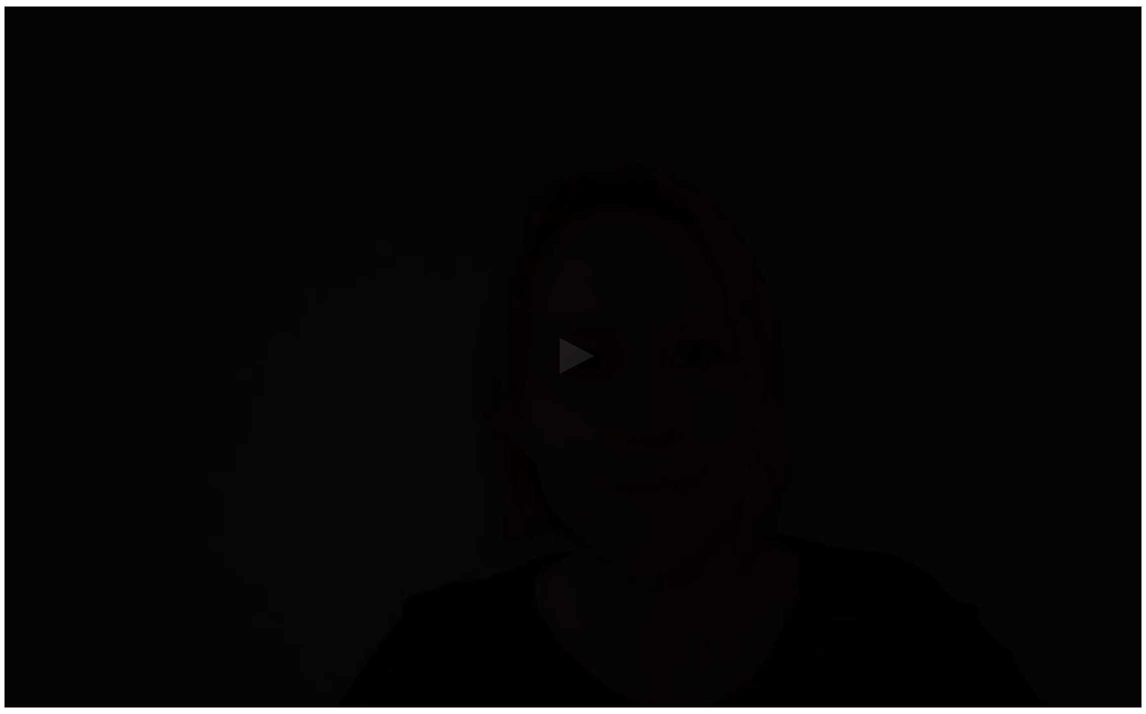
7. Dipole Discovery

Bookmark this page

Exercises due May 21, 2021 19:59 EDT

Dipole Discovery



outcome that you have.

So in summary, we saw that environmental data,

this space-time data, can be viewed

from different perspectives.

We can do simulations.

We can combine them with statistics.

We can model short range connections specifically

for prediction, for example, for sensor selection.

And we can study long range connections, for example,

from the language of networks.

14:19 / 14:19 1.50x

Video

Download video file

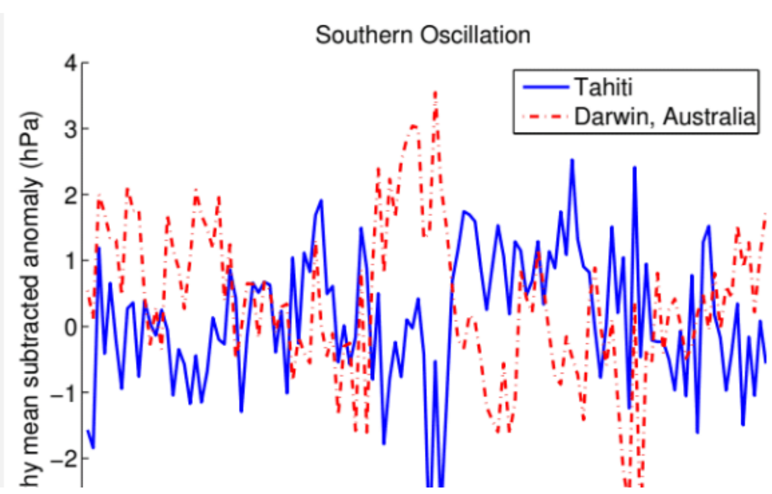
Transcripts

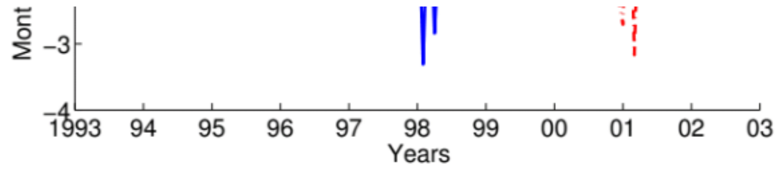
Download SubRip (.srt) file

Download Text (.txt) file

Previously, we have observed that nearby nodes tend to be strongly and positively correlated with each other. Based on our understanding of the environment, these correlations are to be expected. What may not be expected is the presence of negative correlations between the nodes. Such correlations can be evidence of interesting phenomena that warrant further study.

For example, see the figure below, which shows the atmospheric pressure in two widely separated regions on the globe as a function of time.





51: Pressure as a function of time for Tahiti and Darwin.

Note the pressure for both locations in 1995, 1998, and 2000. There is a clear pattern of anti-correlation, which leads to a negative covariance between the pressure variables for each location.

We may wish to know how spatially extended this anti-correlation is. The figure below shows the temperature as a function of position on the globe for two different years.

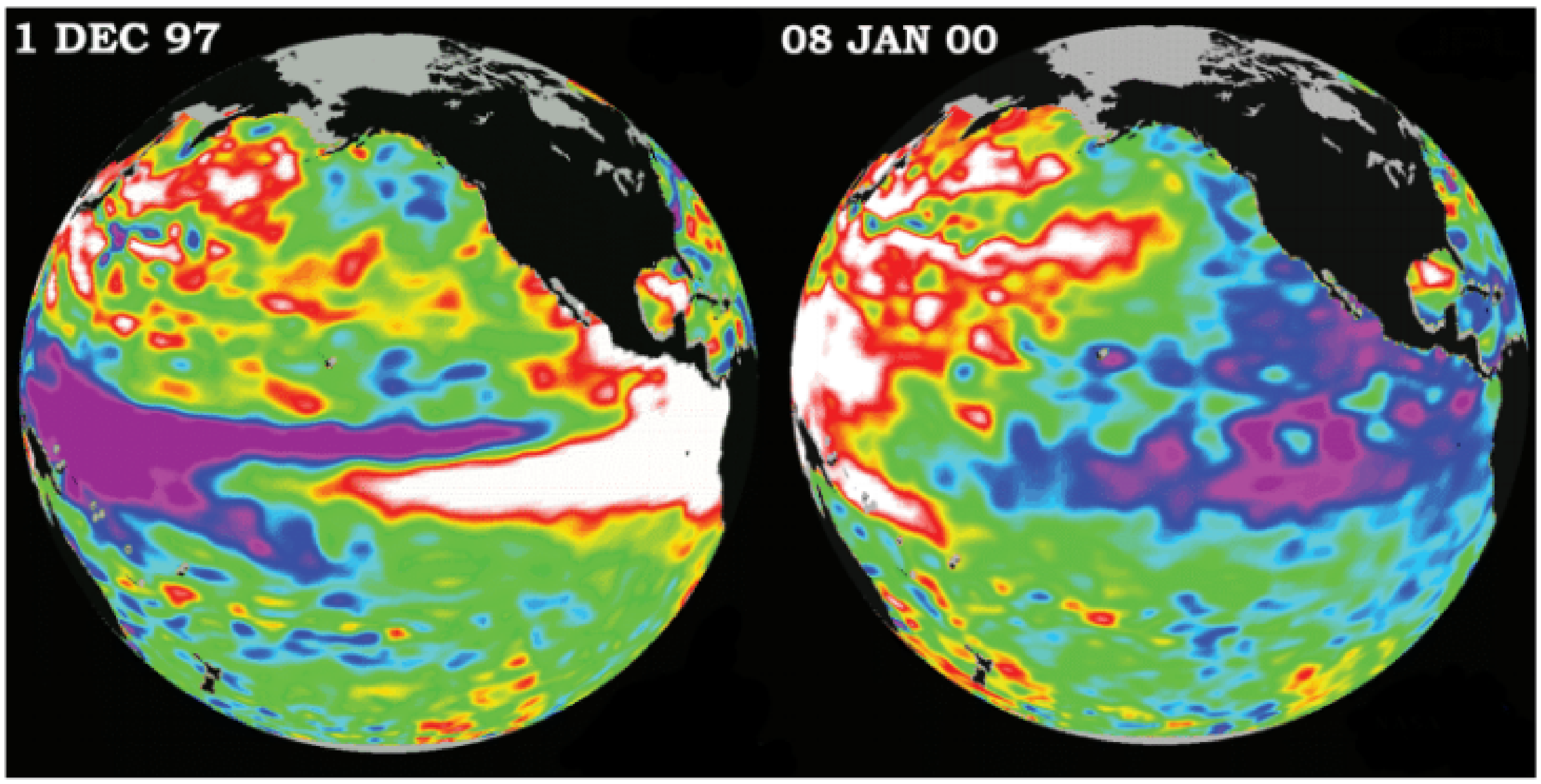


Figure 16.1. *SST anomalies (°C) observed under El Niño conditions (December, 1997; left) and La Niña conditions (January, 2000; right). Reprinted courtesy of NASA/JPL-Caltech.*

52: Temperature across the globe on two days in different years. White-red areas show above average temperatures, blue-purple areas are below average.

On the left hand image, the anti-correlation is evident along the equatorial region. Then, on the right, approximately two years later, the temperatures are reversed in both regions.

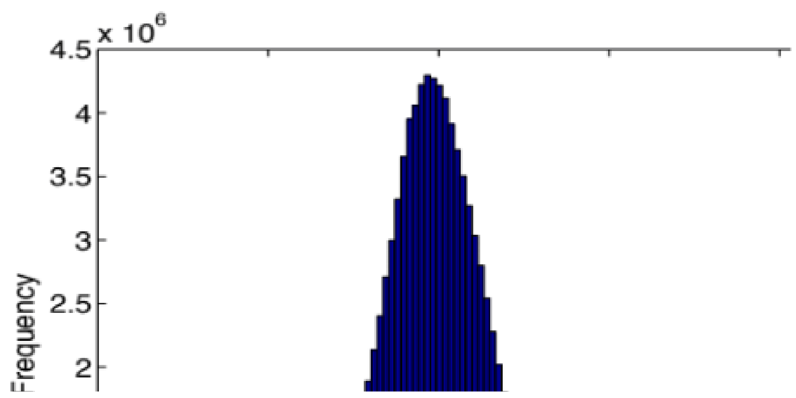
Clearly this is a global phenomena that would have wide ranging impact. It is known as the El Niño–Southern Oscillation, and countries in the region that are dependent on agriculture can be greatly affected by the extreme weather events induced by these changes in climate. This is an example of a dipole.

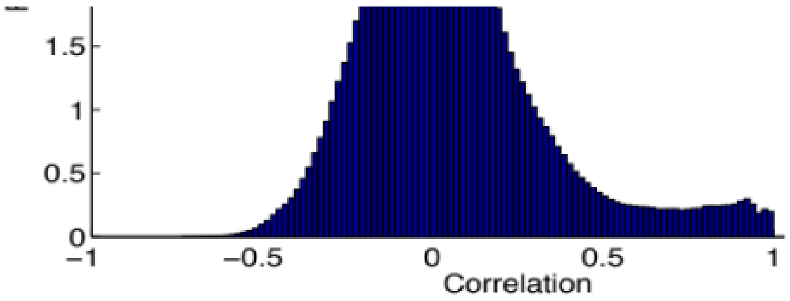
Definition 7.1 A **Dipole** is a quantity and an opposing quantity that are separated by a distance.

The name dipole comes from electric and magnetic dipoles in physics. You can imagine a dipole as a magnet, with a north and south end located a distance apart. Here, the region of above average temperature corresponds to the north end of the magnet, while the region of below average temperature corresponds to the south end. As such, we can see that we will have a dipole any time we have an anti-correlation between two regions separated by a distance.

Automatic discovery of dipoles

The potential ramifications of any dipole-like phenomena behooves us to find these dipoles in our data. The correspondence between dipoles and anti-correlations suggests that we should first start by examining the distribution of correlations:





53: Distribution of the correlation coefficient, ρ, for each node in the graph.

This distribution is shown in the above figure. We can see that if we were to set a threshold at $|\rho| > 0.5$, then we would pick up a large number of positive correlations but very few negative correlations.

Positive correlation tail

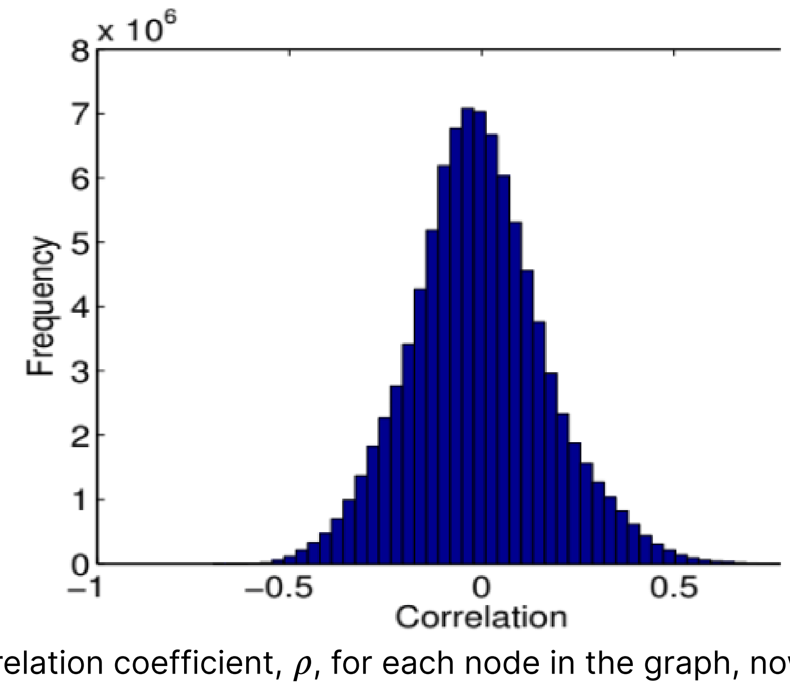
1 point possible (graded)

Why is there a heavy tail of positive correlations?

- In environmental data, correlations can be found all across the globe.
- Nodes are often correlated to their neighbors in environmental data.
- A tail of negative correlations should exist, but is not present in this data.

Submit You have used 0 of 2 attempts

Based on our reasoning for the origin of this positive correlation tail, we can remove it by only considering the correlation coefficient between nodes that are located far apart. The effect of putting this requirement on the correlation distribution is shown in the figure below.



54: Distribution of the correlation coefficient, ρ, for each node in the graph, now with a requirement that the distance between nodes is greater than 5000 km.

Now we can place a threshold at $|\rho| > 0.5$ and get roughly equal numbers of positive and negative correlations. We can also conclude from the above figure that the negative correlations are mostly long range, as this region of the correlation distribution was not changed much by imposing the distance requirement. Thus, we could, as an alternative to placing a distance requirement, simply use different thresholds for the negative and positive correlations, such as $\rho < -0.4$ or $\rho > 0.85$.

Identifying a negative correlation

1 point possible (graded)

Suppose we impose these thresholds and find two nodes that are negatively correlated. Can we conclude that we have found a dipole?

- No, the correlation could have arisen by chance.

- Yes, any strong negative correlation must be real.

Submit You have used 0 of 1 attempt

We can begin identifying a dipole by starting with the most negative correlation in the graph, that belongs to the pair of nodes (a, b). We can strengthen our identification of a dipole by examining the region around the ends of this potential dipole. Call these ends pole A and pole B. We expect from our dipole that:

- Pole A should consist of nodes that are positively correlated to each other.
- Pole A should also consist of nodes that are negatively correlated to nodes in pole B.
- Pole B should consist of nodes that are positively correlated to each other.
- Pole B should also consist of nodes that are negatively correlated to nodes in pole A.

We can define the sets

$$\begin{aligned} \mathcal{A}_+ &= \{v \in V : \rho(v, a) > \rho_+\} \\ \mathcal{A}_- &= \{v \in V : \rho(v, a) < \rho_-\} \end{aligned}$$

where $\rho(v, a)$ is the correlation coefficient between nodes v and a, and V is the set of nodes in the graph. These sets are the sets of nodes that meet a positive correlation threshold ρ_+ with a, and a negative correlation threshold ρ_- with a. The negative threshold, ρ_- should be no stricter than the threshold used to identify the pair (a, b) to begin with, lest we be left with an empty set.

Similarly we define the sets

$$\begin{aligned} \mathcal{B}_+ &= \{v \in V : \rho(v, b) > \rho_+\} \\ \mathcal{B}_- &= \{v \in V : \rho(v, b) < \rho_-\} \end{aligned}$$

as those nodes (anti-)correlated with b.

We can then define pole A as $A = \mathcal{A}_+ \cap \mathcal{B}_-$ and pole B as $B = \mathcal{B}_+ \cap \mathcal{A}_-$.

Dipole sets

1 point possible (graded)

Assuming there is a pair of nodes with a strong negative correlation, such that the above procedure may be used, will this procedure always result in a dipole? Here we define distinct regions as groups of nodes that are separated by distance.

- Yes, we will always get two distinct regions thus forming a dipole between them.
- No, we may end up with more than two distinct regions.
- No, we may end up with zero regions: the intersections of the sets may be empty.

Submit You have used 0 of 2 attempts

Time variance of dipoles

Once we have discovered a dipole, we may wish to examine how it evolves with time.

Suppose we have 50 years of data. We could divide this data into 10 sets of 5 years, then identify the dipoles in each. However, depending on the measurement interval, 5 years may be too little data to accurately estimate the correlation coefficient between nodes.

Instead, we can create overlapping sets. For example, we could create seven sets of 20 years, each offset by 5 years from each other. This increases the amount of data from which the correlation coefficients are derived, thus improving the estimates. When we do this we need to be careful of a few points:

- We should expect some similarity between overlapping sets, as they share data. Thus, observed similarity does not necessarily mean the dipole is stable with time.
- We should not use the presence of a dipole in two overlapping sets to justify the presence of the dipole generally, this would be double-counting our data.

Discussion

Hide Discussion

Topic: Module 5: Environmental Data and Gaussian Processes:Sensing and Analyzing global patterns of dependence / 7. Dipole Discovery

Add a Post

Show all posts by recent activity

Thank you!
I'm looking forward to the next written paper but feeling a little sad to have finished the last set of exercises. Wow, what a course! Th... 3

Previous Next

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate