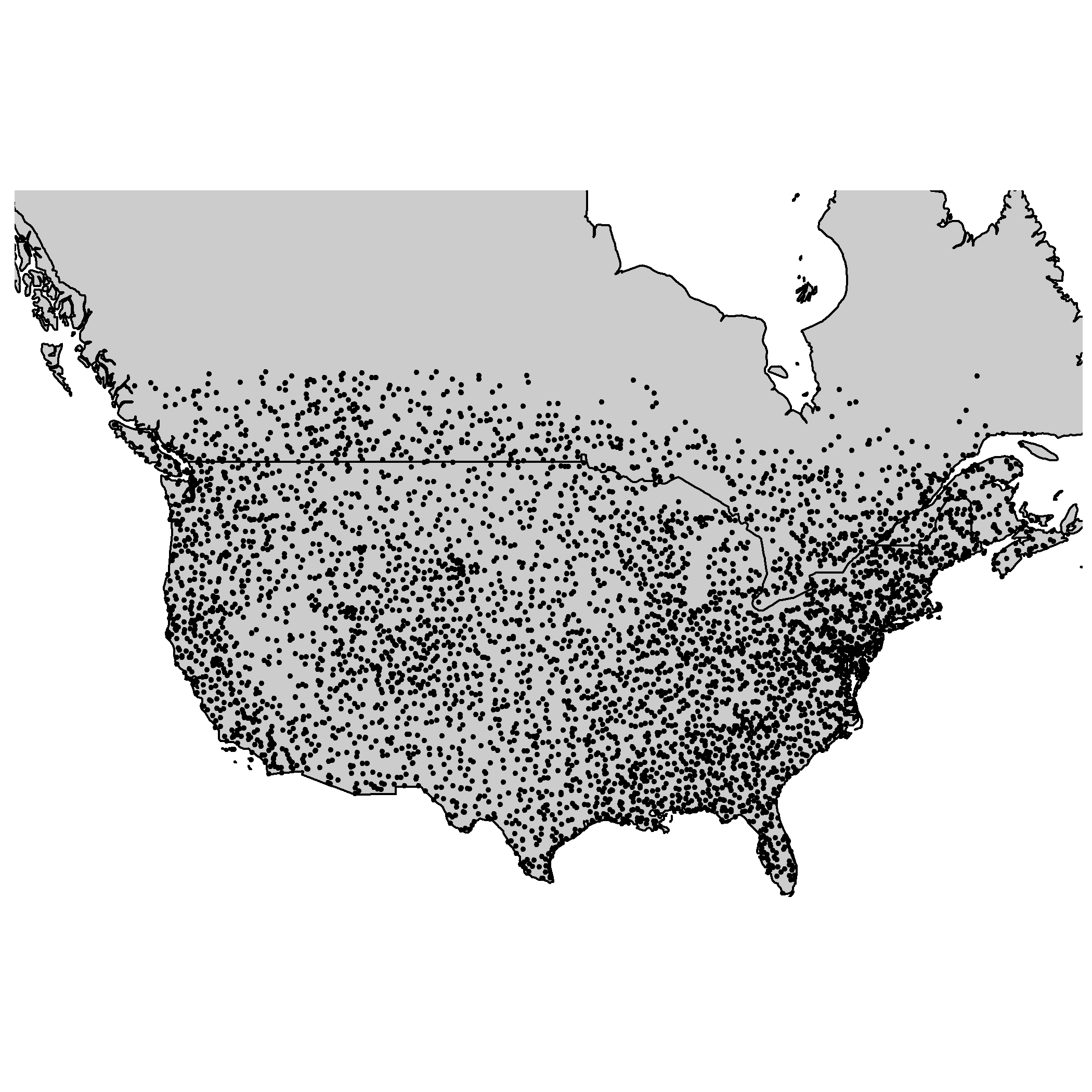
# An application of Fisher Information to bird community data does not reveal distinct regimes in the continental United States

## Introduction

Ecosystems are open, dynamical systems that are most often not easily described using fully parameterized models. Some patterns have emerged in certain statistical mechanics of ecological observations. despite the complexity of most ecological systems. An uptick in recent years of studies of **regime shifts** (Table @ref(tab:glossary)) in ecology has spurred an increase in the number of new methods for detecting ecological regime shifts (Chapter @ref(rdmReview)), some of which have been applied to spatial information [@butitta\_spatial\_2017; @kefi2014early; @sundstrom2017detecting; @guttal2009spatial; @brock\_variance\_2006]. As defined in Table @ref(tab:glossary), a regime shift is largely considered an abrupt and persistent change in a system’s structure or functioning. Following this definition and without considering the pressures (Table @ref(tab:glossary)) associated with the observed regime shift, it is not yet clear whether identifying a ‘spatial regime’ using a snapshot of a system (i.e. using a single or short period of time relative to the time scale of the system dynamics and/or pressures) is pragmatic. A concise and global definition of the spatial regime detection measure is important since observations of non-random spatial processes (e.g., land cover) can manifest as either a rapid shift (e.g. an ecotone) or as a gradual change (e.g., slow mixing along a gradient). Consequently, and because most regime detection measures signal abrupt change, only the former may be identified as “regime shifts” using spatial regime detection measures.

Although it is suggested that statistical and pragmatic methods are advanced more rapidly by bottom-up approaches, i.e. using case studies [see @deangelis2017spatially], there is much work to be done in the way of testing the statistical rigor of spatial regime detection measures. The objective of this chapter is to determine whether the Fisher Information as a regime detection method [Eq. @ref(eq:fiDerivs)] identifies spatial regime boundaries in the bird communities of the continental United States. This chapter is also supported by original software developed for implementation in Program R, which is publicly available (see Appendix @ref(regimeDetectionMeasures)).



Locations of Breeding Bird Survey routes sampled between 1966 and 2017.

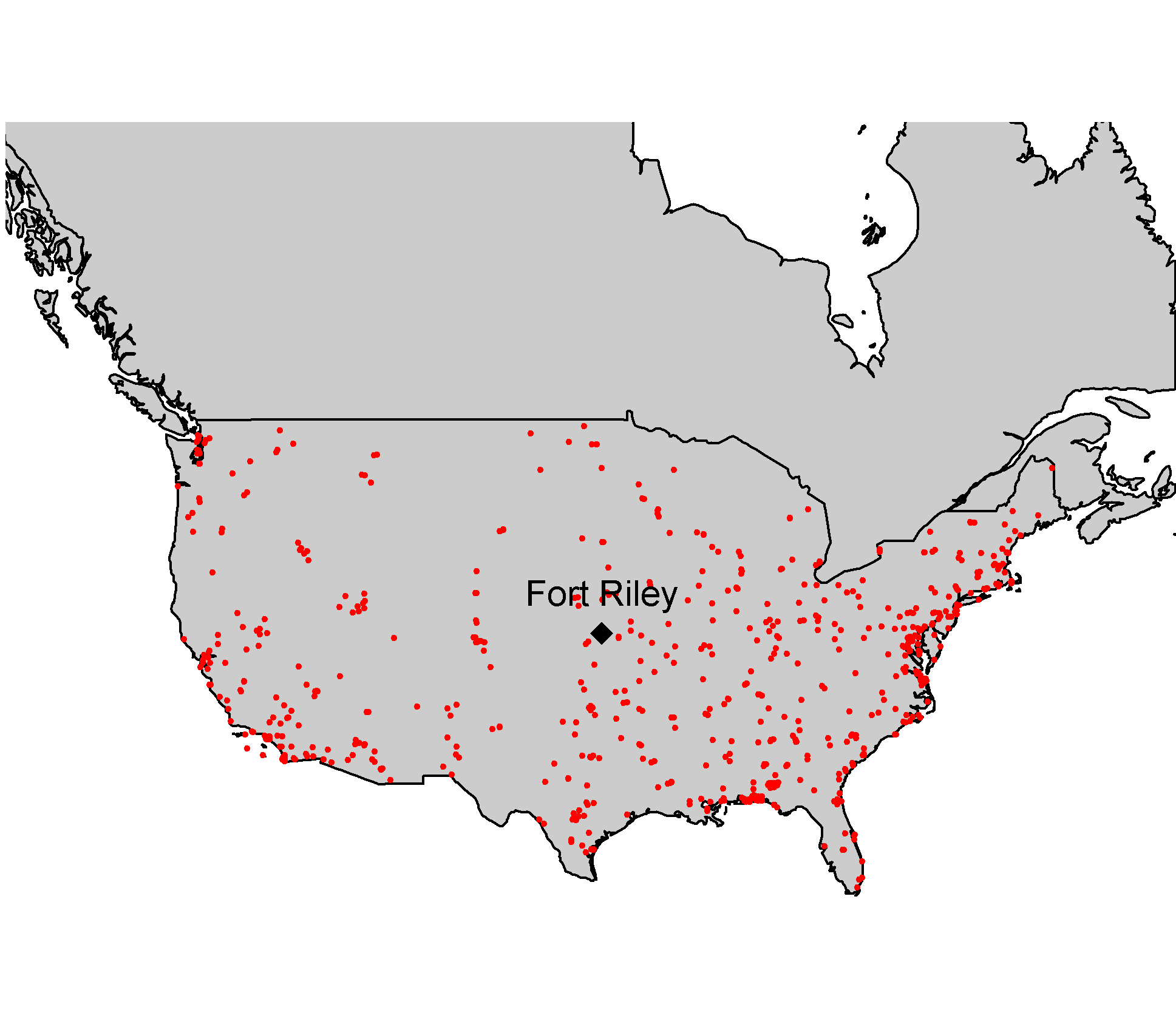
## Data and Methods

### Data: North American breeding bird communities

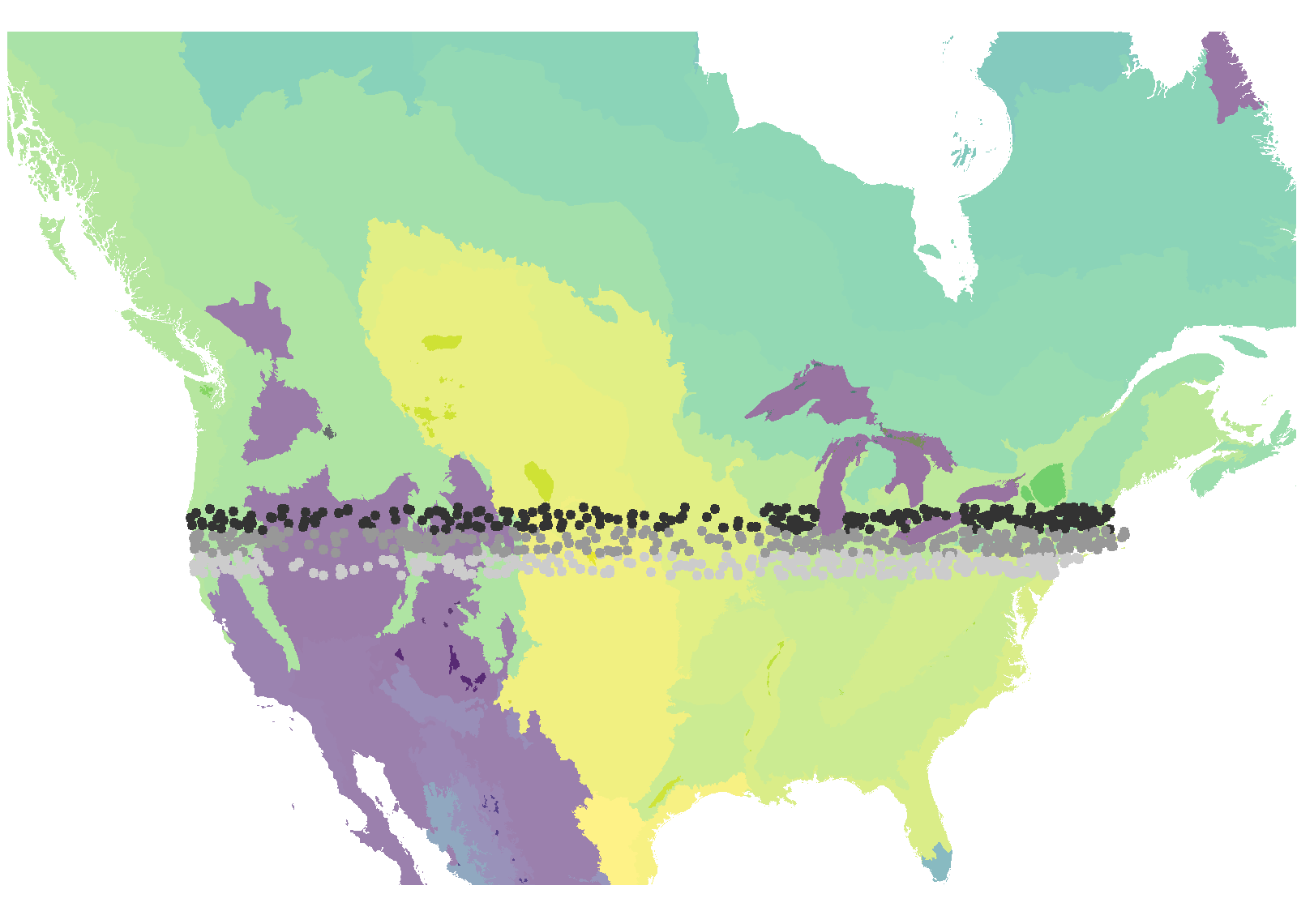
I used community abundance data [@sauer2017results] from long-term monitoring programs to identify spatial and temporal regimes using the Fisher Information (FI) derivatives method (see Eq. @ref(eq:fiDerivs)). The North American Breeding Bird Survey (NABBS) trains citizen scientist volunteers to annually collect data using a standardized roadside, single observer 3-minute point count protocol and has organized data collection annually across North America (Figure @ref(fig:bbsPoints)) since 1966. The roadside surveys consist of 50 point counts (by sight and sound) along mile stretch of road. Due to strict reliance on volunteers, some routes are not covered every year. Additionally, some routes are moved or discontinued due to changing landscape conditions and change in observer safety. Route-year combinations that were missing years but were not discontinued were treated as missing data. Although NABBS volunteers attempt to identify all species as possible, persistent biases exist in this protocol. Despite a standardized survey protocol, some species are difficult to detect using these methods. For example, crepuscular species are less likely to be detected beyond the first few points of the BBS route, given they are most active at sunrise and the survey begins within 30 minutes of sunrise. Further, species which congregate in large groups and are highly mobile (e.g., waterfowl) tend to have less reliable inter-annual abundance estimates given their ability to move long-distances in a short period of time. To remove any potential influence of sampling bias on the Fisher Information result, I removed birds of these types from all analyses: waterfowl, waders, and shore species (BBS AOU numeric codes 0000 through 2880).

### Study area

Although the NABBS conducts surveys throughout much of North America (most of the United States, Canada, and Mexico), coverage of the boreal forests of Canada are sparse in space, and many routes in Mexico have fewer than 25 years of observations. For these reasons I limited analyses largely to the continental United States and parts of Southern Canada (see Figure @ref(fig:bbsPoints)).



Locations of focal U.S. military bases, Eglin Air Force Base (AFB) and Fort Riley Military Base.



Three East-West running transects analyzed in this Chapter overlayed against the Omernick Ecoregion boundaries map.

#### Focal military base

The Mission of the U.S. Department of Defense is to provide military forces to deter war and protect the security of the country, and a primary objective of individual military bases is to maintain military readiness. To maintain readiness, military bases strictly monitor and manage their natural resources. Military bases vary in size and nature, and are heterogeneously distributed across the continental United States (See Figure @ref(fig:ewRoutesUsedHere)). The spread of these bases (Figure @ref(fig:basesOfInterestMap)), coupled with the top-down management of base-level natural resources presumably influences the inherent difficulties associated with collaborative management within and across military bases and other natural resource management groups (e.g., state management agencies, non-profit environmental groups).

Much like other actively managed landscapes, military bases are typically surrounded by non-managed lands, or lands not managed specifically for natural resource or ecological biodiversity or conservation. Natural resource managers of military bases face environmental pressures within and surrounding their properties, yet their primary objectives are very different. Natural resource managers of military bases, whose primary objective is to maintain military readiness, are especially concerned with if and how broad-scale external forcings might influence their lands. Prominent concerns include invasive species, wildlife disease, and federally protected species (personal communication with Department of Defense natural resource managers at Eglin Air Force and Fort Riley military bases). For these reasons, natural resource managers attempt to create buffers along their perimeters (e.g., live fire/ammunition suppression, wide fire breaks). Identifying the proximity of military bases to historic and modern ecological shifts may provide insight into the effectiveness of their natural resource management efforts.

The NABBS routes chosen for analyses in this Chapter lie within or near Fort Riley military base (located at approximately , ; Kansas, USA). Fort Riley (Figure @ref(fig:basesOfInterestMap)) is a useful reference site for this study. Woody encroachment of the Central Great Plains over the last century has triggered shifts in dominant vegetative cover and diversity [@ratajczak2018abrupt] in the area surrounding Fort Riley military base [@van2009causes]. This phenomena should present itself as a regime boundary if Fisher Information is a reliable spatial regime detection measure.

#### Spatial sampling grid

Fisher Information has been applied to empirical data as a spatial regime detection measure in recent years [@sundstrom2017detecting; @eason2019information]. The authors of @sundstrom2017detecting used the Fisher Information binning method to demonstrate the utility of this method as an indicator of spatial regime boundaries, suggesting that rapid changes in the resulting value of Fisher Information as calculated for spatially adjacent sites should indicate ‘regime changes’. @sundstrom2017detecting identified sampling sites which transected multiple ecoregions, resulting in a transect which zigzagged across a region of the Midwestern United States [@sundstrom2017detecting]. I identified sites using a gridded system across the continental United States and parts of Canada to ameliorate potential bias associated with handpicking NABBS routes. The gridded system comprises East-West running transects transects, ameliorating potential sampling bias as the transect location and widths were designed to capture large-scale shifts in bird communities at regular intervals. This spatial sampling grid approach also allows for raster stacking, or layering data layers (e.g., vegetation, LIDAR, weather), providing an opportunity to identify potential relationships with abiotic drivers, should regime shifts be observed in the avifauna data. This spatial sampling method also provides a simple vector for visualizing changes in the Fisher Information over space-time. For brevity, I present visual results of only three, spatially-adjacent, East-West running transects (Figure @ref(fig:ewRoutesUsedHere)) at multiple time periods.

### Calculating Fisher Information (FI)

Fisher Information, , was developed in 1922 by Ronald Fisher as a measure of the amount of information that an observable variable, X, reveals about an unknown parameter, . Fisher Information is a measure of indeterminacy (Fisher 1922) and is defined as, where is the probability density of obtaining the data in presence of . The Fisher Information measure (FIM) is used to calculate the covariance matrix associated with the likelihood, . Fisher Information is described as Extreme Physical Information [@frieden1995lagrangians; @frieden\_non-equilibrium\_2002], a measure that has been used to track the complexity of systems in many scientific disciplines including, physics, cancer research, electrical engineering, and, recently, complex systems theory and ecology

Fisher Information as gathered from observational data provides insight as to the dynamic order of a system, where an orderly system is one with constant (i.e., unchanging) observation points, and one whose nature is highly predictable. A disorderly system is just the opposite, where each next data point is statistically unpredictable. In ecological systems, patterns are assumed to be a realization of ecosystem order; therefore, one should expect orderliness in a system with relatively stable processes and feedbacks. Orderliness, however, does not necessarily infer long-term predictability. Equation @ref(eq:fiGeneral1922) is next adapted to estimate the dynamic order of an entire system, , as

where is the probability density for . Here, a relatively high Fisher Information value () infers higher dynamic order, whereas a lower value (approaching zero) infers less orderliness. To limit the potential values of in real data, we can calculate the amount of Fisher Information by re-expressing it in terms of a probability amplitude function [@mayer\_applications\_2007; @fath\_regime\_2003]:

A form specific to the probability density function of distance traveled by the entire system, which I call the ‘derivatives’ method, is defined as [@mayer\_applications\_2007, eq. 7.12]:

where T is the number of equally spaced time points over which the data are integrated. Numerical calculation of using the binning method (Eq. @ref(eq:fiAmp) and @ref(eq:fiDerivs)) each incorporate a moving-window procedure for calculating the probability of the system, , as being in one of an unidentified number of states (). Although previously applied to spatially-explicit terrestrial community data,the binning method requires multiple parameters to be defined *a priori*, which have been shown to influence inference based on the metric. I therefore calculated FI using the derivatives equation [see Chapter @ref(fiGuide)].

The binning procedure allows for a single point in time or space to be categorized into more than one state, which violating the properties of alternative stable states theory. The size of states [see @eason\_evaluating\_2012] measure is required to construct . In the case of high dimensional data, a univariate binning procedure of is not intuitive (i.e., reducing a multivariate system to a single probability distribution rather than constructing a multivariate probability distribution). Importantly, when using community or abundance data, rare or highly abundant species can influence the size of states criterion, thus influencing the assignment of each point into states. Finally, Eq. @ref(eq:fiAmp) assumes equal spacing (in space or time) between sampling points. Each of these violations can be avoided by using Eq. @ref(eq:fiDerivs) [@cabezas\_towards\_2002; @fath\_regime\_2003] to calculate the Fisher Information measure (see Chapters @ref(fiGuide), @ref(velocity) for detailed discussions on this topic). Briefly, derivatives method (Eq. @ref(eq:fiDerivs)) estimates the trajectory of the system’s state by calculating the integral of the ratio of the system’s acceleration and speed in state space [@fath\_regime\_2003]. Here, I use the derivatives method (Eq. @ref(eq:fiDerivs)) to calculate Fisher Information for all East-West transects (see Figure @ref(fig:ewRoutesUsedHere)) at decadal intervals (years 1980, 1990, 2000, and 2010). Justification for using this method is provided in detail in Chapter @ref(fiGuide).

### Interpreting and comparing Fisher Information across spatial transects

#### Interpreting Fisher Information values

Interpretation of FI, like the interpretation of numerous other regime detection measures, is currently a qualitative effort. Fisher Information is proposed as an indicator of system orderliness, where periods of relatively high values of FI indicate the system is in an “orderly” state, possibly fluctuating around a single attractor. A rapid change in FI is proposed as an indicator of a change in a system’s orderliness, suggesting a potential reorganization phase. Whether Fisher Information can identify a switch among basins of attraction within a single, stable state remains unknown, as does the number of states which a system can occupy. When a system occurs within any number of states equally, i.e., is equal for each state, both the derivative, (, and are zero. As , we infer the system is approaching a stable state, and as the system is showing no preference for a single stable state and is on an unpredictable trajectory. Eq. @ref(eq:fiAmp) bounds the potential values of Fisher Information at , whereas Eq. @ref(eq:fiGeneral1922), Eq. @ref(eq:fi73c), and Eq. @ref(eq:fiDerivs) are positively unbounded . If the Fisher Information is assumed to represent the probability of the system being observed in some state, , then the absolute value of the Fisher Information index is relative within a single datum (here a single datum is a spatial transect). It follows that Fisher Information should be interpreted relatively, but not absolutely.

Here I define a potential regime change as a point(s) having a non-zero derivative, and at which relatively large changes (manifested as either a sharp increase or decrease) in FI occurs. Regime shifts are identified as data changing from one state to another, thus, rapid shifts in the value of FI should indicate the locations of these shifts in the time *and* space, at which the system undergoes reorganization. Spatial and temporal Fisher Information calculation does not vary, but interpretation of either differ in that a spatial analysis will identify a spatial regime boundary [@sundstrom2017detecting] within a single time period, whereas temporal analysis identifies the point in time at which the system undergoes a regime shift. I follow published recommendations for interpreting the Fisher Information results in the context of identifying regime shifts [e.g., @karunanithi\_detection\_2008; @eason\_evaluating\_2012; @fath\_regime\_2003].

#### Interpolating results across spatial transects

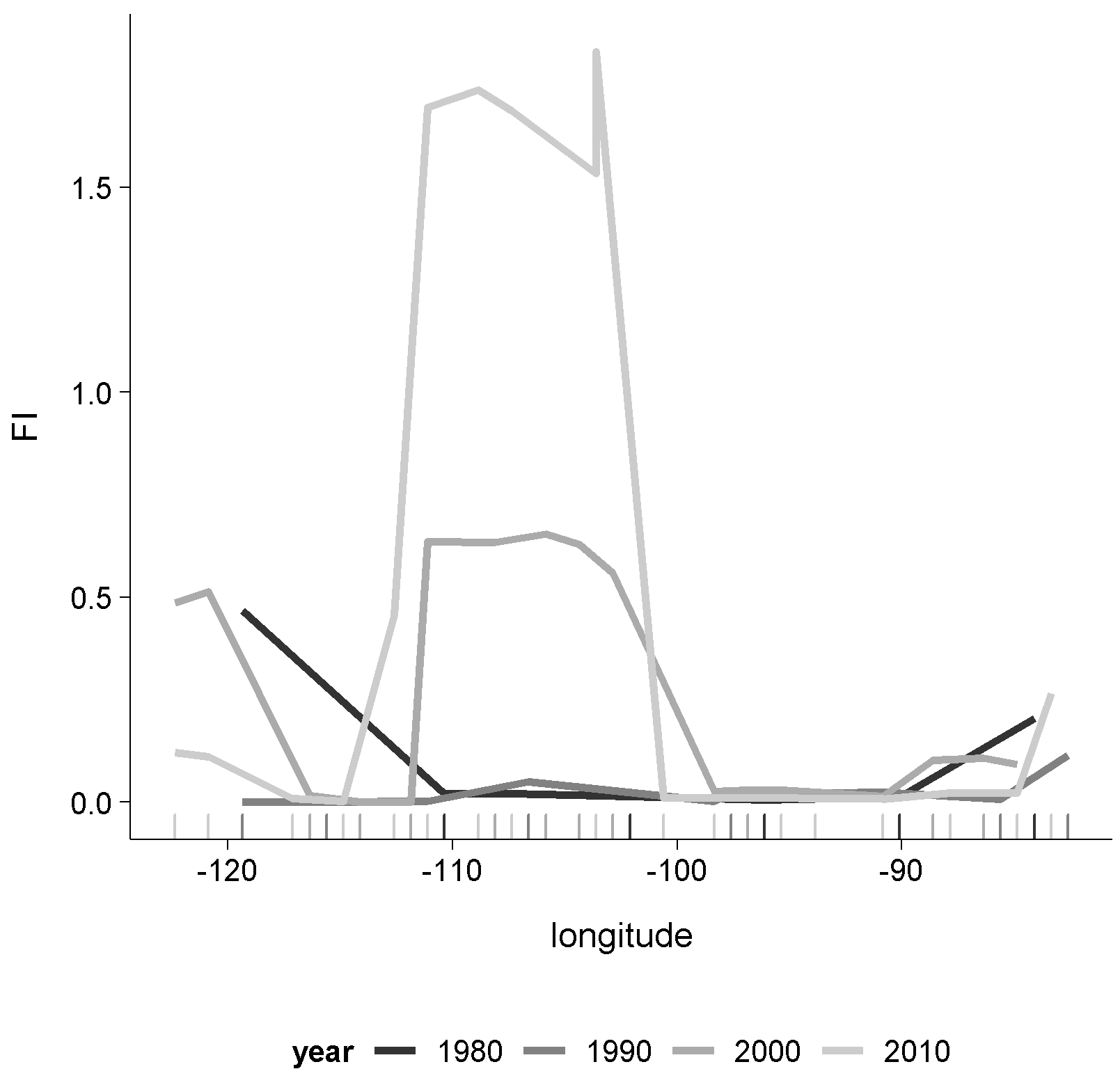
NABBS are not regularly spaced, and pairwise correlations of adjacent transects (see Figure @ref(fig:ewRoutesUsedHere)) is not possible without either (1) binning the Fisher Information calculations using a moving-window analysis, or (2) interpolating the results to regularly-spaced positions in space. To avoid potential biases associated with the former option (i.e. choosing window size, location of data aggregation), I linearly interpolated the calculated Fisher Information within each spatial transect to 50, evenly-spaced points along the longitudinal dimension. The 50 longitudinal points to which I interpolated were the same across each spatial transect, while latitude varied across transects. I used the function stats::approx() (with argument rule=1) to linearly approximate the Fisher Information. I did not interpolate values beyond the longitudinal range of the original data (i.e., no extrapolation).

#### Spatial correlation of Fisher Information

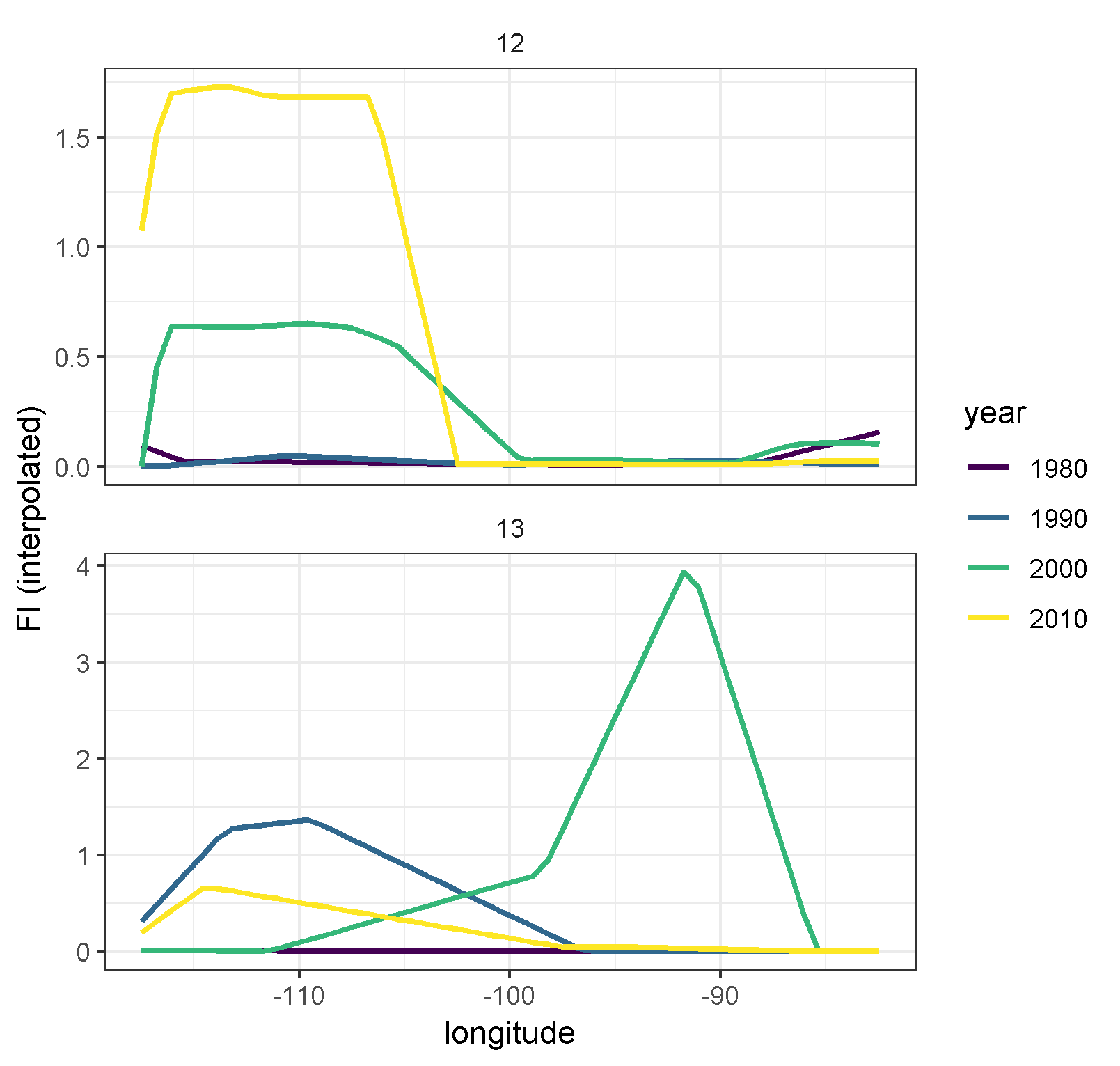
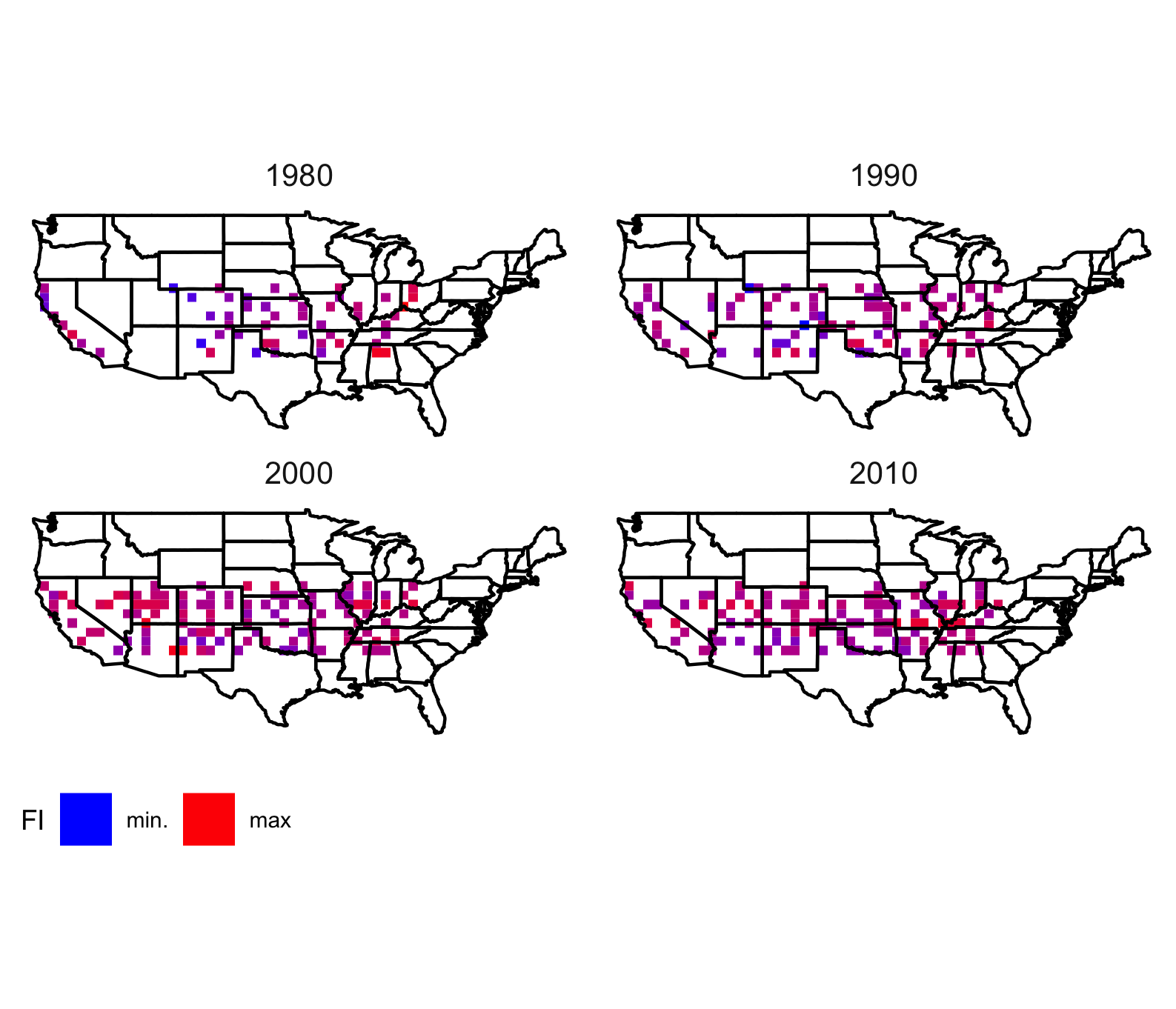
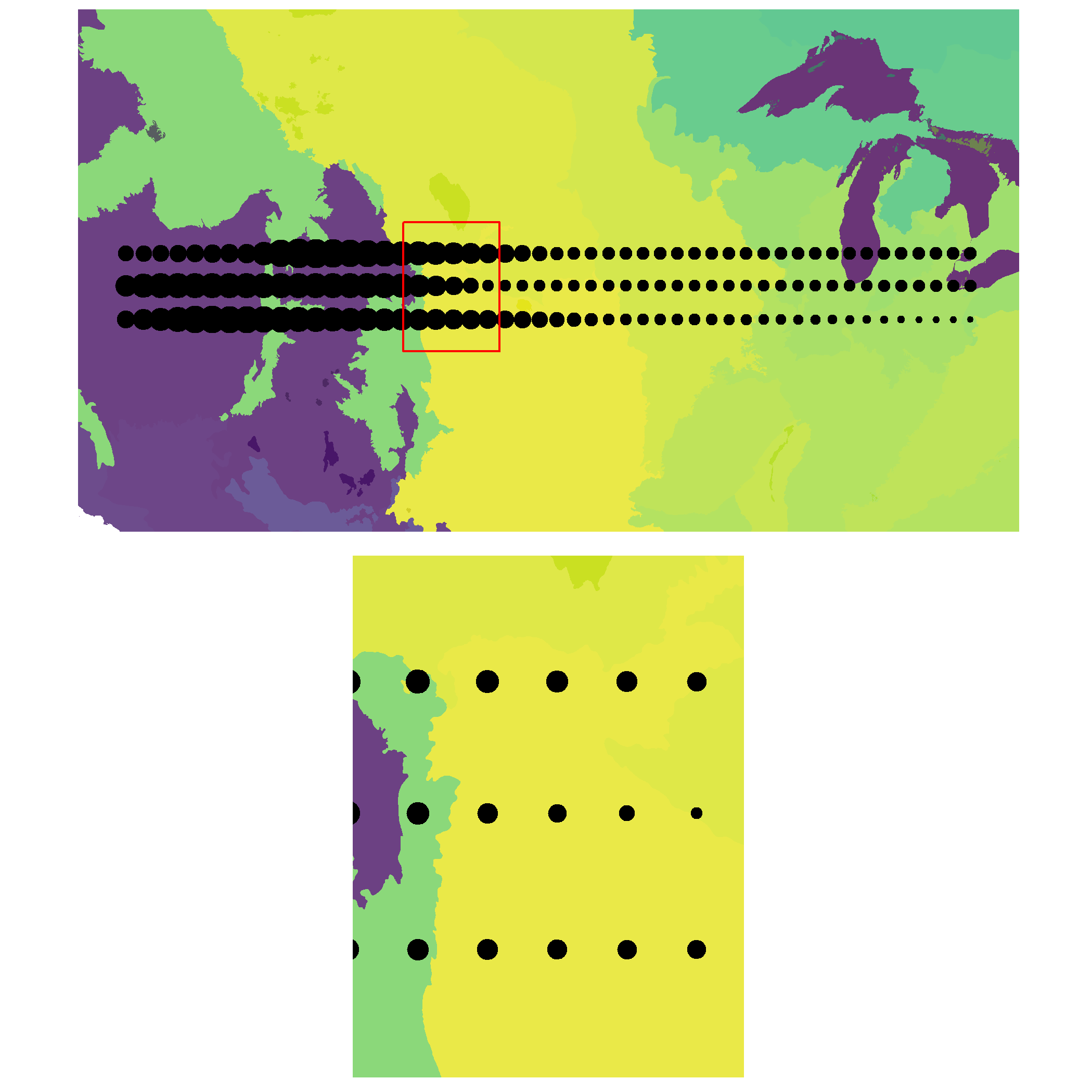
If Fisher Information captures and reduces information regarding abrupt changes in community structure across the landscape, then it follows that the values of Fisher Information should be spatially auto correlated. That is, the correlation of FI values should increase as the distance between points, both within and among transects, decreases. Further, direct comparison of FI across routes is not possible since FI (Eq. @ref(eq:fiDerivs)) is a relative value with no upper limit (i.e. can take on any value between and ). In other words, FI values calculated are **not** relatively comparable outside of a single spatial transect (Figure @ref(fig:ewRoutesUsedHere)). Fisher Information **is**, however, directly comparable within each spatial transect (e.g., @ref(fig:ewRoutesUsedHere)). For these reasons, we can identify spatial regime shifts both within and among spatial transects by using pairwise correlations among two transects (e.g., @ref(fig:ewRoutesUsedHere)) to determine whether values of FI are consistent across space. Here, I calculate the pairwise correlation (Pearson’s) among each pair of adjacent spatial transects (e.g., Figure @ref(fig:ewRoutesUsedHere)). I removed a pair of points if at least one point was missing an estimate for Fisher Information. This occurs when the original longitudinal range of one transect exceeded the range of the adjacent pair.

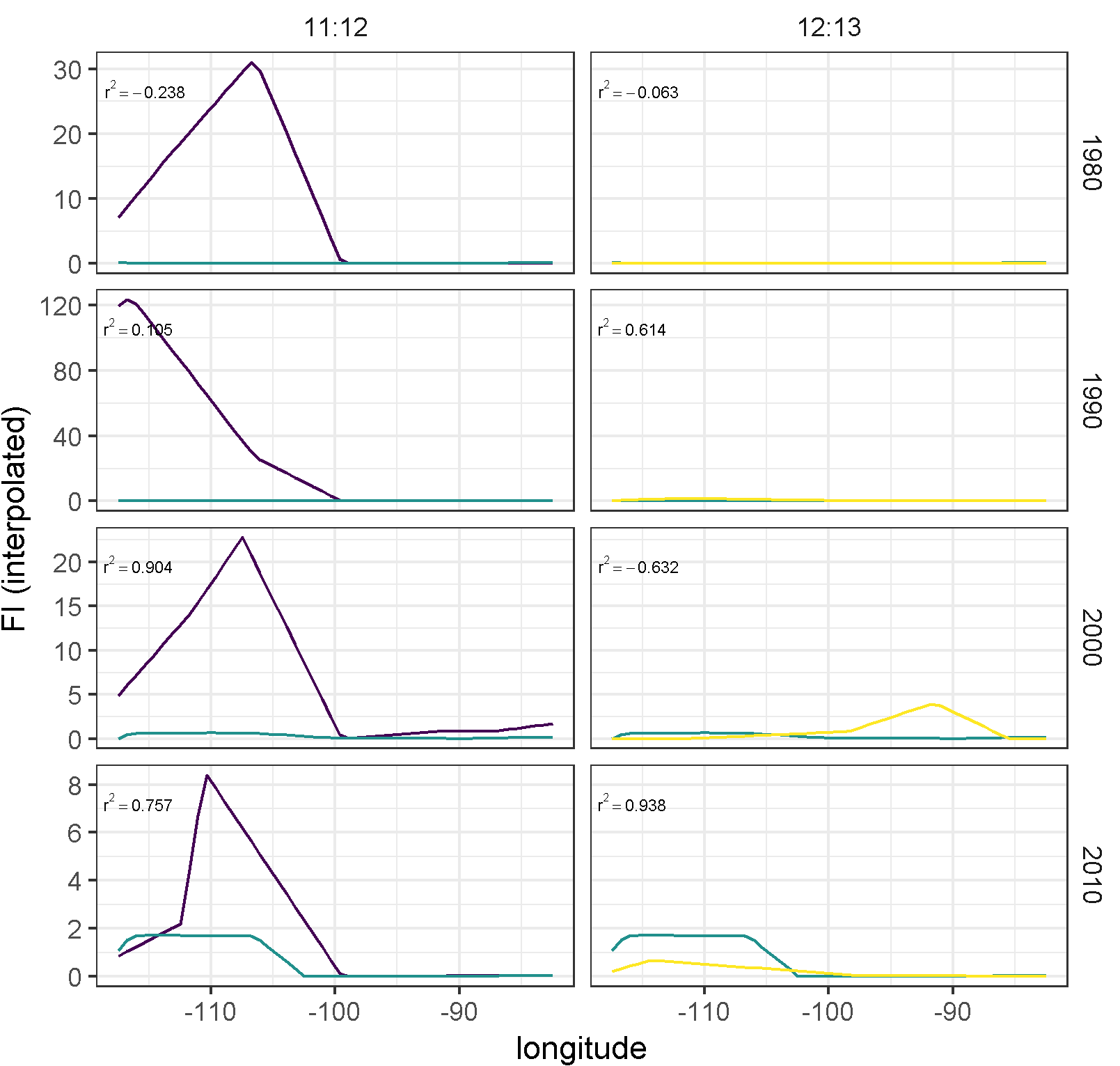
## Results

### Fisher Information across spatial transects

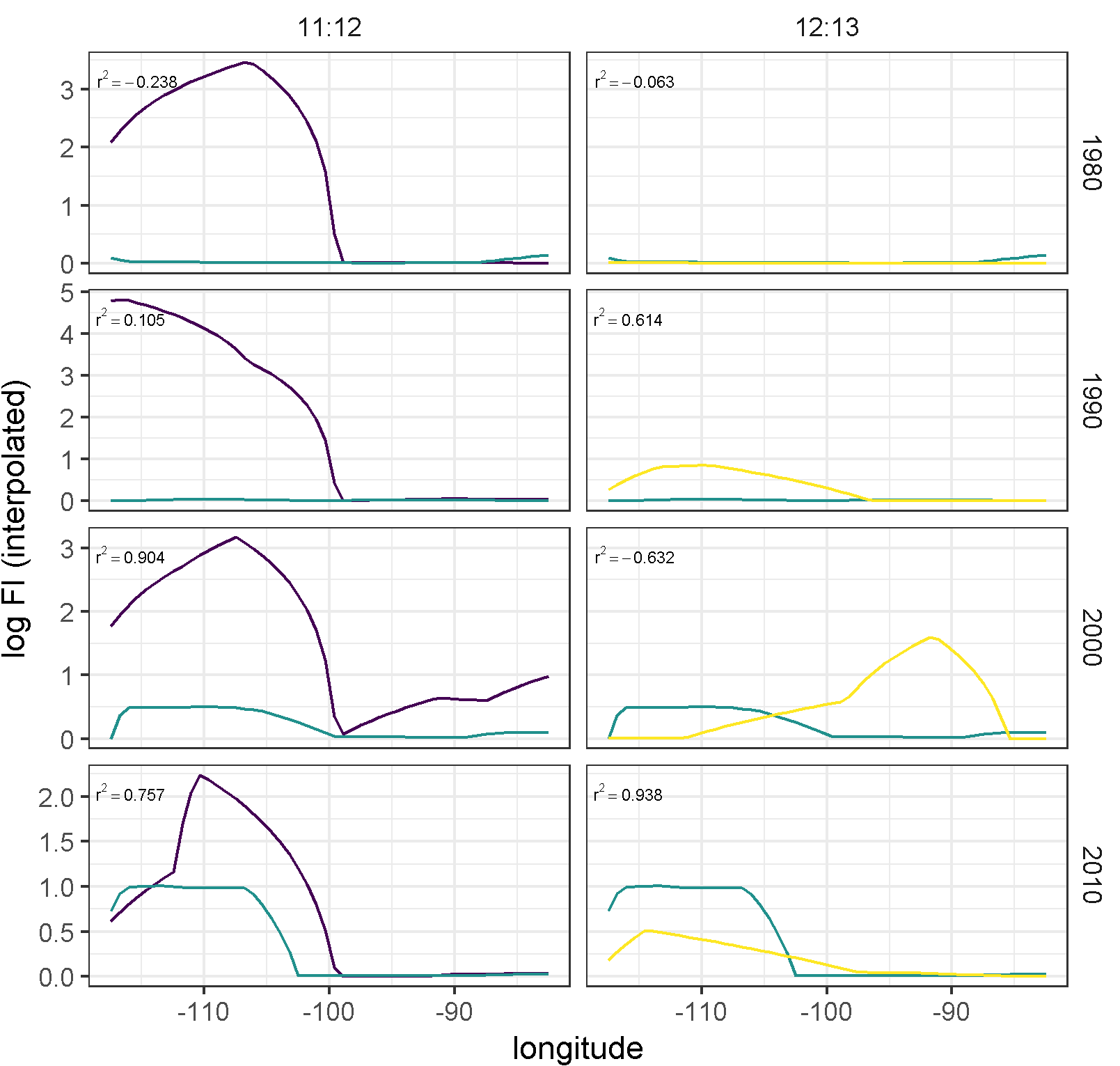


The change in the Fisher Information values along a single, East-West-running spatial transect (Transect number 12) over time.

As suggested earlier, rapid increases or decreases in FI are posited indicate a change in system orderliness, potentially suggesting the location of a regime shift. Using this method yields inconclusive results regarding the location of ‘spatial regimes’ (Figure @ref(fig:fi1Tsect)). Of the three spatial transects analyzed in this chapter (see Figure @ref(fig:ewRoutesUsedHere)), Figure @ref(fig:fi1Tsect) is representative of the lack of pattern observed in the Fisher Information values across all analyzed transects. I did not identify patterns of spatially contagious abrupt changes in the Fisher Information values within or among spatial transects.  \begin{landscape}  \end{landscape} 



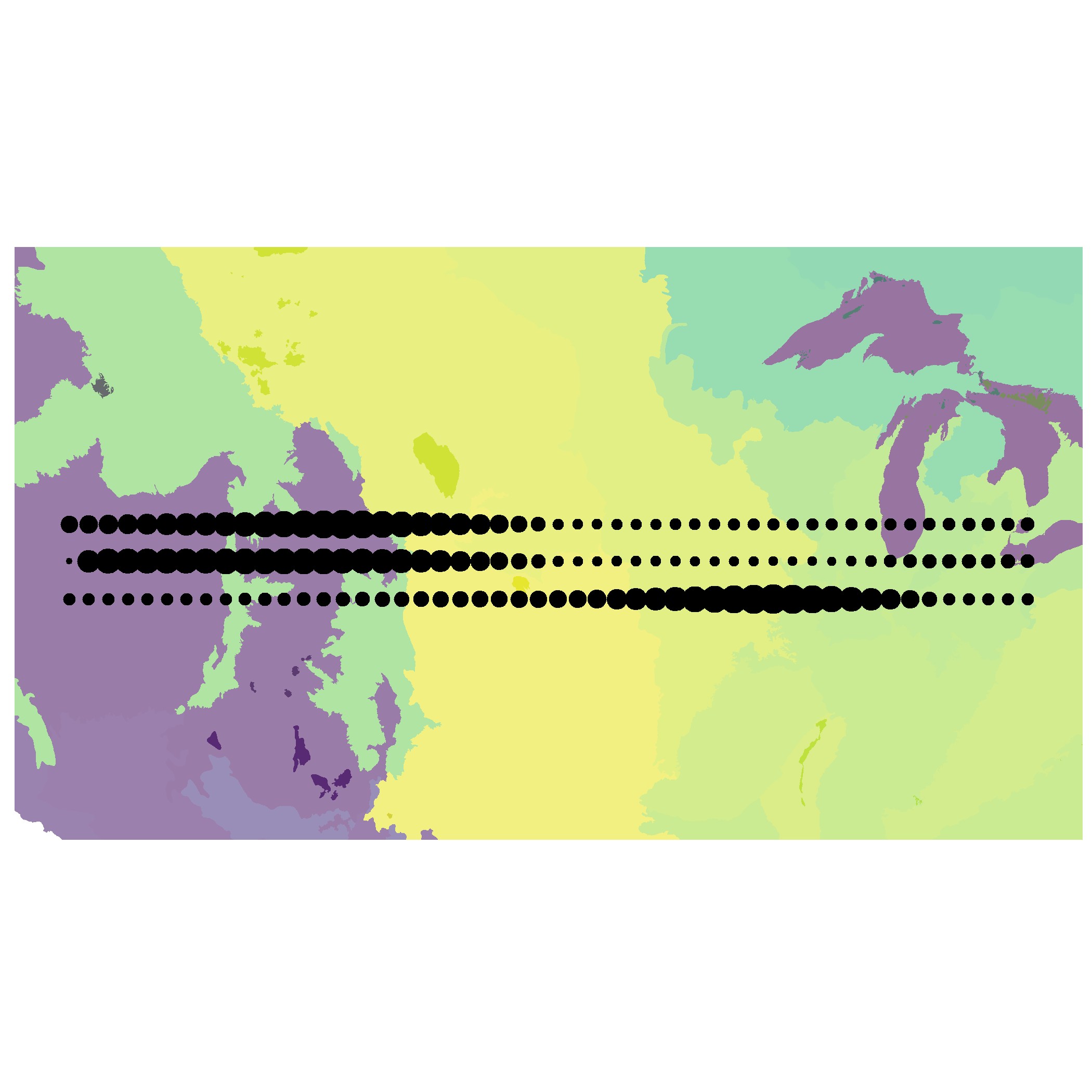
Pairwise relationships of Fisher Information (interpolated values) of spatially adjacent transects over time do not exhibit expected patterns of high postive correlation. Pairs were compared (column) at select sampling years (rows), and pair-wise correlations among paired transects are presented. Large, positive correlations indicate Fisher Information signals similarly at adjacent spatial transects.



Pairwise relationships of Fisher Information (interpolated values on the log-scale) of spatially adjacent transects over time. Pairs were compared (column) at select sampling years (rows), and pair-wise correlations among paired transects are presented. Large, positive correlations indicate Fisher Information signals similarly at adjacent spatial transects.

### Spatial correlation of Fisher Information

This study did not identify spatial correlation of the Fisher Information results among most of the spatially adjacent transects (Figures @ref(fig:corPlotTsectsInterp) and @ref(fig:corPlotTsectsInterpLog))). For spatially-adjacent transects (e.g, transects 11 and 12, or 12 and 13 in Figures @ref(fig:corPlotTsectsInterp) and @ref(fig:corPlotTsectsInterpLog)), we should expect high and positive correlation values, and these values should stay consistent across time *unless* the spatial transects were separated by an East-West running physical or functional boundary. This is not, however, what I expect in our East-West running transects (Figure @ref(fig:ewRoutesUsedHere)), as the spatial soft-boundaries limiting the distribution and functional potential of avian communities are largely North-South (Figure @ref(fig:ewRoutesUsedHere)). Note spatial transects in Figure @ref(fig:ewRoutesUsedHere) overlap multiple, large spatial ecoregion boundaries, such that we should expect our data to identify these points (boundaries).

Upon initial investigation, there are no consistent signs of broad-scale patterns in FI across space (Figure @ref(fig:fiEcoregion))[[1]](#footnote-44). If Fisher Information is an indicator of spatial regime boundaries, we should expect to see large changes in its value (in either direction) near the edges of functional spatial boundaries (e.g., at the boundaries of ecoregions). No clear regime changes appeared in areas where we might expect rapid changes (e.g., along the 105th meridian West, where a sharp change in altitude occurs). 

Numerical investigation of the spatial correlation among adjacent transects also yielded no clear patterns. I did not identify any obvious correlation with changes in FI values and functional potential (using Omernick Ecoregion Level 2; see Figure @ref(fig:fiEcoregion)). However, in years 2000 and 2010 the transects 11 and 12, and 12 and 13 were highly positively correlated (Figure @ref(fig:corPlotTsectsInterpLog)). Rather than abrupt changes in Fisher Information I found gradual changes (e.g., see results for years 2000 and 2010 in Figs. @ref(fig:fiEcoregion) and @ref(fig:fiEcoregion00).

## Discussion

The Fisher Information measure was introduced as a method to avoid analytical issues related to complex and noisy ecological data [@karunanithi\_detection\_2008; @fath\_regime\_2003] and was recently suggested as an indicator of spatial regimes [@sundstrom2017detecting; @eason2019information]. Contrary to expectations, I did not consistent abrupt changes in the Fisher Information metric (Eq. @ref(eq:fiDerivs)), which would indicate a regime shift of sorts, in the avian communities. Further, there was an absence of autocorrelation among the spatially adjacent transects in my study area, suggesting that the Fisher Information may not be a suitable metric for identifying abrupt changes in bird communities at this scale.

Although the Fisher Information equation (Eq. @ref(eq:fiDerivs)) used in this study is a relatively straightforward and fairly inexpensive computational calculation, extreme care should be taken when applying this index to empirical data. Fisher Information is capable of handling an infinite number of inputs (variables) and, given sufficiently low window size parameters, can technically calculate an index value for only two observations. It is important that the user understands the assumptions of identifying regime shifts or abrupt changes when using this method, as rigorous testing of its efficacy is necessary (but see Chapter @ref(resampling)). The sampling design of the North American Breeding Bird Survey data in this Chapter was designed to avoid subjective decisions present in a previous application [@sundstrom2017detecting].

There are three primary assumptions required when using Fisher Information to estimate relative orderliness within ecological data [@mayer\_applications\_2007]: (i) the order or state(s) () of the system is observable; (ii) any observable change in the information observed in the data represents reality and the variables used in the analyses will not produce false negatives; and, (iii) changes in presumed to be regime shifts do not represent the peaks of cyclic (periodic) patterns. Assumption (i) is one of philosophical debate and is thus not controllable. To attempt to control for false negatives or false positives that may result from violating assumption (ii), the user of this metric should take care in their selection of state variables. In the the case of a high dimensional data, relativization of state variables and/or a state variable reduction technique may be useful. However, Fisher Information does not convey information on how specific variables relate to the calculated index. Finally, we can take measures to account for cyclic behavior [assumption (iii)] in the data by ensuring integration periods capture at one full cycle of the system and, given sufficiently high number of observations, increasing the integration period may also alleviate some issues related to irreducible error, or white noise.

The lack of patterns identified using Fisher Information may be influenced by a mismatch among the ecologically relevant scales and the temporal resolution and extent of our data may influence the ability of this index to capture large-scale changes in whole bird communities. Aside from the typical biases associated with the BBS data (e.g., species detection probability, observer bias), there are additional considerations to be made when using these data to identify ‘spatial regime shifts’. Breeding Bird Survey routes are spaced apart so as to reduce the probability of observing the same individuals, but birds which fly (especially in large flocks) overhead to foraging or roosting sites have a higher probability of being detected on multiple routes. We have, however, removed these species (waders, shorebirds, waterfowl, herons) from analysis. Regardless, this study assumes there is potential for each unique BBS route to represent its own state. If routes were closer together, it is more probable that the same type and number of species would be identified on adjacent routes. Therefore, if this method does not detect slight changes in nearby routes which occupy the same ‘regime’, then it follows that the method is sensitive to loss or inclusion of new species, which are spatially bounded by geological and vegetative characteristics. What new information does this give us about the system? Fisher Information reduces and removes the dimensionality of these systems, which may omit information or signals integral to understanding the ecological processes at play.

Effective regime detection measures should provide sufficient evidence of the drivers and/or pressures associated with the identified regime shifts [@mac2014scrutiny]. The Fisher Information index, while collapsing a wealth of data into a single metric, does not allow the user to relate the resulting value to the original data, unlike other dimension reduction techniques. For example, the loadings, or the relative influence of variables on the ordinate axes, can be derived from a Principal Components Analysis–this cannot be achieved using Fisher Information. If Fisher Information clearly suggested a spatial regime boundary or shift, a before-and-after post-hoc analysis of the regional community dynamics might confirm the regime shift occurrence.

A rapid change in either direction (increase or decrease) of the Fisher Information value is proposed as an indicator of ecological regime shifts, or a change in the orderliness of a system [@mayer\_fisher\_2006; @eason\_evaluating\_2012]. After calculating the Fisher Information for each spatial transect (Figure @ref(fig:ewRoutesUsedHere)) during each sampling year in this study, I used pairwise correlation to determine whether spatial autocorrelation existed among pairs of spatial transects. If some set of points are close in space and are *not* separated by some physical or functional boundary (e.g., an ecotone, high altitude rock formations), then the Fisher Information calculate should exhibit a relatively high degree of spatial autocorrelation that is consistent over time. It follows that the correlation coefficient of spatially adjacent transects should be similar, diverging only as the distance between the transects differs and/or a functional or physical boundary separates them. Contrary to these expectations, I did not find evidence of such abrupt changes within nor across the East-West running spatial transects. Several questions remain regarding the application of regime shift detection methods to spatially-explicit data. If signals of regime shifts do exist, the results of this study suggest the Fisher Information metric may not be ideal for identifying them. This study provided an objective evaluation of the Fisher Information metric as a spatial regime detection measure. Future work on the following areas may improve our understanding of if and how Fisher Information may provide insights of ecological regime shifts in spatial and/or temporal data:

1. Sensitivity of Fisher Information to data quality and quantity (this is explored in Chapter @ref(resampling)).
2. What, if any, advantages does FI have over other density estimation techniques?
3. Does FI provide signals in addition to or different than geophysical and vegetative (e.g. LIDAR) observations (data)?
4. Relationship of Fisher Information to likelihood ratio-based unsupervised change-point detection algorithms [e.g., ChangeFinder; @liu2013change].
5. How does Fisher Information perform relative to other regime detection measures (see Chapter @ref(resampling))?

1. Here, shape size indicates the relative value of the scaled and centered Fisher Information results. Red box (top panel) indicates the extent of the results presented in the bottom panel. [↑](#footnote-ref-44)