves. Fig 5 shows how one quadratic model appears in the province.

Having modelled both these two dimensions, we arrive at a model of that we represent schematically as

$$\underline{T}_a(\textit{tlong}, \textit{tlat}, \textit{climate}) = \sum_{k=1}^{16} \lambda_{i(j)} \underline{s}_{i(j)} r_{i(j)}(\textit{tlong}, \textit{tlat}) y_{i(j)}(\textit{climate})$$

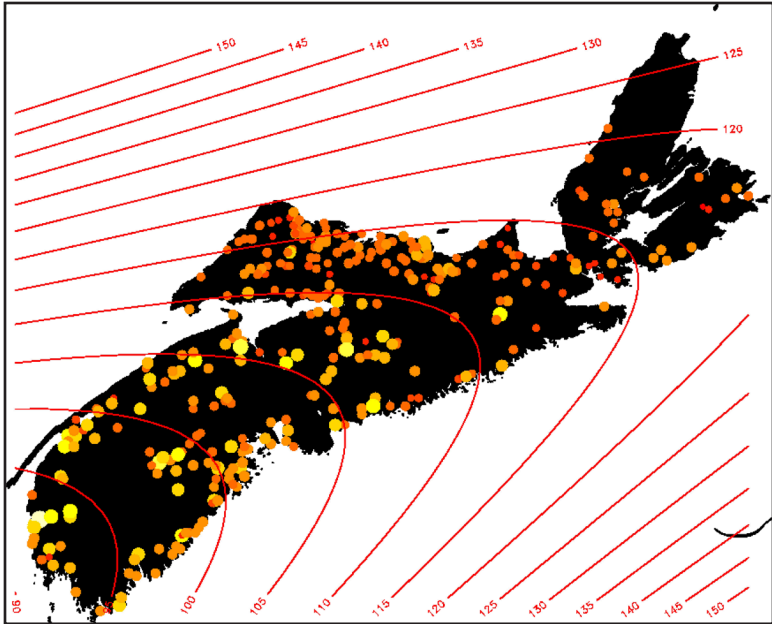


Fig 5 This illustrates our projections of First Arrival Time for Mayflower for the year 1901. Numbers represent day since beginning of the year. One can see the advantage of the quadratic FAT model over a linear model: the isoFATs bend over the highlands of Nova Scotia and wrap from coast to coast.

where time is embedded in the climate variables. Details of the modeling procedure are provided in an appendix.

RESULTS

Because we now have a continuous model of species as a function of location and time, our model can be used to forecast FATs for any year for which we have the same climate data, at any location.

How the Model Performs

Model predictions for the FATs of the original data are given in the succession of small multiples shown in Fig 6. The data and model fits are available for visual comparison for all five species across all 23 years.

As one can plainly see, some species are better reconstructed generally than others. For example, it is clear that mayflower is better reconstructed than blackberry, as a rule. That being said, in

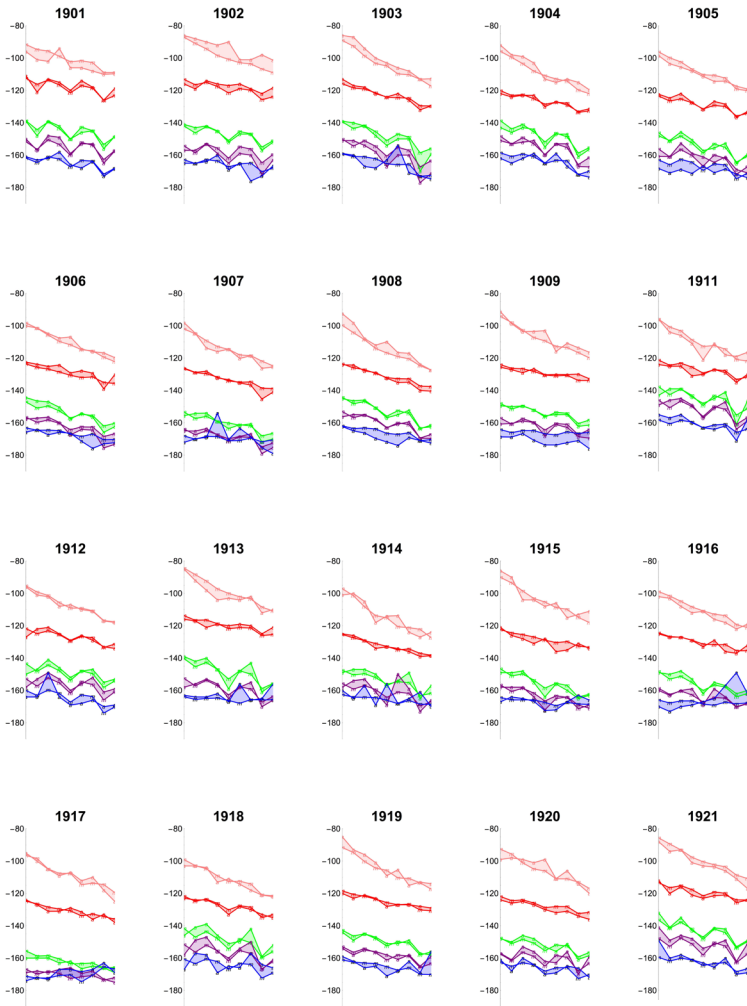


Fig 6 Model results (comparison with data) for the FATs from 1901-1923. We use Tufte's "small multiples" (Tufte 2004) to help us gauge the effectiveness of our model over the entire span of the data.



some years, blackberry may be better predicted than mayflower, etc. Average error in a species prediction, by region, is a reasonable “diagnostic” to provide for using model predictions in the future. The average error across the entire data set is 2.7 days.

Year, region, and species means of absolute errors across the years are given in Tables 3, 4, and 5. As one can see, strawberry is best reconstructed, and blackberry worst. Among regions, all are roughly similar, although the combined regions (4 and 5 in the original data) have the largest mean absolute deviation: 3.11 days.

Comparison with Predictions of Others

A recent paper in the *Proceedings* (Hill and Garbary 2013) caught our eye, entitled “Early spring flowering in Nova Scotia: an extreme spring is reflected in advanced flowering”. It features data on many species of flowering plants, but in particular it included recent data (2012) on two of our species: mayflower and strawberry. They provided coordinates for their search (latitude and longitude), so we were able to generate predictions of first flowering of those two species. Our predictions were within five days in each of the cases (see Table 6), and biased toward predicting earlier FATs, rather than later ones. Hill and Garbary note that “...since all of these records were based on opportunistic sampling across a wide geographic area (northern Cape Breton to the Annapolis Valley), we do not claim that the dates represent first flowering of the species.” The implication is that the observed dates should be later than the actual FAT. For that reason,

Table 3 Error by region (in days).

Region	Average Error (Days)
Yarmouth and Digby	2.77
Shelburne, Queens, and Lunenburg	2.45
Annapolis and Kings	2.37
Hants and Colchester	2.44
Halifax and Guysboro	2.81
Cobequid Slope	2.76
Northumberland Straits Slope	2.71
Richmond and Cape Breton	2.75
Bras d’Or and Inverness Slopes	3.11

Table 4 Error by species (in days).

Species	Average Error (Days)
Mayflower	3.02
Strawberry	1.88
Apple	2.22
Lilac	2.67
Blackberry	3.64

Table 5 Error by year (in days). 1918, 1914, 1922, 1903, and 1923 were the most poorly reconstructed. The run from 1904-1913 was fairly well reconstructed.

Year	Average Error (Days)
1901	2.17
1902	3.04
1903	3.29
1904	2.17
1905	2.59
1906	2.50
1907	2.46
1908	2.08
1909	2.65
1911	2.69
1912	2.43
1913	2.64
1914	3.19
1915	2.59
1916	2.82
1917	2.30
1918	3.11
1919	2.34
1920	2.55
1921	2.50
1922	3.24
1923	3.76

Table 6 Predictions against data (Hill and Garbary 2013), including predictions for 2012 (for which “temperatures for the late winter and spring of 2012 in Nova Scotia were considerably elevated relative to climate normal”) and for the years adjacent to 2012. As one can see, our predictions suggest that 2012 was, indeed, anomalous, and that flowering should have occurred about six days earlier due to the warmer temperatures.

Location	Species	2012 Prediction	2012 Observation	2011 Prediction	2013 Prediction
Antigonish	Mayflower	April 10	April 15	April 16	April 17
Antigonish	Strawberry	April 25	April 28	May 1	May 2
Annapolis	Strawberry	April 17	April 16	April 21	April 23

the bias in our results towards earlier dates (in Antigonish, at least) is encouraging.

A publication referenced in the Hill and Garbary paper, entitled “Spring-flowering herbaceous plant species of the deciduous forests of eastern Canada and 20th century climate warming” (Houle 2007) performed regressions of flowering times of 18 different plants from eastern Canada, from 1900 to 2000. “...Results show a 2-6 days advance in flowering date over 100

years, depending on the region considered (corresponding to a 2-3 days advance per 1°C); these values are somewhat lower than those published in other studies, but still support the increasing body of literature on the effects of climate warming on plant phenology.”

We highlight the predictions of Houle’s work, which are in exact accord with our own conclusions: 2-6 days advance per century (Figs 7 & 8). He proceeded by regressing on data; we proceeded by regressing on model predictions. Although we used different species, our general agreement is satisfying.

DISCUSSION

We have produced a model that predicts First Appearance Times (FATs) of flowering of five plant species, in Nova Scotia, Canada. The model is a function of average monthly temperatures and winter/spring ice coverage in Newfoundland.

The model was generated by fitting a tensor model to the tensor of data with three dimensions: species, region of Nova Scotia, and year, with five species, nine regions, and 22 years of data. The data were first cleaned and a few values imputed using a technique based on the Singular Value Decomposition (SVD).

The data tensor was decomposed into a sum of rank-one tensors, each of which is an outer-product of three vectors (one in each dimension – species, region, and year). For each rank-one tensor, a model was constructed using linear regression on two of the dimensions: region and year. The region model was a function of longitude and latitude; the year model was a function of average monthly temperatures and winter/spring ice coverage in Newfoundland. The rank-one tensors were then replaced with a vector of five functions, each of which represented one species. When summed in the same fashion as the rank-one tensor decomposition, we produced a set of five functions, each of which predicts FAT based on these geographic and climatic variables.

CONCLUSIONS

Our models as shown in Figs 7 and 8 suggest that climate change over the past 100 years (increasing temperatures, in particular) are causing some plants to come into flower earlier (2-6 days earlier,

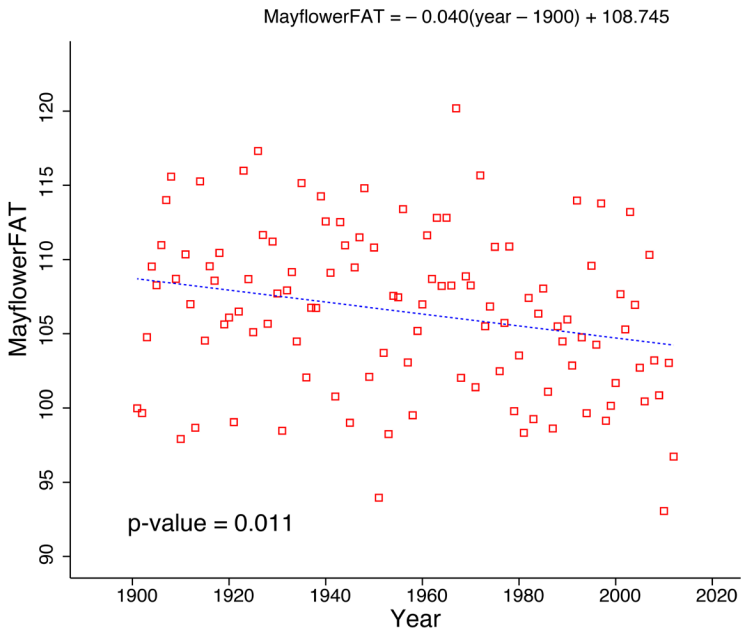


Fig 7 Model projections for Mayflower for one hundred years, showing a significantly earlier flowering time (about 4 days/year).

with all five plants that we studied showing a significant trend to earlier flowering).

FUTURE WORK

Moving forward, we will:

Use a simpler and faster alternative method of tensor decomposition than the one we used here, to see how results compare.

Add additional summary data from MacKay's published reports. In addition, we plan to add two species which overlap with the Hill and Garbary paper (Hill and Garbary 2013), and are well-represented in the MacKay data, to see if our predictions continue to hold up well.

Include comparisons of our model predictions with other data in the literature, including Plant-Watch data (Liette Vasseur 2001), and even MacKay's own data (we did not include his summaries prior to 1901, because they were not completely formulated in terms of his citizen science project prior to 1901).

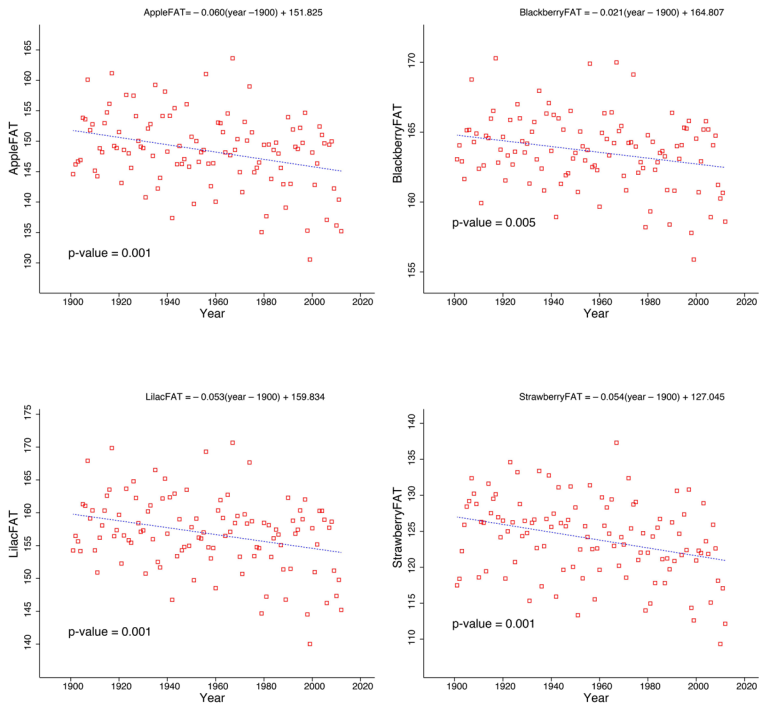


Fig 8 The other four species also showed significantly earlier flowering over the 100 years.

Acknowledgments We received invaluable assistance with

- data (Adam Fenech, Brian T. Hill);
- thoughtful suggestions (Teresa Devor);
- grant support (NKU's College of Arts and Sciences Collaborative Faculty and Student Grant, 2016, and NKU's CINSAMUR-STEM Program);
- student research support from NKU's Department of Mathematics and Statistics (Dr. Roger Zarnowski, chair);
- assistance with typing this document (Taylor House);
- and, perhaps most importantly, inspiration (Long thanks Culbertson for inspiring this project).

REFERENCES

- Austen, I.** (2000). Grassroots science. *Canadian Geographic*. p. 76.
- CaraDonna, P.J., Iler, A.M., & Inouye, D.W.** (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences* 111(13): 4916-4921, doi:10.1073/pnas.1323073111.
- Fenech, A., MacIver, D., Auld, H., & Beal, S.** (2005). Integrated Mapping Assessment, In Impact of climate on changes in the seasonal timing of life cycle events of Eastern Canada from 1901 to 1923, p. 186. Environment Canada, Toronto, Ontario, Canada.
- Hern, T., & Long, C.** (1991). Visualization in Teaching and Learning Mathematics, MAA Notes, in vol. 19, Viewing some concepts and applications in Linear Algebra, pp. 173-190, Mathematical Association of America, Washington, DC.
- Hill, N.M., & Garbary, D.J.** (2013). Early spring flowering in Nova Scotia: an extreme spring is reflected in advanced flowering. *Proceedings of the Nova Scotian Institute of Science* 47(2): 211-220.
- Houle, G.** (2007). Spring-flowering herbaceous plant species of the deciduous forests of Eastern Canada and 20th century climate warming. *Canadian Journal of Forest Research* 37(2), 505–512. doi:10.1139/X06-239.
- Kolda, T.G., & B.W.** (2009). Tensor decompositions and applications, *SIAM Review* 51(3): 455-500. doi: 10.1137/07070111X.
- Liette Vasseur, P.J.M., & Guscott, R.L.** (2001). Monitoring of spring flower phenology in Nova Scotia: Comparison over the last century. *Northeastern Naturalist* 8(4): 393-402.
- MacKay, A.H.** (1898). Phenological observations, Canada, 1896. *Proceedings and Transactions of the Nova Scotian Institute of Science* 9(3): 268-274.
- MacKay, A.H.** (1899). Phenological observations in Nova Scotia and Canada, 1898. *Proceedings and Transactions of the Nova Scotian Institute of Science* 10(1): 91-109.
- MacKay, A.H.** (1901). Phenological observations, Canada, 1896. *Proceedings and Transactions of the Nova Scotian Institute of Science* 10(3): 379-398.
- MacKay, A.H.** (1902a). Phenological observations in Canada. *The Canadian Record of Science* 8(2): 71-84.
- MacKay, A.H.** (1902b). Local 'nature' observations. *Journal of Education* 4(1): 24-27.
- MacKay, A.H.** (1902c). Reports on phenological observations, 1901. *Journal of Education* 4(1): 58-63.
- MacKay, A.H.** (1903). Phenological observations in Nova Scotia and Canada, 1901. *Proceedings and Transactions of the Nova Scotian Institute of Science* 10(4): 486-501.
- MacKay, A.H.** (1911). Phenological phenomena, Canada, 1910. *Proceedings and Transactions of the Royal Society of Canada* 5(3).

- MacPhie, J.P.** (1914). *Pictonians at home and abroad: Sketches of professional men and women of Pictou county - its history and institutions*, Pinkham Press, Boston, Massachusetts U.S.A.
- Nova Scotian Institute of Science.** (1992). Cumulative author and subject indices to *Proceedings of the Nova Scotian Institute of Science*, Volumes 1-39 (1863 to 2008) inclusive, *Proceedings of the Nova Scotian Institute of Science* 39(4): 149-247.
- Piers, H.** (1930). Obituary notice of Alexander Howard MacKay, B.A., B.Sc., LL.D., F.R.S.C. educationist and scientist, 1848-1929, in *Proceedings of the Nova Scotian Institute of Science*, 17(4): 97-102.
- Sheehan, N.M.** (1974). Alexander H. MacKay, social and educational reformer. In: *Profiles of Canadian Educators*, pp. 253-270.
- Strang, G.** (1993). The fundamental theorem of linear algebra. *The American Mathematical Monthly* 100(9): 848-855.
- Tufte, E.R.** (2004). *The Visual Display of Quantitative Information* (Second Edition). Graphics Press, Cheshire, Connecticut.
- Webb, K., & Marshall, I.** (1999). Ecoregions and ecodistricts of Nova Scotia. Tech. Rep., Crops and Livestock Research Centre, Research Branch, Agriculture and Agri-Food Canada, Truro, Nova Scotia; Indicators and Assessment Office, Environmental Quality Branch, Environment Canada, Hull, Quebec.
- Wood, B.A.** (1994). Constructing Nova Scotia's 'scotchness': the centenary celebrations of Pictou academy in 1916. *Historical Studies in Education* 6(2): 281-302.
- Wu, Y.S., & Tang, C.L.** (2011). Atlas of ocean currents in Eastern Canadian waters. *Canadian Technical Report of Hydrography and Ocean Sciences* No. 271. 94 pp.

APPENDIX A: DATA TREATMENT AND ANALYSIS DETAILS

MODEL DETAILS

1. We begin with the data tensor, which we can envision as a three-dimensional stack of first arrival times (FATs) for five species, nine regions, and 22 years.
2. From T we create the "anomaly tensor" T_a , by subtracting off from each year the mean matrix of the time series of the five species and nine regions (Table 7):

$$T_a = T - M_{5 \times 9} \otimes \underline{1}_{22}$$

Table 7 The mean matrix M , of mean FATs of the time series of FATs for the five different species of plants over the nine regions (moving generally southwest to northeast, hence illustrating the general trend toward later flowering along that axis).

mayflower	strawberry	apple	lilac	blackberry
94.4	121.5	145.8	156.4	163.7
98.6	123.8	146.9	156.5	164.4
102.2	123.4	145.0	154.2	163.4
106.2	125.9	149.5	157.5	164.9
108.3	128.4	155.8	163.9	167.7
111.9	127.2	152.5	159.3	166.6
112.7	129.0	151.0	158.3	166.0
116.4	133.0	161.9	169.2	168.4
117.9	131.8	158.3	165.5	167.5

The matrix crossed with the one-vector of length 22 represents a tensor where each slice along the temporal dimension is the same – containing the mean FATs for each species in each region. Thus, T_a represents not FATs, but rather the difference in FAT from the mean FAT for that species in that region. As a consequence, the mean of each temporal column of anomaly tensor T_a is 0.

Our objective is to model T by modeling the anomaly tensor T_a and the matrix M separately: M as a function solely of latitude and longitude, and T_a as a function of latitude, longitude, and climate variables.

- Latitude and longitude are “tilted” to align with the shape of Nova Scotia, which runs generally southwest to northeast (discussed in the text). So our regressions use “tlong” and “tlat”, defined as

$$\text{tlong} = \cos(20) (\text{longitude} + 65.8094) - \sin(20) (\text{latitude} - 41.1496)$$

$$\text{tlat} = \sin(20) (\text{longitude} + 65.8094) + \cos(20) (\text{latitude} - 41.1496)$$

The choice of a 20-degree tilt was made by visual examination; the centroid was simply chosen to lie within Nova Scotia, along what we perceived to be the major and minor axes.

- To model M we first decompose M using the SVD, producing the usual rank five decomposition

$$M = U\Lambda V^T = \sum_{i=1}^5 \lambda_i \underline{u}_i \otimes \underline{v}_i$$

The \underline{u}_i we may construe as species-related, and we do not model those; the \underline{v}_i are region-related, and we model by regression on latitude, longitude, and their squares.

The first three singular products contained \underline{v}_i which produced a significant $p < .05$ model, whereas the last two did not. Therefore, the first three outer products only were used to create the vector function of latitude and longitude (a mean surface, or rather a set of five of

them – one for each species). We give below an explicit representation of M as a vector model, composed of a sum of three pieces, each comprised of the product of a singular value weighting a dot product of two vectors, all of which scales the vector of species weights \underline{u}_i on that outer-product pair:

$$\begin{aligned} & \underline{M}(tlong, tlat) \\ &= 967.6(-0.318 - 0.022tlat^2 + 0.00696tlat \\ &\quad - 0.00511tlong) \langle -0.334, -0.394, -0.471, -0.497, -0.514 \rangle \\ &+ 16.96(-0.571 - 0.0149tlat^2 + 0.379tlat \\ &\quad + 0.251tlong) \langle 0.844, 0.0229, 0.0905, -0.148, -0.506 \rangle \\ &+ 7.064(0.0966 - 0.786tlat) \langle -0.314, -0.206, 0.506, 0.505, -0.590 \rangle \end{aligned}$$

5. The model for the anomaly tensor T_α is a little more complicated. We decomposed it, using the procedure described in the text, to obtain

$$T_\alpha = \sum_{i=1}^{16} \lambda_i \underline{s}_i \otimes \underline{r}_i \otimes \underline{y}_i$$

where \underline{s}_i , \underline{r}_i , and \underline{y}_i , are the i^{th} species, region, and year vectors, respectively. They are essentially singular vectors of a tensor version of the singular value decomposition.

We reduced our study to just the first sixteen products by considering only those outer-products that showed some dependence on climate or region variables. However, not all of the first sixteen showed dependence on both climate and region: twelve of the sixteen products produced models that were simultaneously significant for \underline{r}_i and \underline{y}_i (otherwise a particular outer-product in the sum is replaced by the 0 tensor, which is the null model – these were products 6, 9, 10, and 11 – ordered by their relative importance in the decomposition).

The vector model form FAT anomaly thus becomes a sum of twelve vector functions,

$$T_\alpha = \sum_{i=1}^{16} \lambda_i \underline{s}_i \otimes \underline{r}_i \otimes \underline{y}_i$$

where the models for region vectors (and their weights), obtained by linear regression, are given in Tables 8 and 9 below; the particular values of λ used are given in Table 10. Table 11 contains the weights on each outer-product by species. $i(j)$ represents the j^{th} retained index skipping the four listed in Table 10 as insignificant. “Climate” is used here as shorthand, by which we mean temperature and sea ice extent values, which were obtained using relevant data from the years

over which the FAT dates were collected. Temperature records were obtained from Environment Canada for Parrsboro, Halifax, Sydney, and Yarmouth, and then averaged to obtain an estimate for Nova Scotia generally. These locations had the longest records contemporaneous with our FAT data, with some few missing values. Newfoundland ice values were obtained from Brian T. Hill, formerly of Environment Canada, “an estimation of the sea ice extent in the winter months off the east coast of Newfoundland and over the Grand Banks.” (Hill, personal communication)

6. In the end, then, the vector model for T (FAT itself, for the five species as a function of location and climate variables) is obtained simply as the sum of two length-five vector models:

$$\underline{T}(tlong, tlat, climate) = \underline{M}(tlong, tlat) + \underline{T}_\alpha(tlong, tlat, climate)$$

Table 8 Results of backward-stepwise regressions for the 12 significant region vector models. The model $r_{i(j)}(tlong, tlat)$ is obtained by taking the inner-product of the vector given by $(1, tlat, tlong, tlat^2, tlong^2)$ and the vector given by the coefficients listed below (missing coefficients are 0). Note the absence of any linear dependence on the square of the tilted latitude in the anomaly tensor (although the mean surface contains it as a predictor).

j	outer-product index $i(j)$	intercept	$tlat$	$tlong$	$tlat^2$	$tlong^2$
1	1	0.316	0.112			
2	2	0.331				
3	3	0.215				-0.0271
4	4	0.326		-0.131		
5	5	-0.0175				
6	7	-0.479		0.0687		
7	8	-0.114	0.759			
8	12	-0.0109				
9	13	0.0295				
10	14	0.114	-0.730			
11	15	0.503		-0.505		0.0734
12	16	-0.0230				

Table 9 Results of backward-stepwise regressions for the 12 significant year vector models. July and September temperatures which occur in the models are for the year prior. The model $y_{it(j)}$ (climate) is determined by taking the inner-product of the two vectors indicated below, which creates a scalar function of the chosen climatic variables.

Outer-product index			Predictors Coefficients				
1	<i>I</i> 0.751	<i>Apr Ice</i> 2.896E-7	<i>Jan Temp</i> 8.526E-3	<i>Mar Temp</i> -5.350E-2	<i>Apr Temp</i> -8.155E-2	<i>May Temp</i> -0.142	<i>Sep Temp</i> 4.461E-2
2			<i>I</i> -0.767	<i>Apr Temp</i> -0.102	<i>May Temp</i> 0.131		
3			<i>I</i> -0.757	<i>Mar Ice</i> -1.109E-6	<i>Apr Ice</i> 1.123E-6	<i>May Temp</i> 9.731E-2	
4	<i>I</i> 1.26	<i>Jan/Feb Ice</i> 1.517E-6	<i>Mar Ice</i> -1.883E-6	<i>Feb Temp</i> 1.290E-6	<i>Mar Temp</i> 9.865E-2	<i>Apr Temp</i> -0.111	<i>Jul Temp</i> -9.307E-2
5				<i>I</i> -1.86	<i>Jun Temp</i> 0.144		
7				<i>I</i> 0.401	<i>Jan/Feb Ice</i> -1.389E-6		
8	<i>I</i> -3.00	<i>Jan/Feb Ice</i> 9.872E-7	<i>Mar Ice</i> -2.212E-6	<i>Feb Temp</i> -8.615E-2	<i>Mar Temp</i> -5.168E-2	<i>Apr Temp</i> 7.938E-2	<i>Jul Temp</i> 0.152
12		<i>I</i> 2.10	<i>Apr Ice</i> 1.052E-6	<i>Jul Temp</i> -0.141			
13			<i>I</i> -1.33	<i>Jun Temp</i> 0.103			
14		<i>I</i> -0.189	<i>Jan Temp</i> -5.205E-2	<i>Mar Temp</i> 5.658E-2			
15		<i>I</i> 0.518	<i>Mar Temp</i> 7.20E-2	<i>Apr Temp</i> -0.117			
16			<i>I</i> -0.344	<i>Jan/Feb Ice</i> 1.191E-6			

Table 10 Singular values, and those that we included based on their simultaneous dependence on both place and climate.

Outer-product index	Singular Value	Used (Y/N)
1	124.419	Y
2	73.0337	Y
3	40.7310	Y
4	32.0933	Y
5	31.4943	Y
6	28.3814	N
7	26.9661	Y
8	26.5059	Y
9	20.8977	N
10	20.7605	N
11	18.4311	N
12	18.2133	Y
13	17.3022	Y
14	16.9779	Y
15	15.4649	Y
16	14.9188	Y

Table 11 Species weights for each of the twelve outer-products used in the model. These might be studied to interpret the outer-product pairs biologically, as one would do in a factor analysis.

mayflower	strawberry	apple	lilac	blackberry
0.482	0.492	0.526	0.459	0.195
0.667	0.300	-0.334	-0.504	-0.315
-6.301E-2	-0.113	-0.188	-0.284	-0.931
-0.440	-0.161	-0.647	-0.594	-9.621E-2
0.224	9.209E-2	-1.057E-2	-0.168	-0.955
-0.677	0.616	0.173	-0.172	-0.321
-0.520	-0.414	-0.348	-0.366	-0.550
0.440	0.123	0.236	0.184	-0.838
0.552	0.205	-0.267	-0.720	-0.252
-0.865	-9.634E-2	-0.125	0.476	2.309E-2
-0.817	-8.398E-2	0.537	0.194	-7.051E-3
-0.406	-0.424	-0.416	-0.471	-0.511

DATA MANIPULATION

- Implementations of our SVD imputation is available in three different programming languages—Mathematica, Python, and R:
www.nku.edu/~wilkinson/links/MacKay/software/
- Our original data set of FATs of the five chosen species transcribed from the Proceedings, 1901-1923, is contained in this csv file of First Arrival Times: norsemathology.org/laura/DatesofArrival.csv
- Our final data set of FATs is (following cleaning, region combining, and data imputation) is contained in this csv file of First Arrival Times: norsemathology.org/laura/DatesofFAT.csv

ADDITIONAL CODE FOR REPLICATING RESULTS

- Mathematica file containing graphics of SVD factors:
norsemathology.org/laura/SVDCoefficientsUpdate.nb

