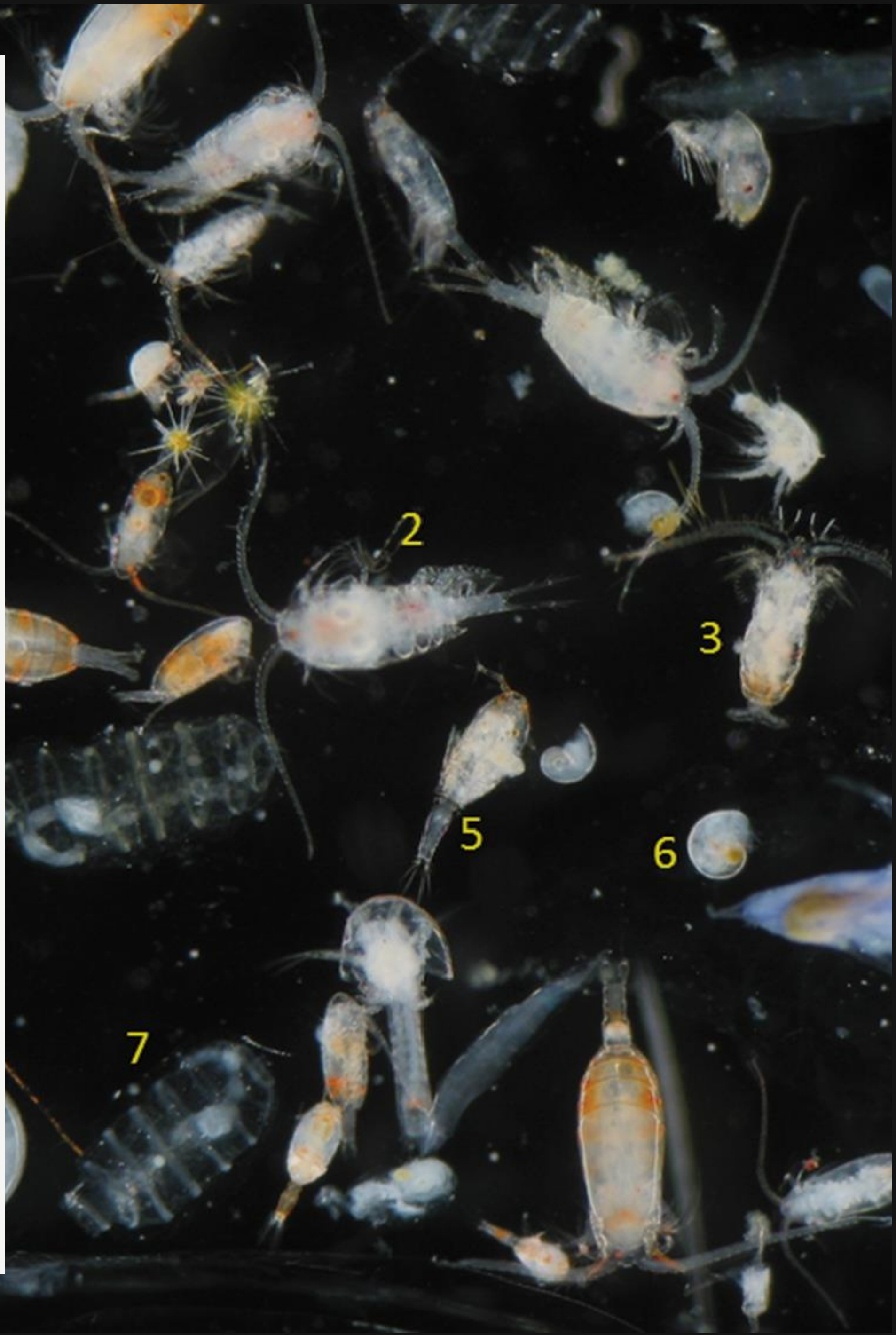
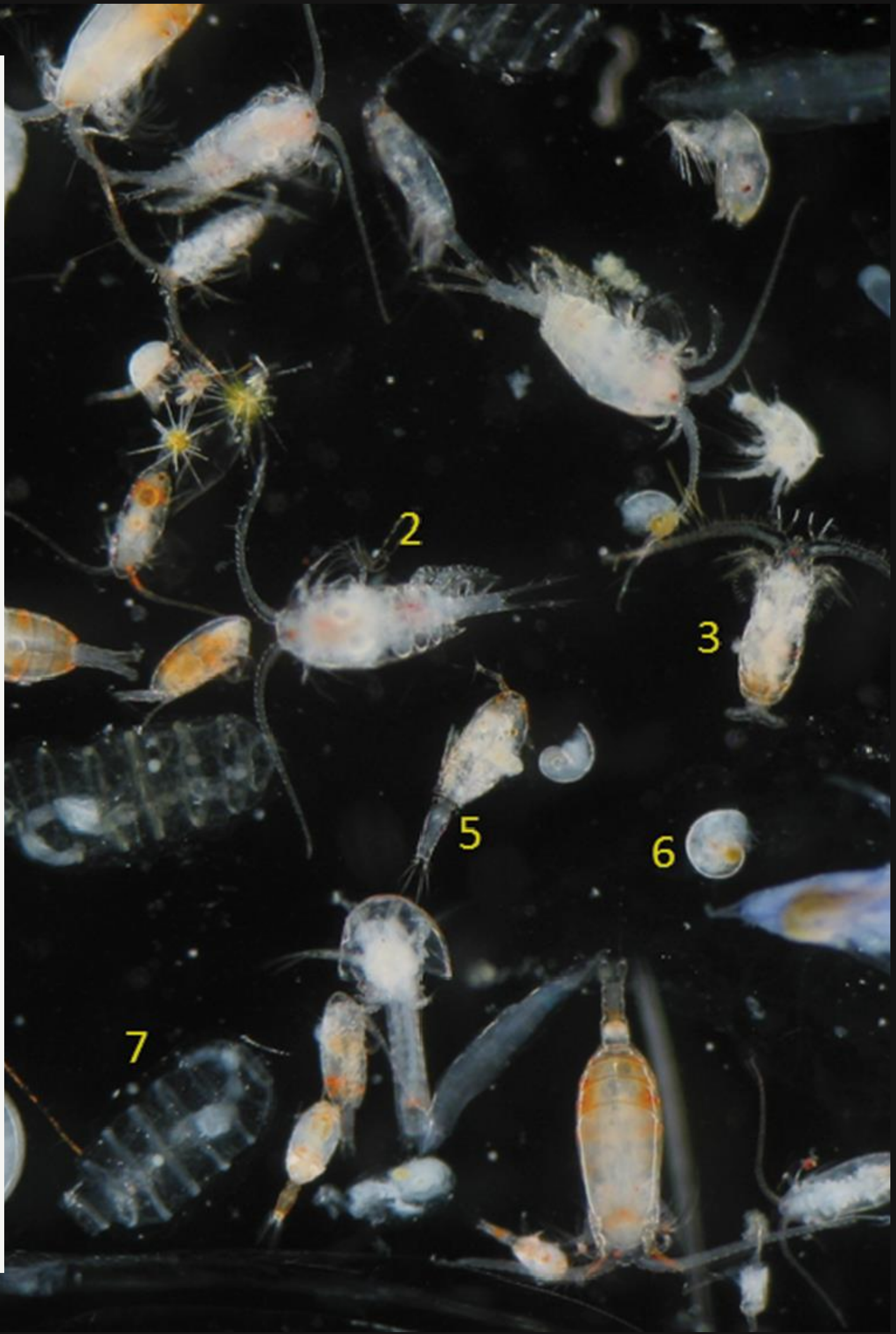
FW 599 Special Topics: Multivariate Analysis of Ecological Data in R

Lecture 13: Spatial and Temporal Data Series

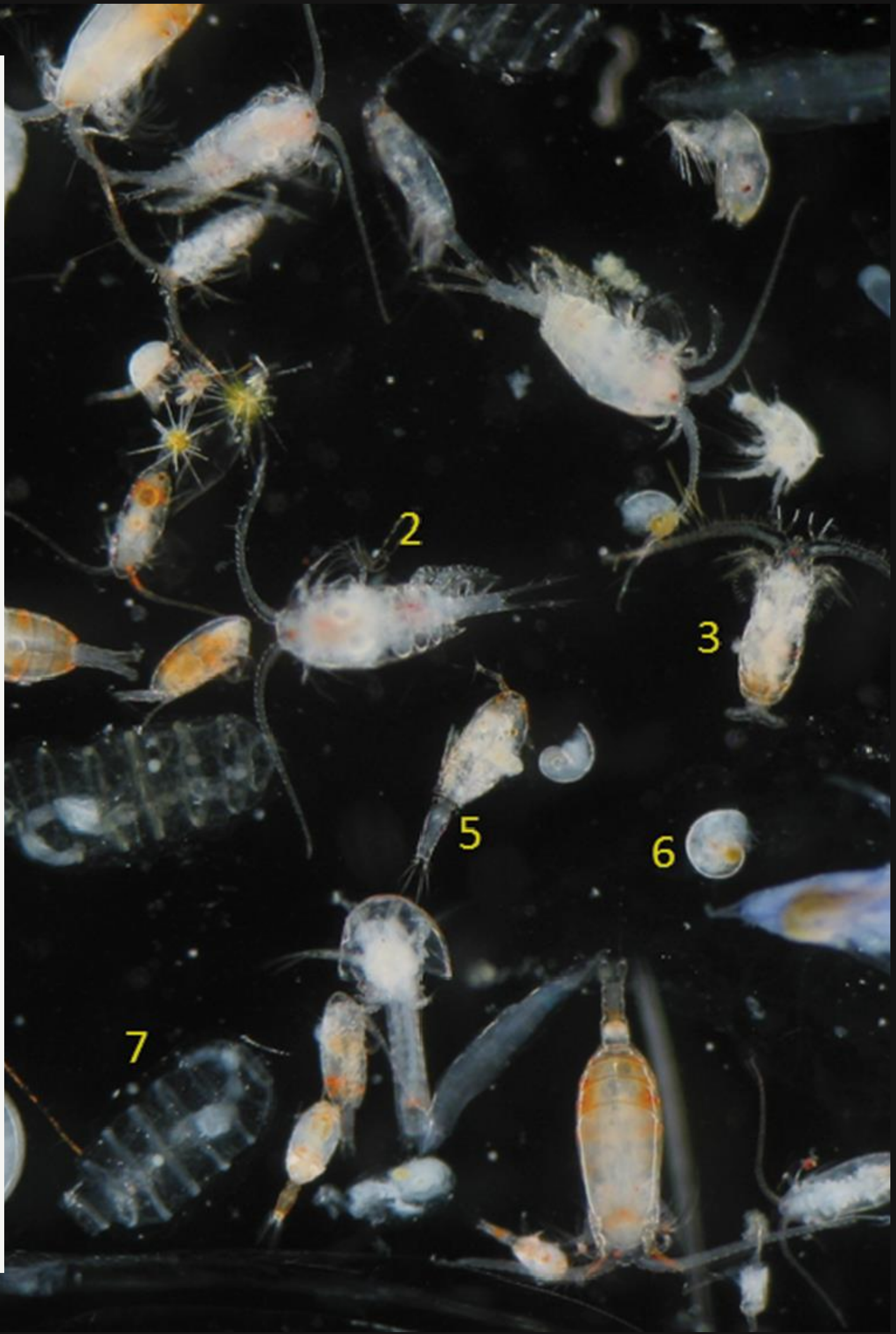
Thursday, November 14, 2024



Lecture 13: Spatial and Temporal Data Series

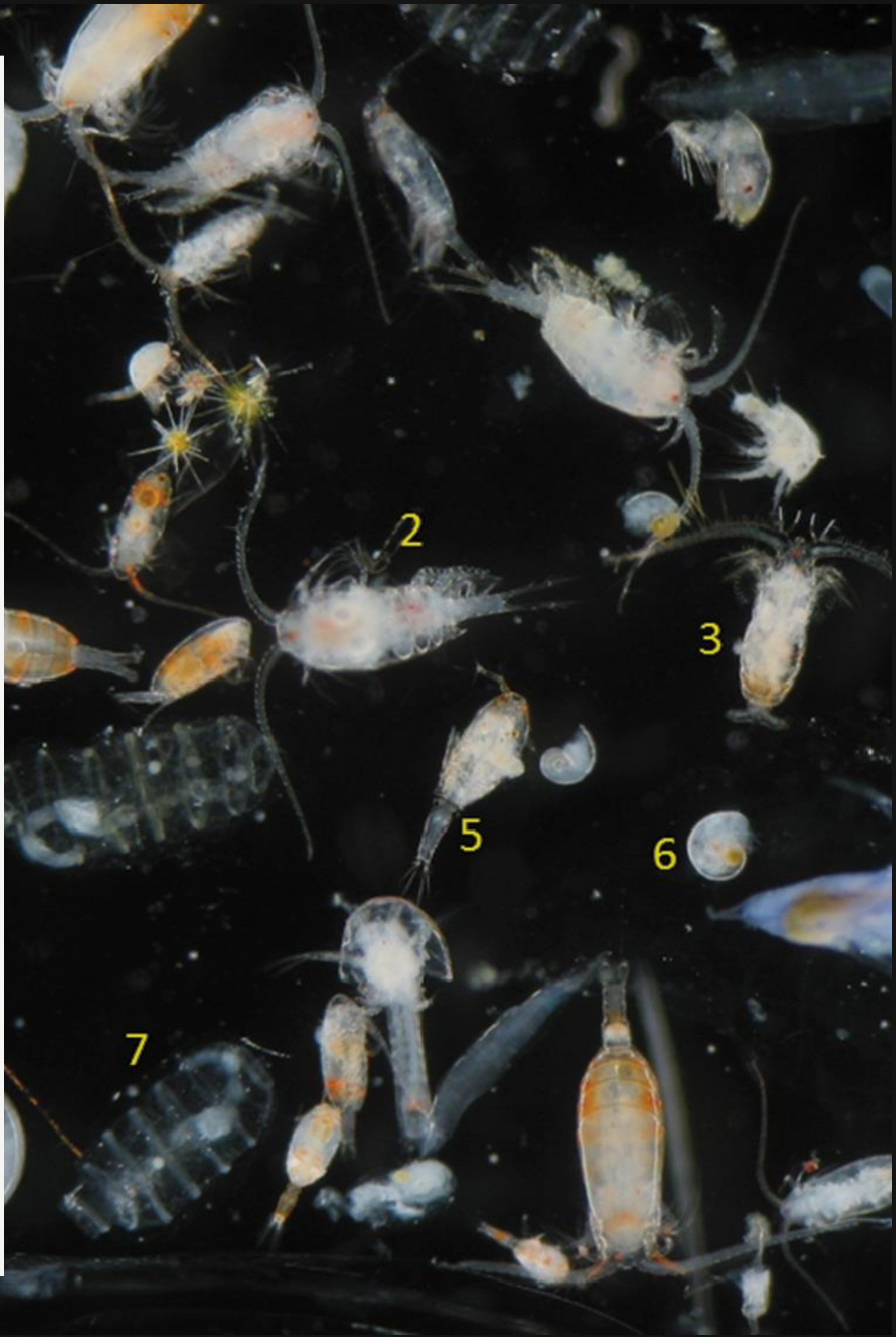


Ecological Data Series



Ecological Data Series

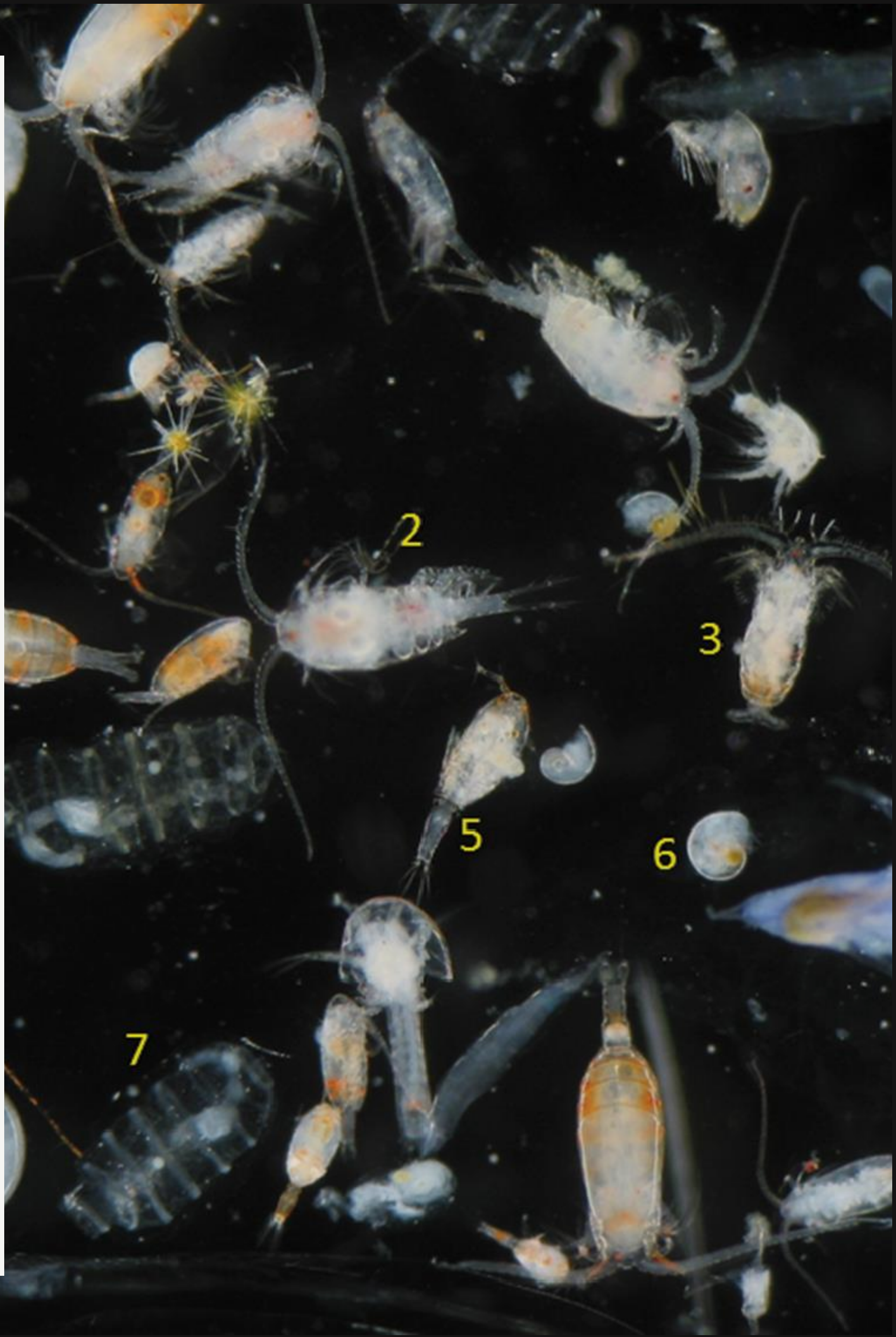
Ecological **data series** are continuous or discrete variables sampled over time or along transects in space.



Ecological Data Series

Ecological **data series** are continuous or discrete variables sampled over time or along transects in space.

i.e., *ordered* along a temporal or spatial axis.



Ecological Data Series

Ecological **data series** are continuous or discrete variables sampled over time or along transects in space.

i.e., *ordered* along a temporal or spatial axis.

Ecological Data Series

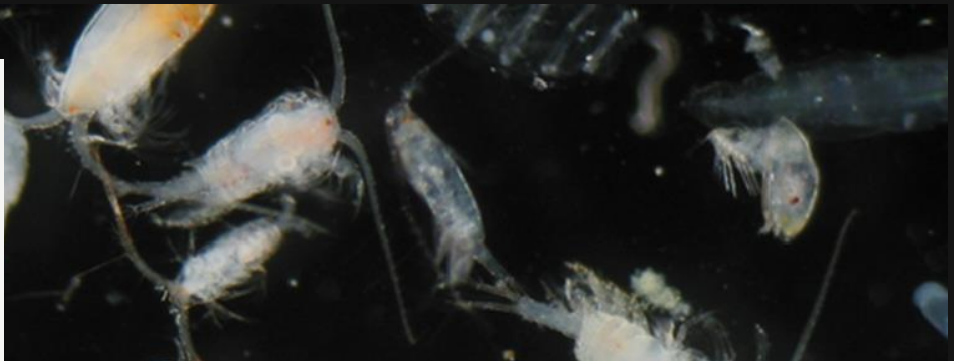
Ecological **data series** are continuous or discrete variables sampled over time or along transects in space.

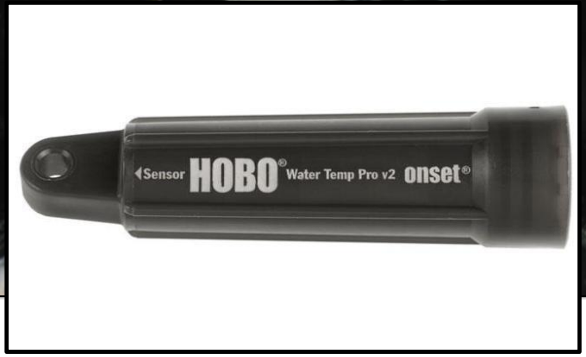
i.e., *ordered* along a temporal or spatial axis.

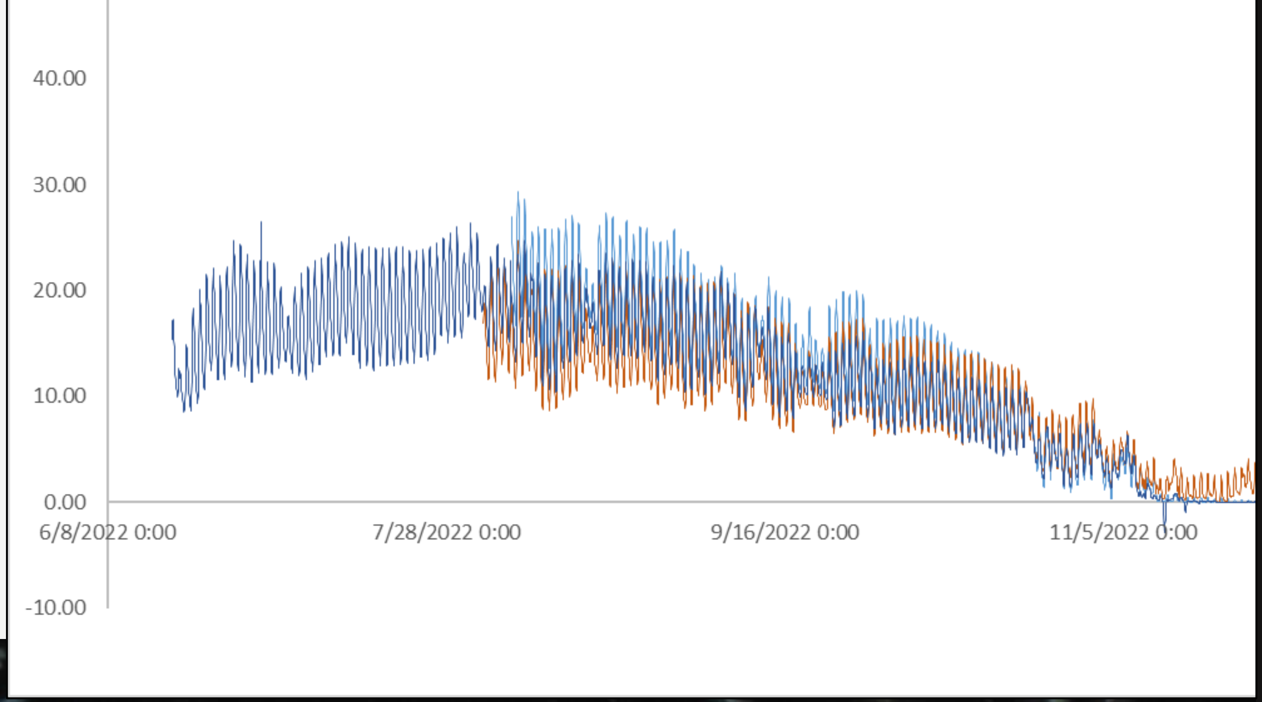
Ecological Data Series

Ecological **data series** are continuous or discrete variables sampled over time or along transects in space.

i.e., *ordered* along a temporal or spatial axis.

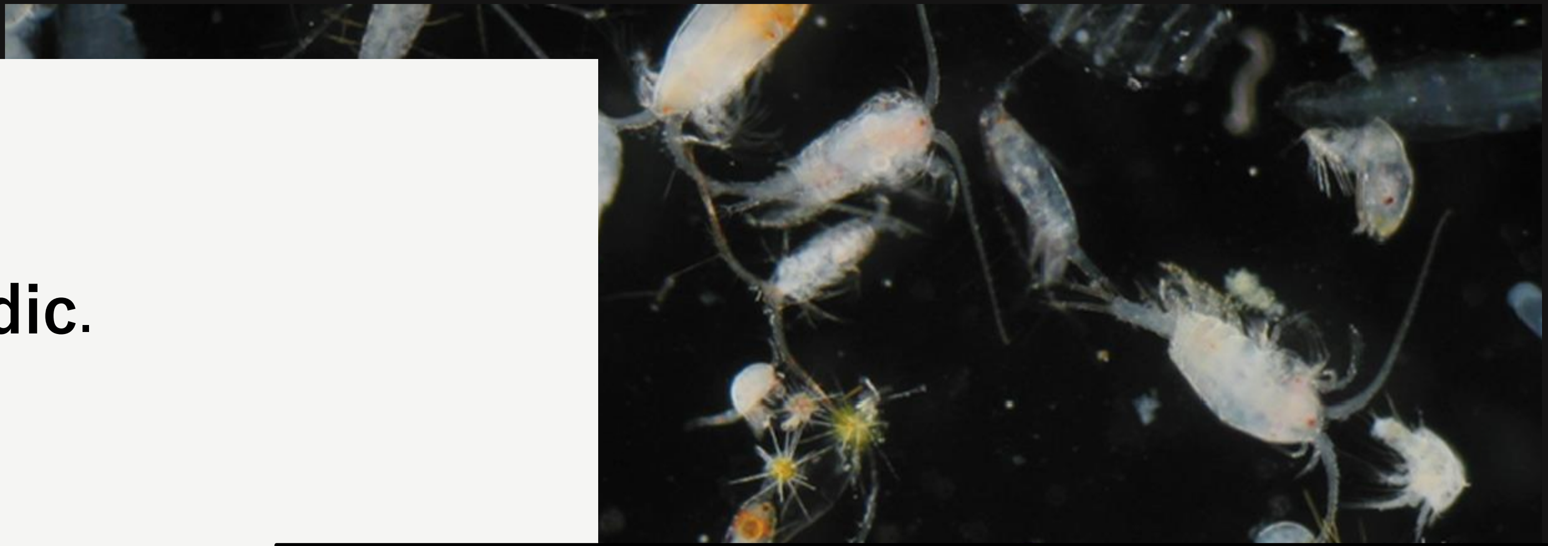


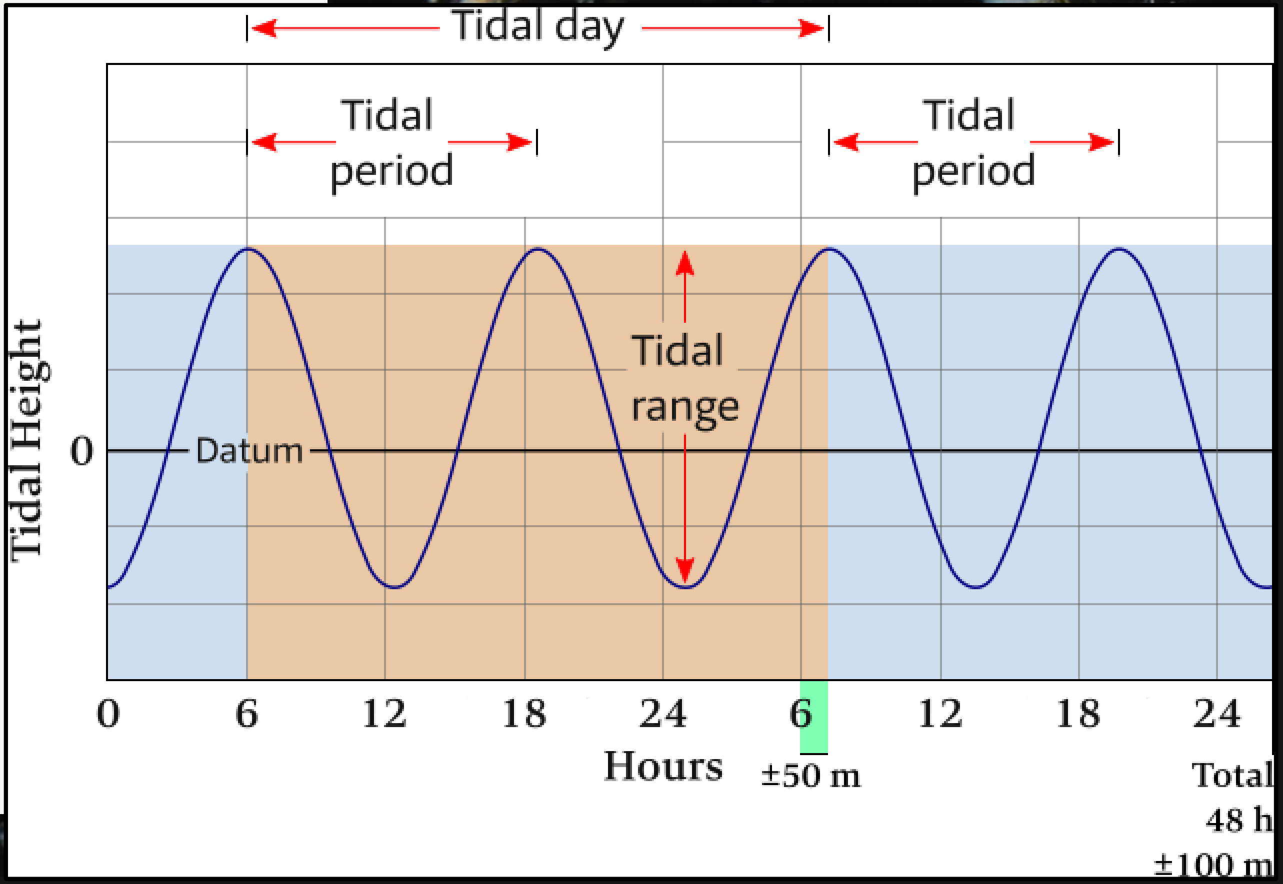




Ecological Data Series

Temporal data are often **periodic**.

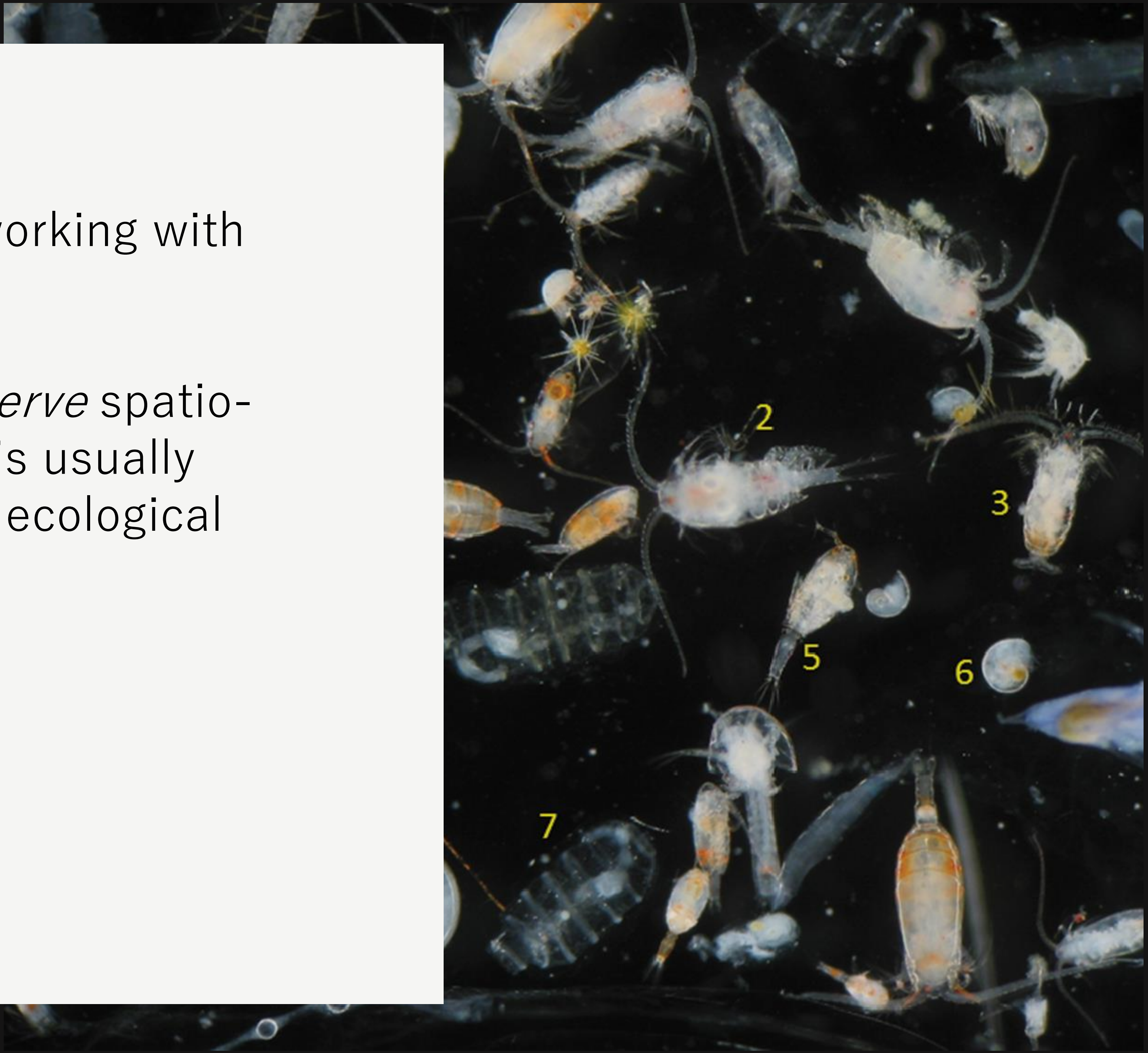




Ecological Data Series

Special considerations when working with data series:

1) Sampling design must *preserve* spatio-temporal variability (which is usually minimized in other types of ecological sampling)

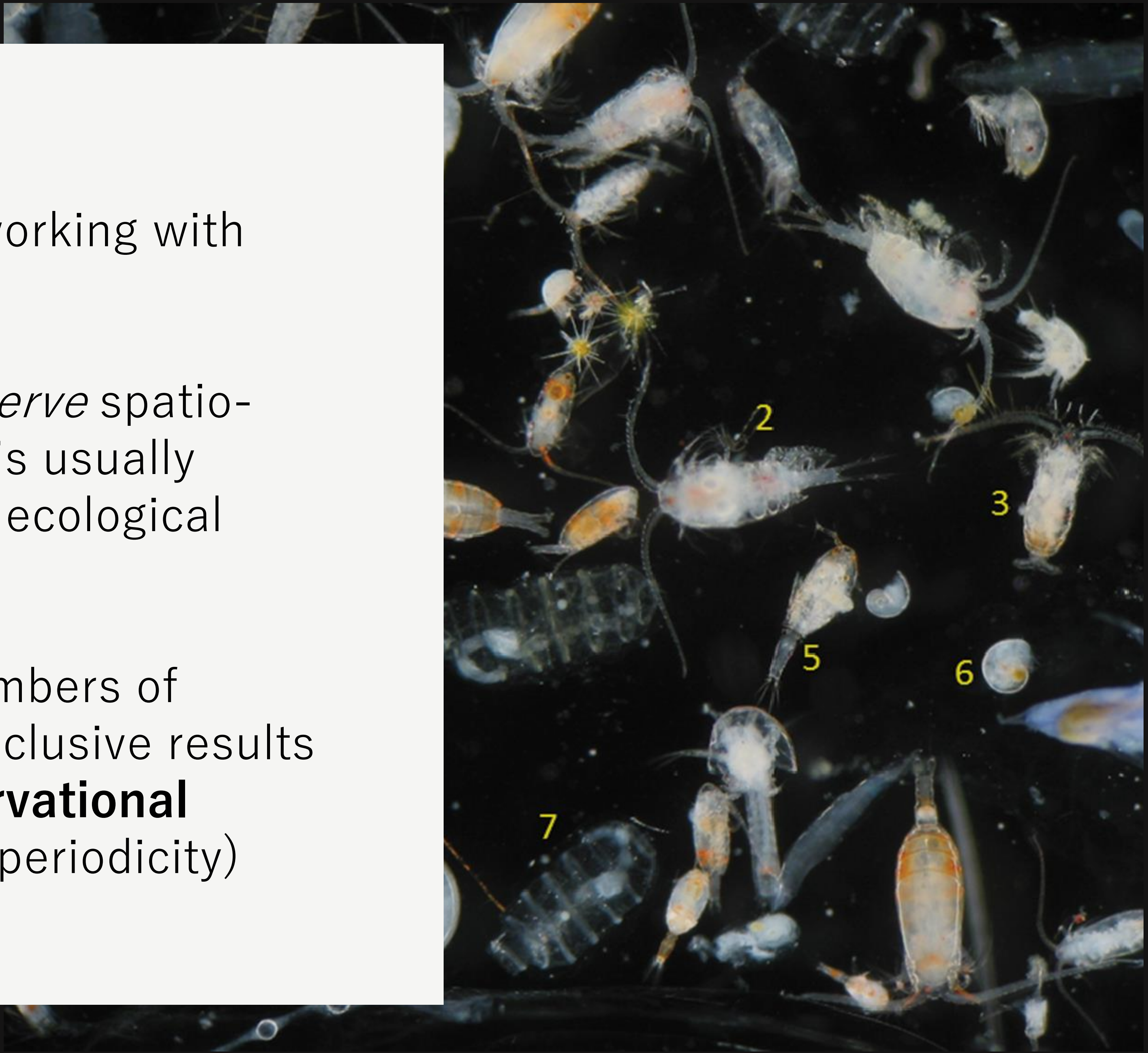


Ecological Data Series

Special considerations when working with data series:

1) Sampling design must *preserve* spatio-temporal variability (which is usually minimized in other types of ecological sampling)

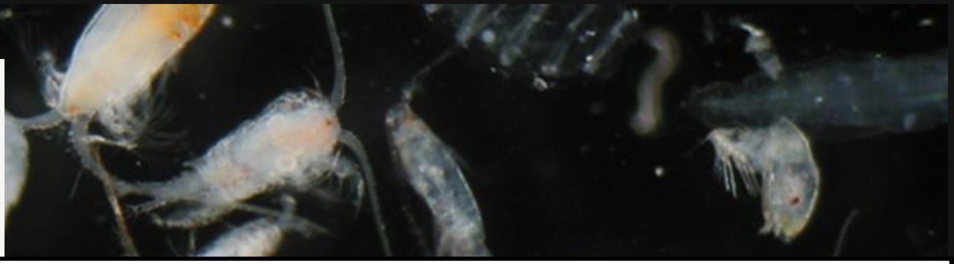
2) Sampling requires large numbers of observations to provide conclusive results (must have sufficient **observational window** to capture trends, periodicity)

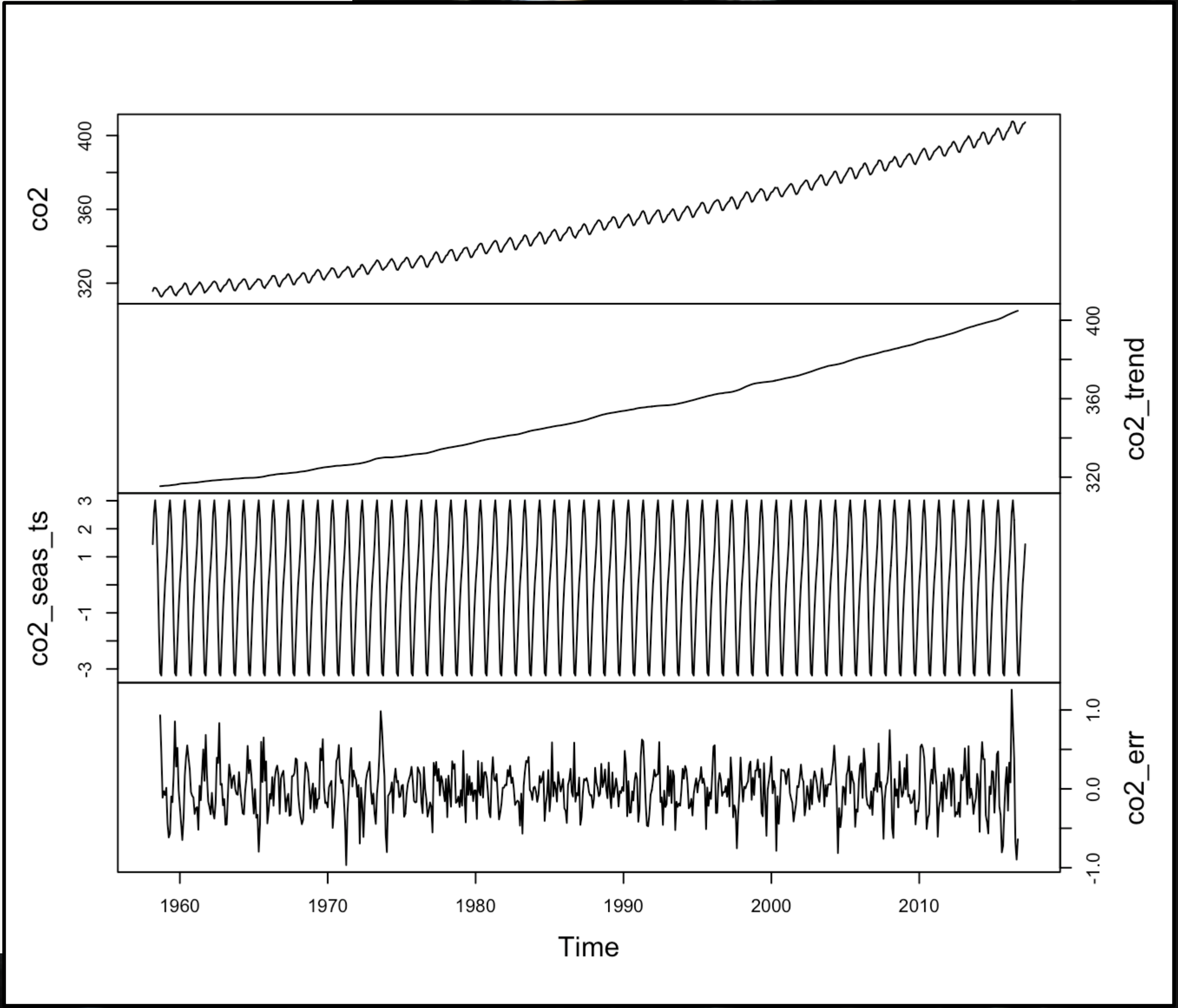


Ecological Data Series

Components:

- Periodic signal
- Trend
- Noise



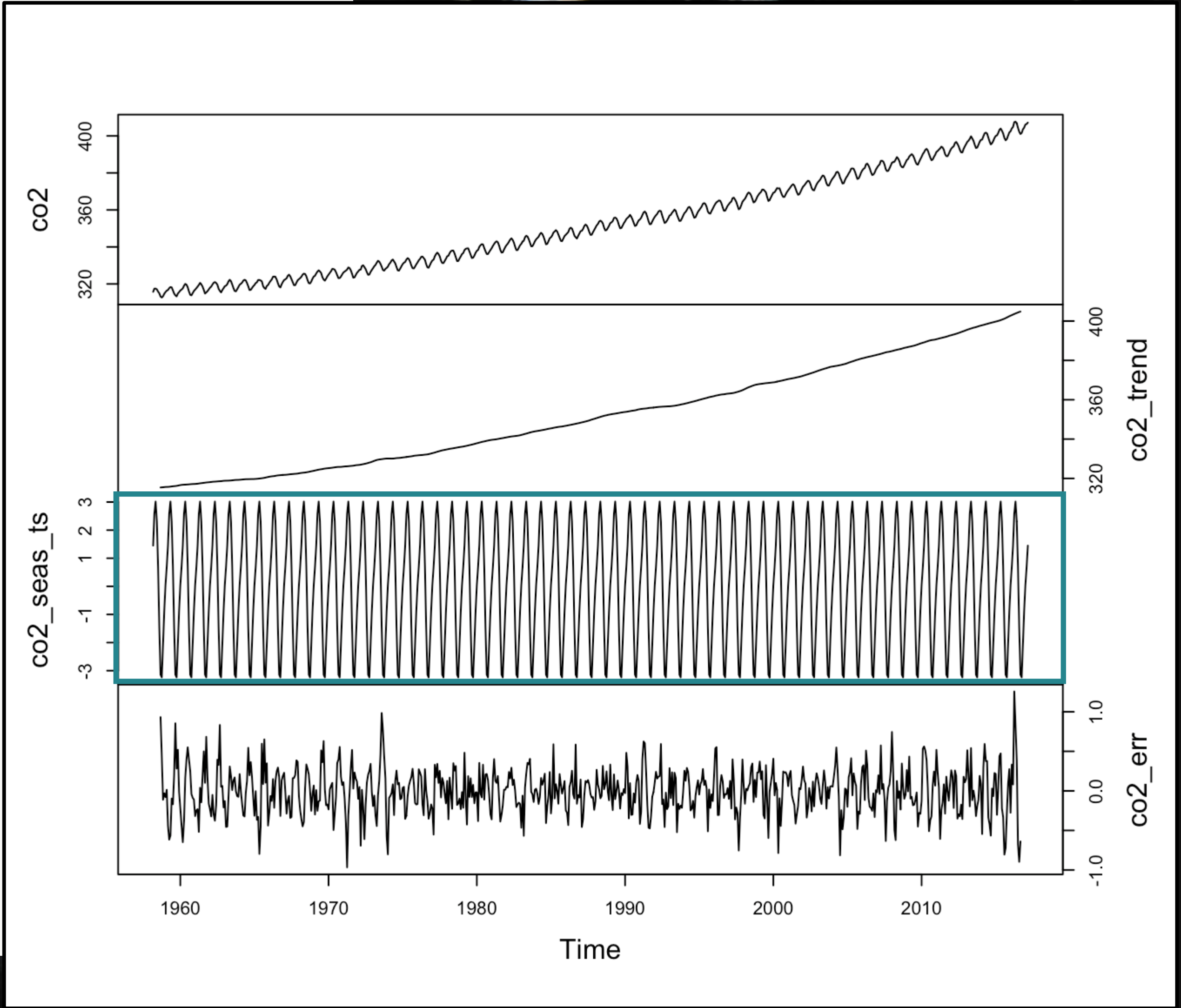


Ecological Data Series

Components:

- Periodic signal
- Trend
- Noise

Goal: Examine *systemic* variability

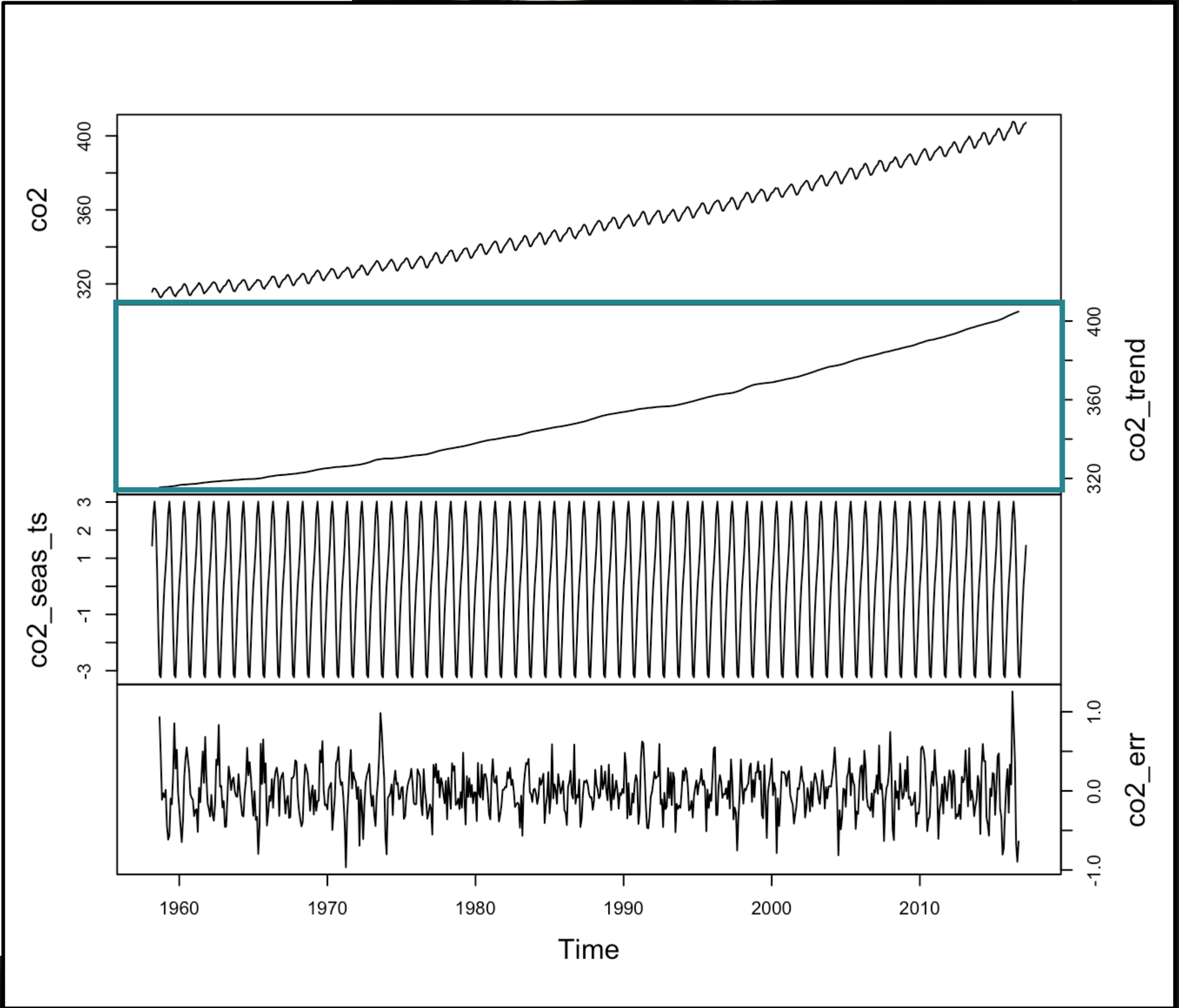


Ecological Data Series

Components:

- Periodic signal
- Trend
- Noise

Goal: Examine *deterministic* change

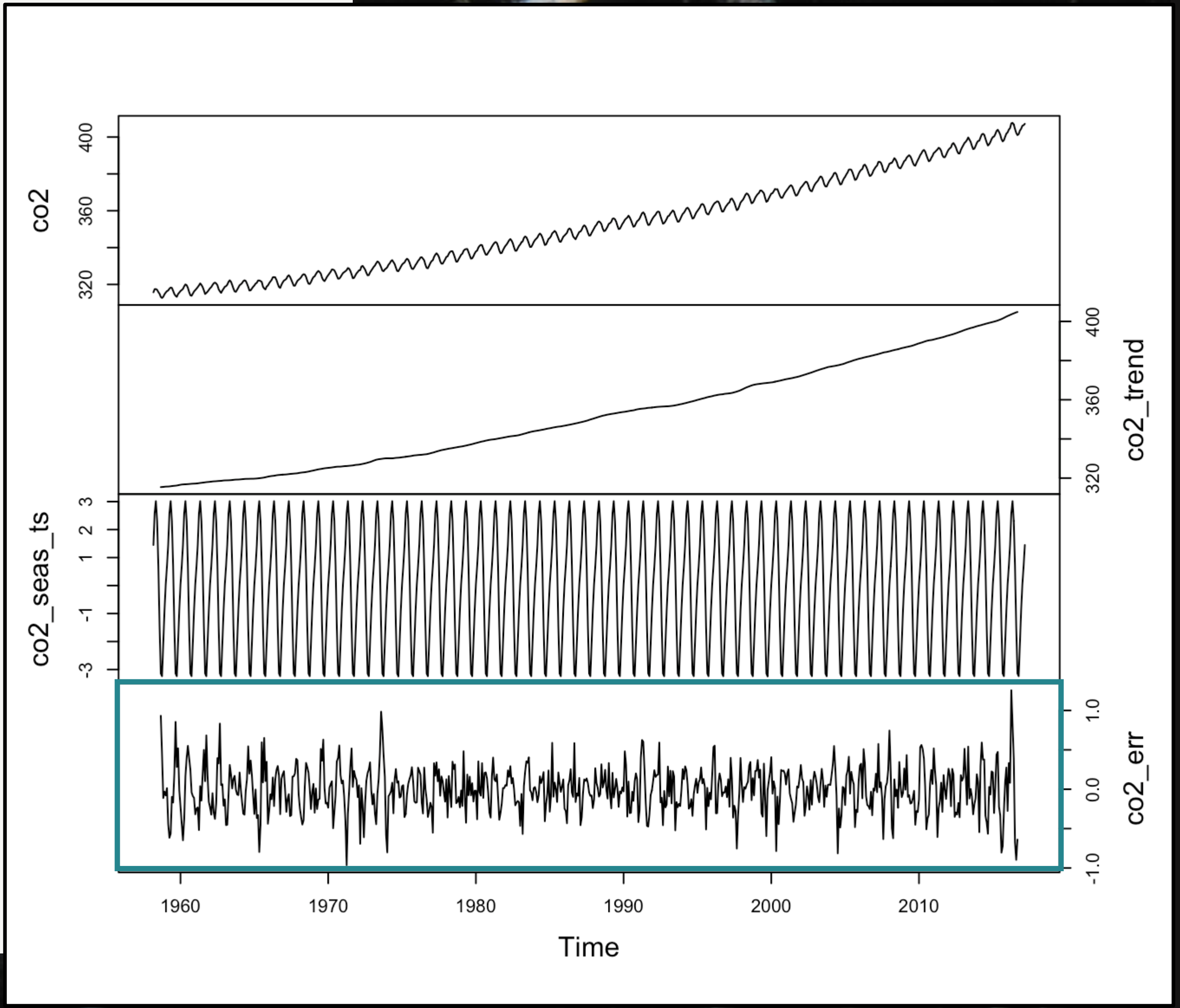


Ecological Data Series

Components:

- Periodic signal
- Trend
- Noise

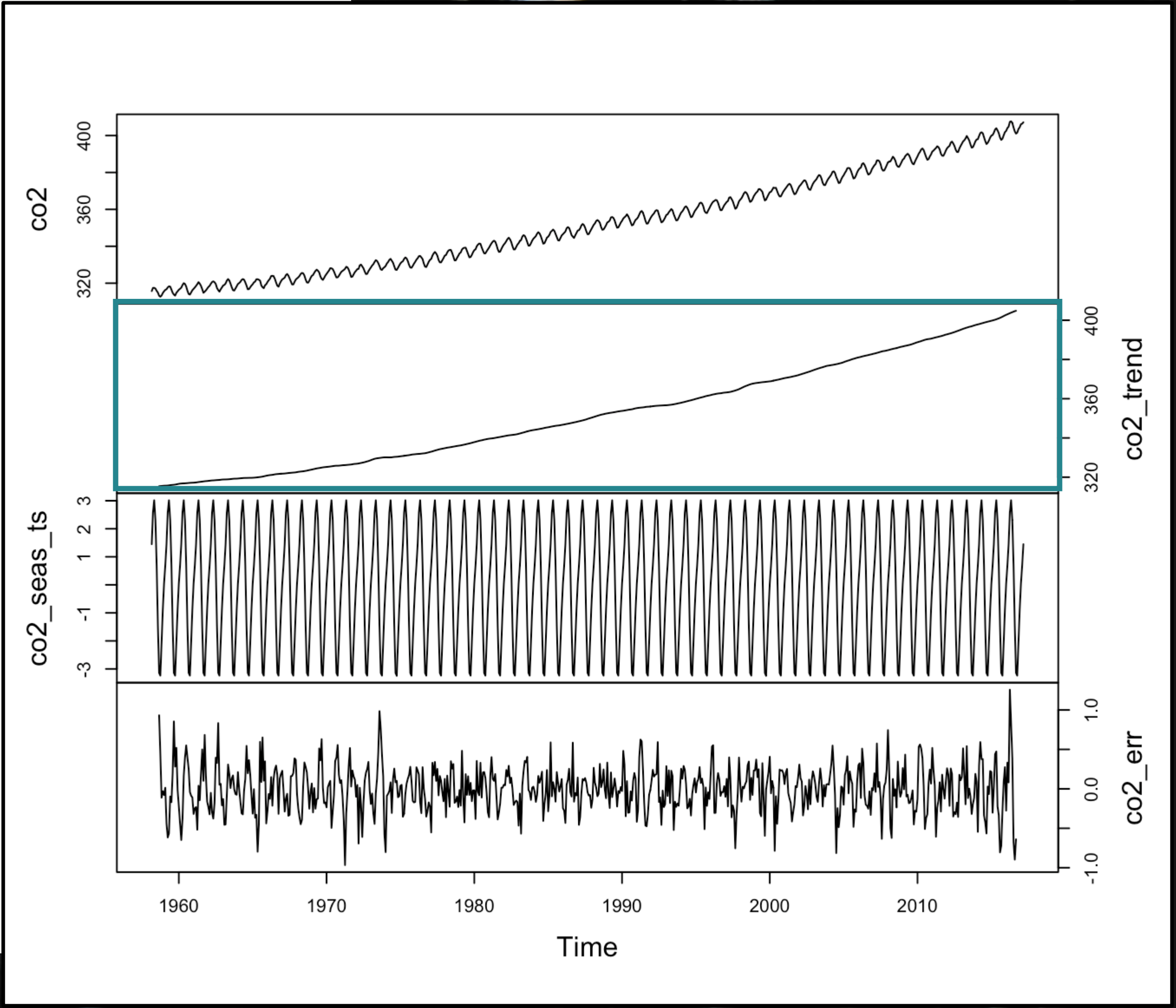
Goal: Examine *random* fluctuations



Ecological Data Series

Examining a periodic component of a data series must be conducted on **stationary** data.

To do this, the dataset must be **detrended**.

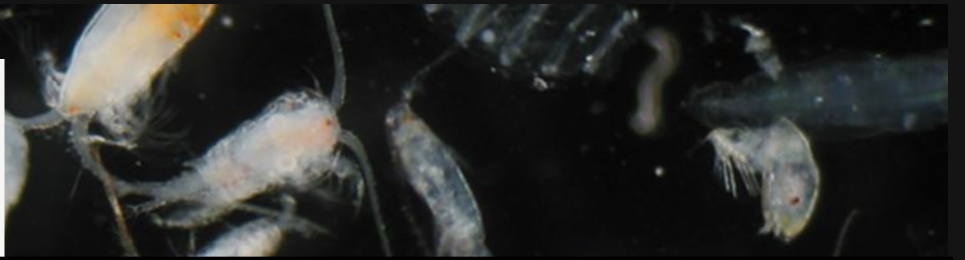


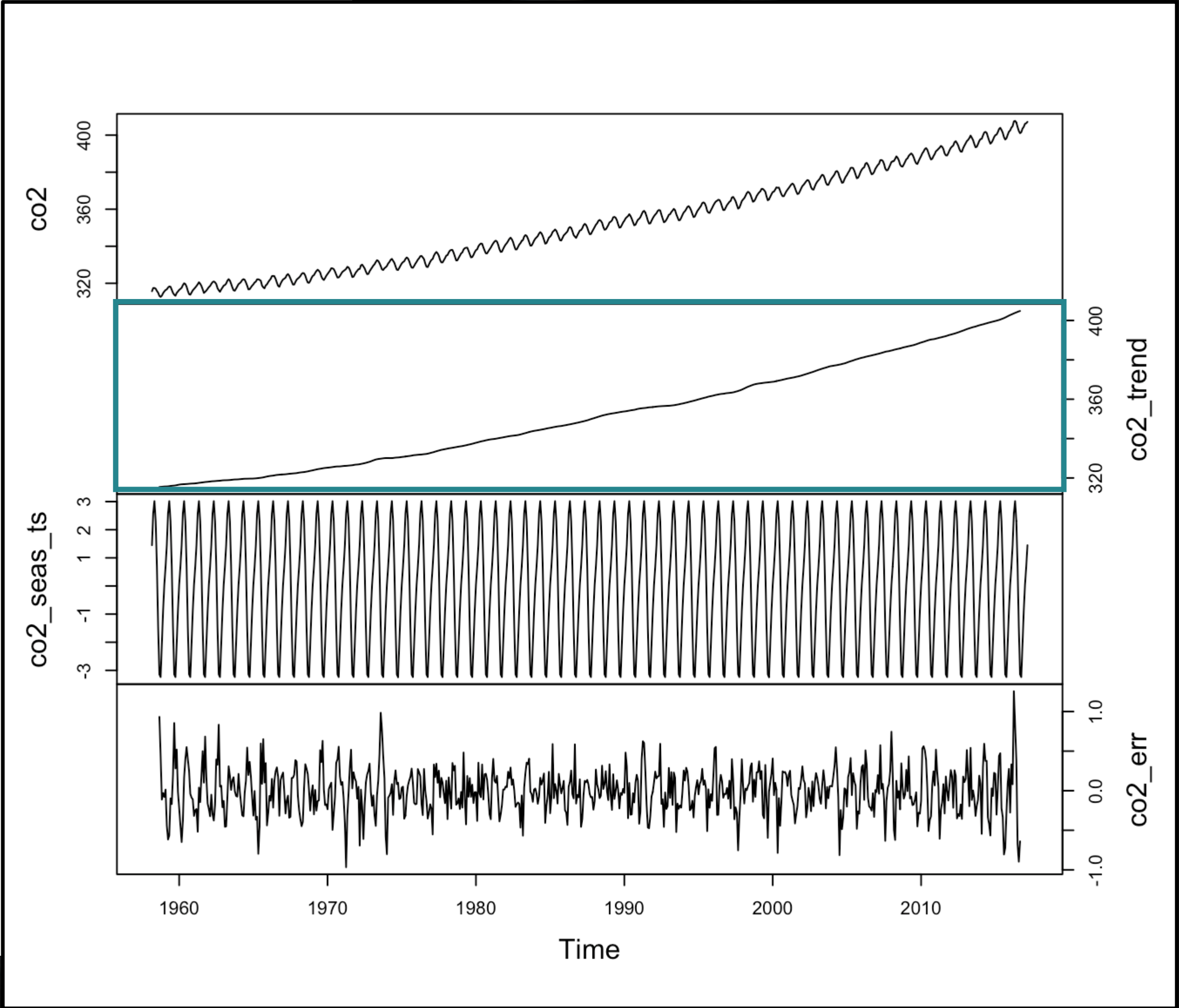
Ecological Data Series

Examining a periodic component of a data series must be conducted on **stationary** data.

To do this, the dataset must be **detrended**.

Removal of noise through **filtration** may also be necessary.





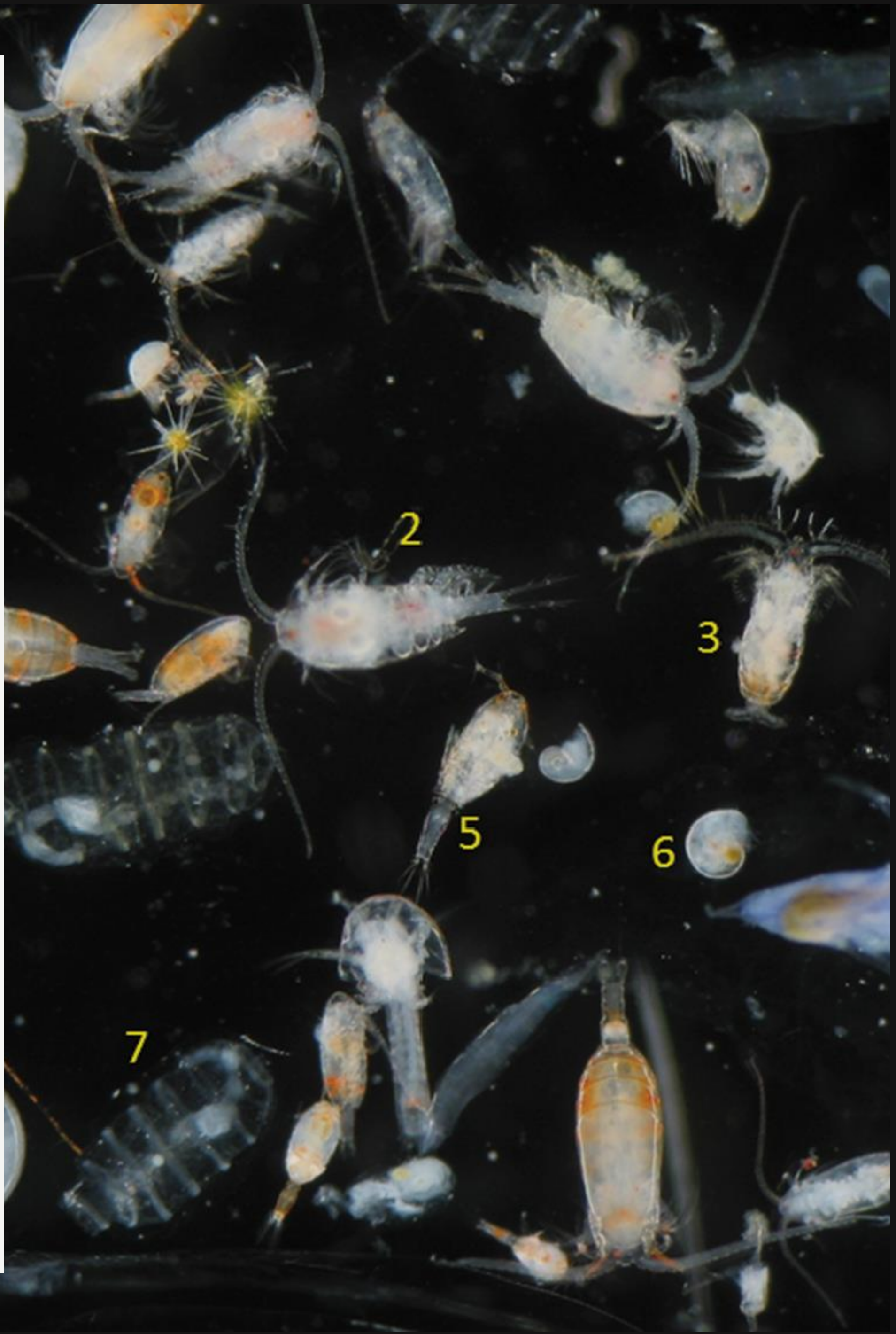
Ecological Data Series

Objectives of the analysis of ecological data series:

1) Identify characteristic periods

2) Characterize the trend

3) Examine random fluctuations (noise)



Ecological Data Series

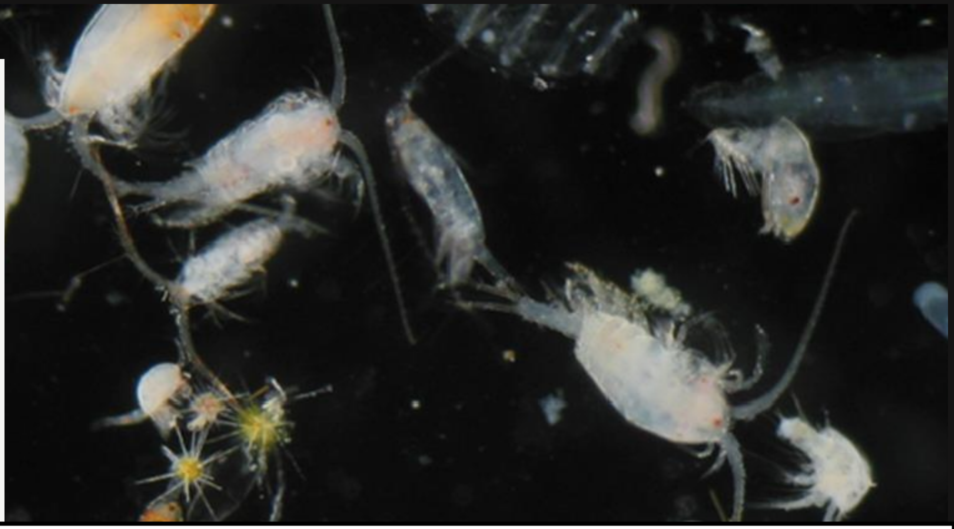
Objectives of the analysis of ecological data series:

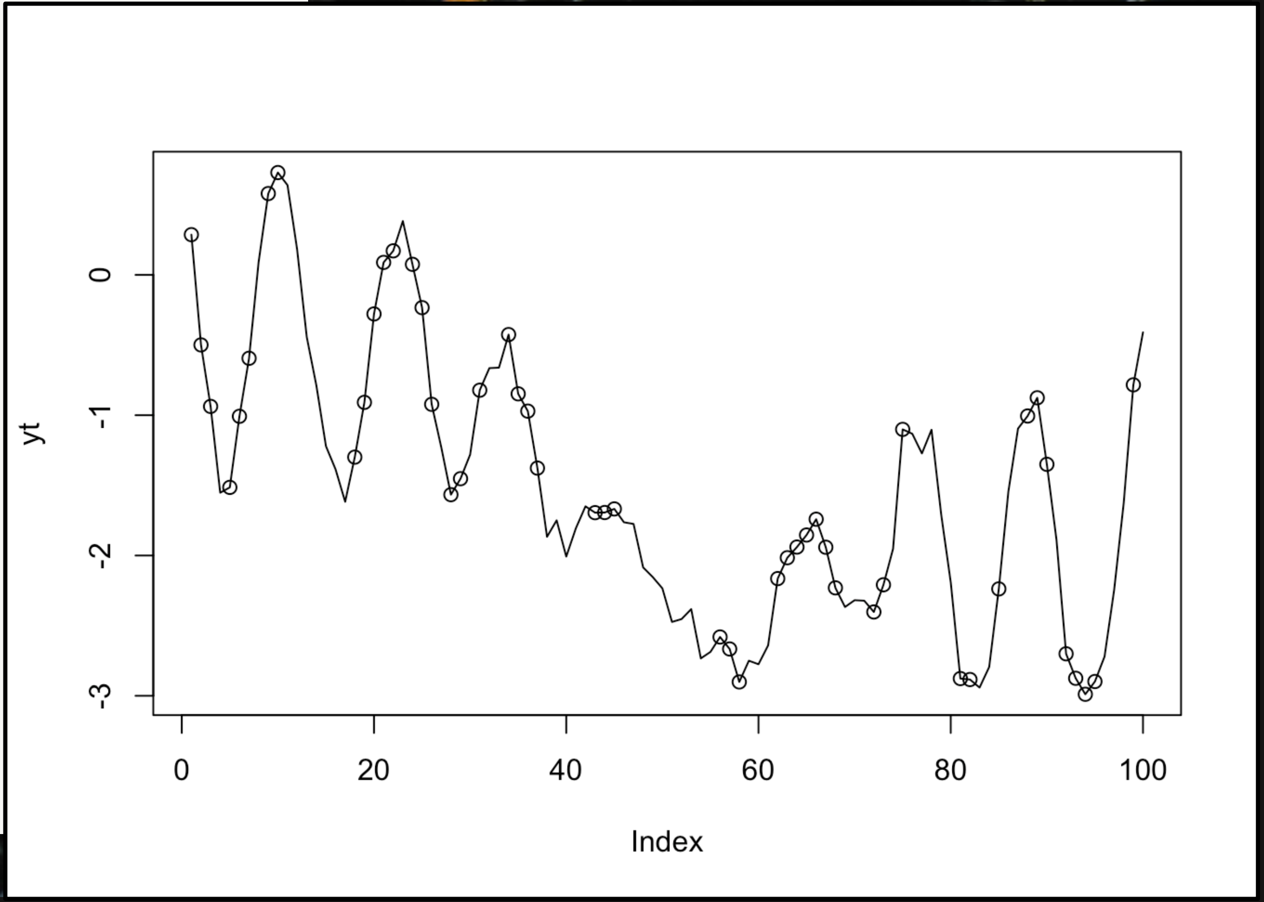
1) Identify characteristic periods

2) Characterize the trend

3) Examine random fluctuations (noise)

4) **Detect discontinuities**





Ecological Data Series

Objectives of the analysis of ecological data series:

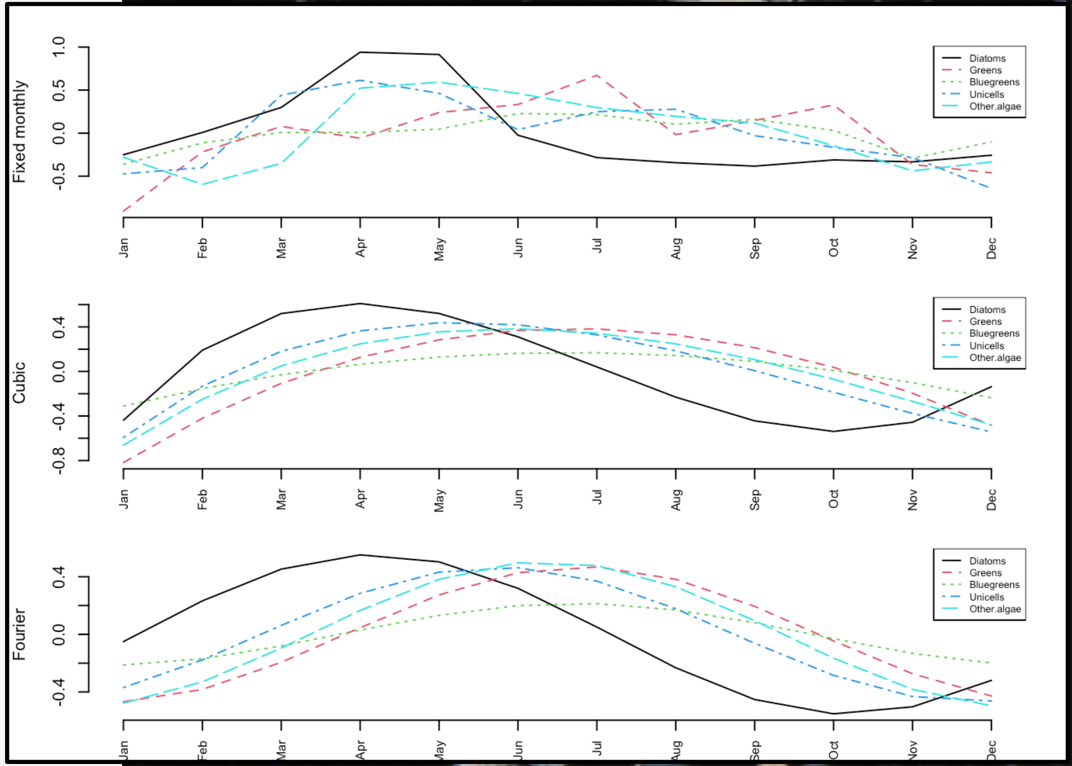
1) Identify characteristic periods

2) Characterize the trend

3) Examine random fluctuations (noise)

4) Detect discontinuities

5) **Correlate changes in one series with another**



Ecological Data Series

Objectives of the analysis of ecological data series:

1) Identify characteristic periods

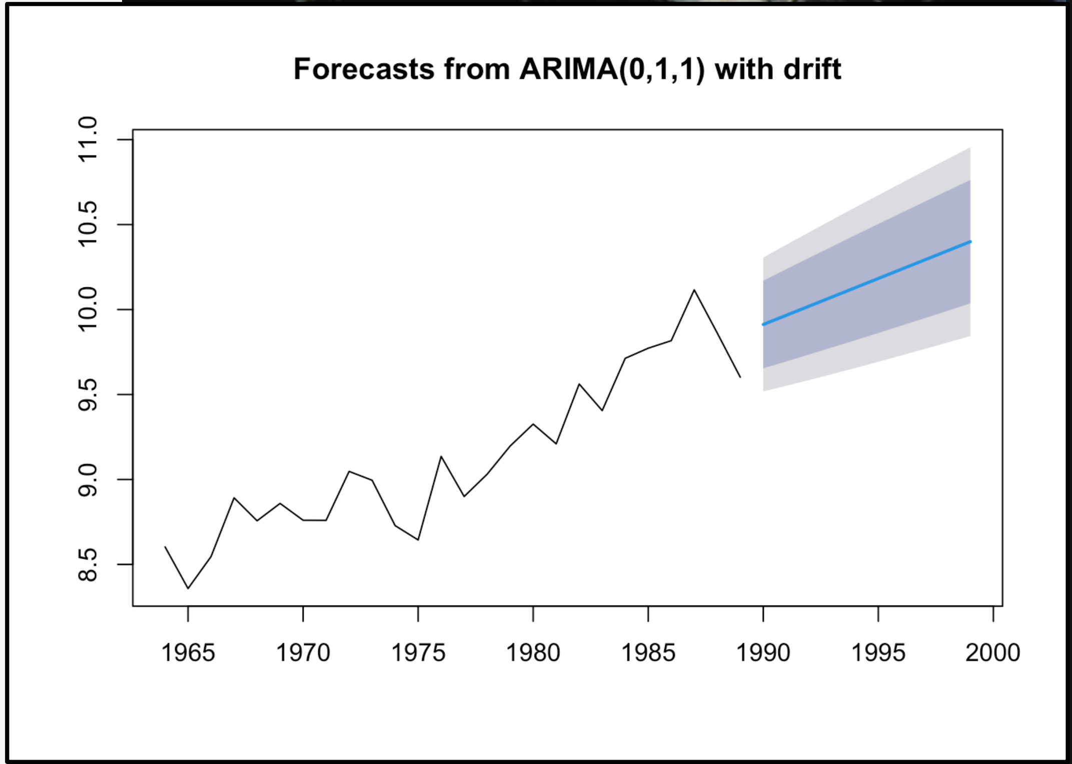
2) Characterize the trend

3) Examine random fluctuations (noise)

4) Detect discontinuities

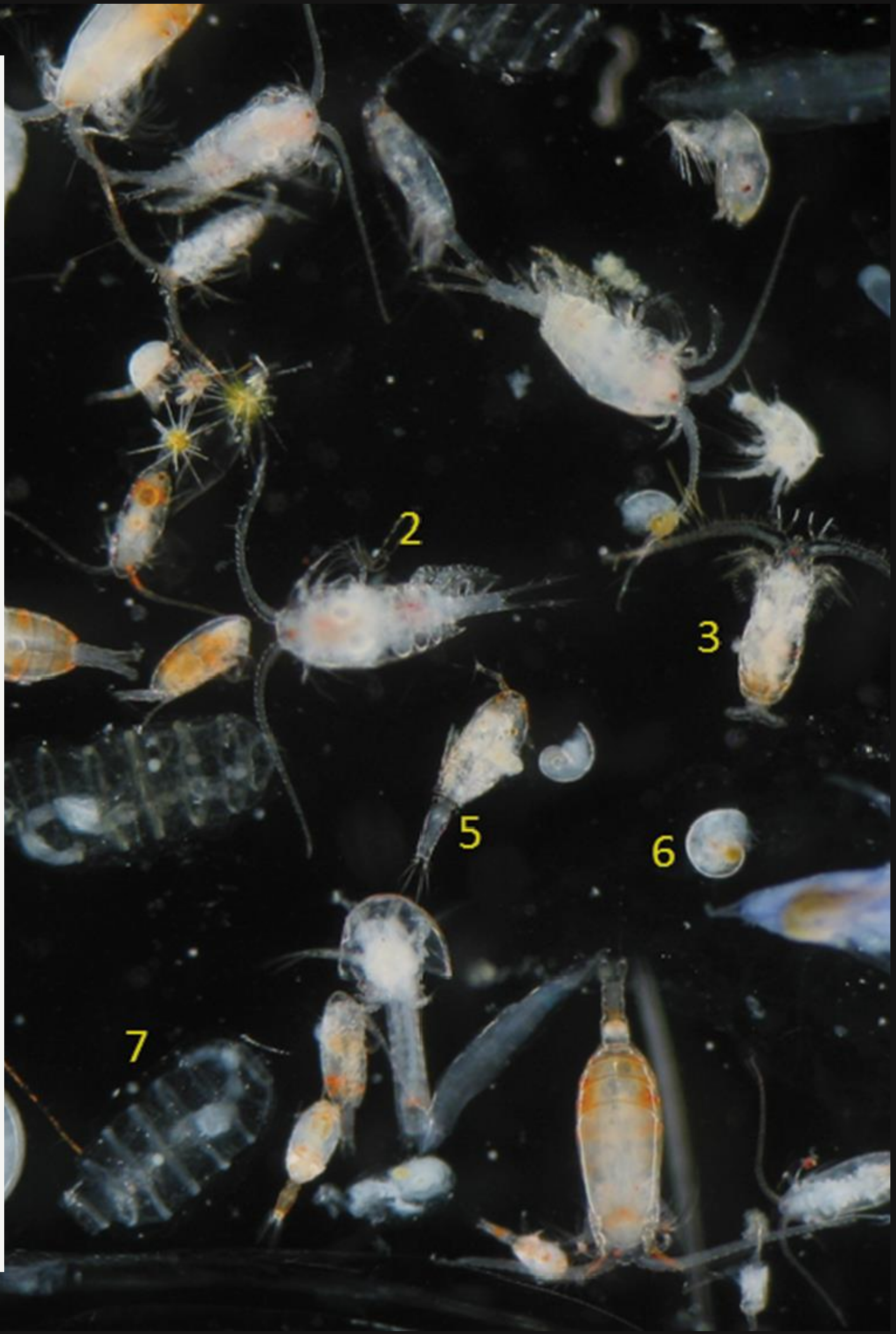
5) Correlate changes in one series with another

6) **Forecast the data series**



Ecological Data Series

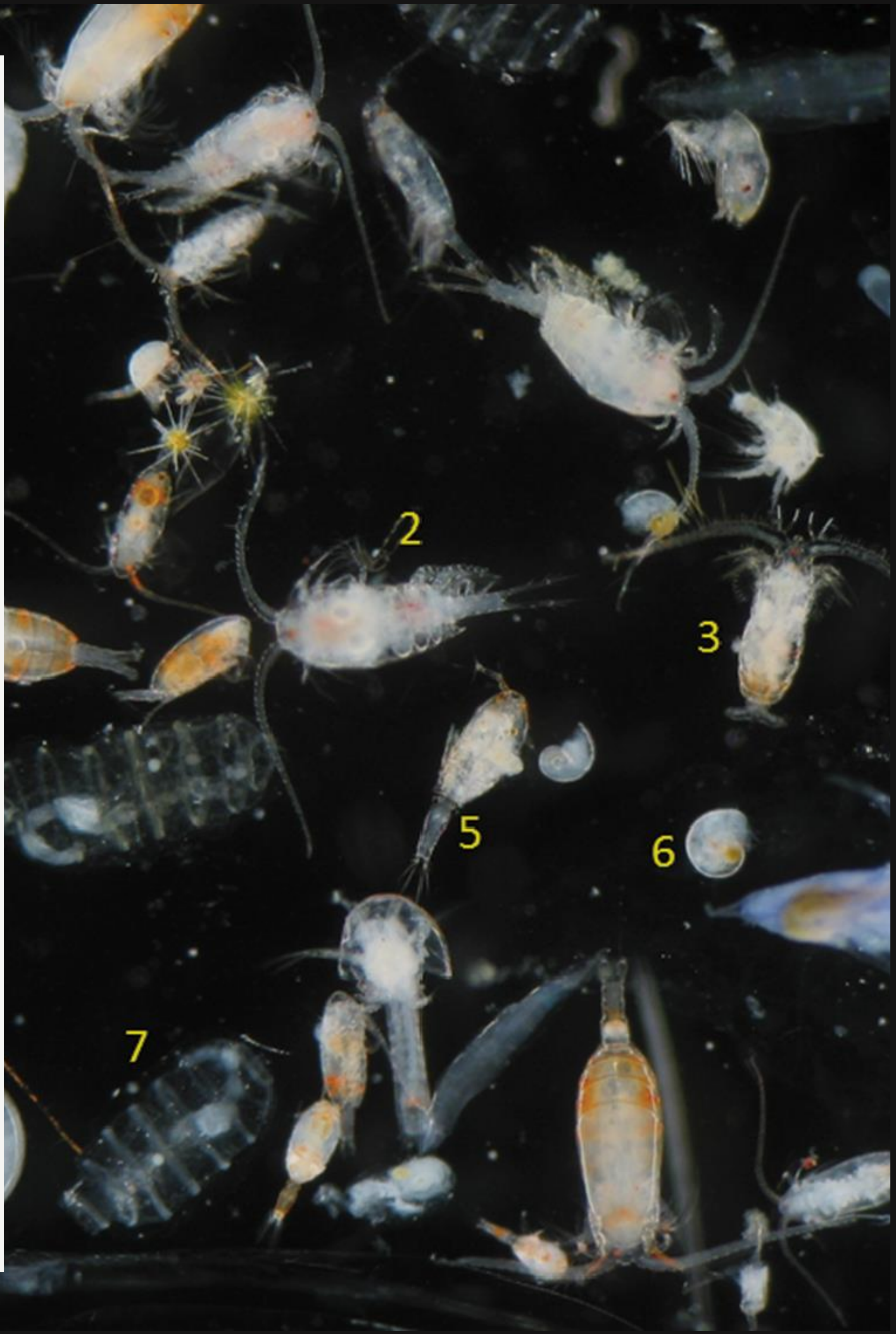
Is there a trend???



Ecological Data Series

Is there a trend???

1) Regress data series on time (or distance)

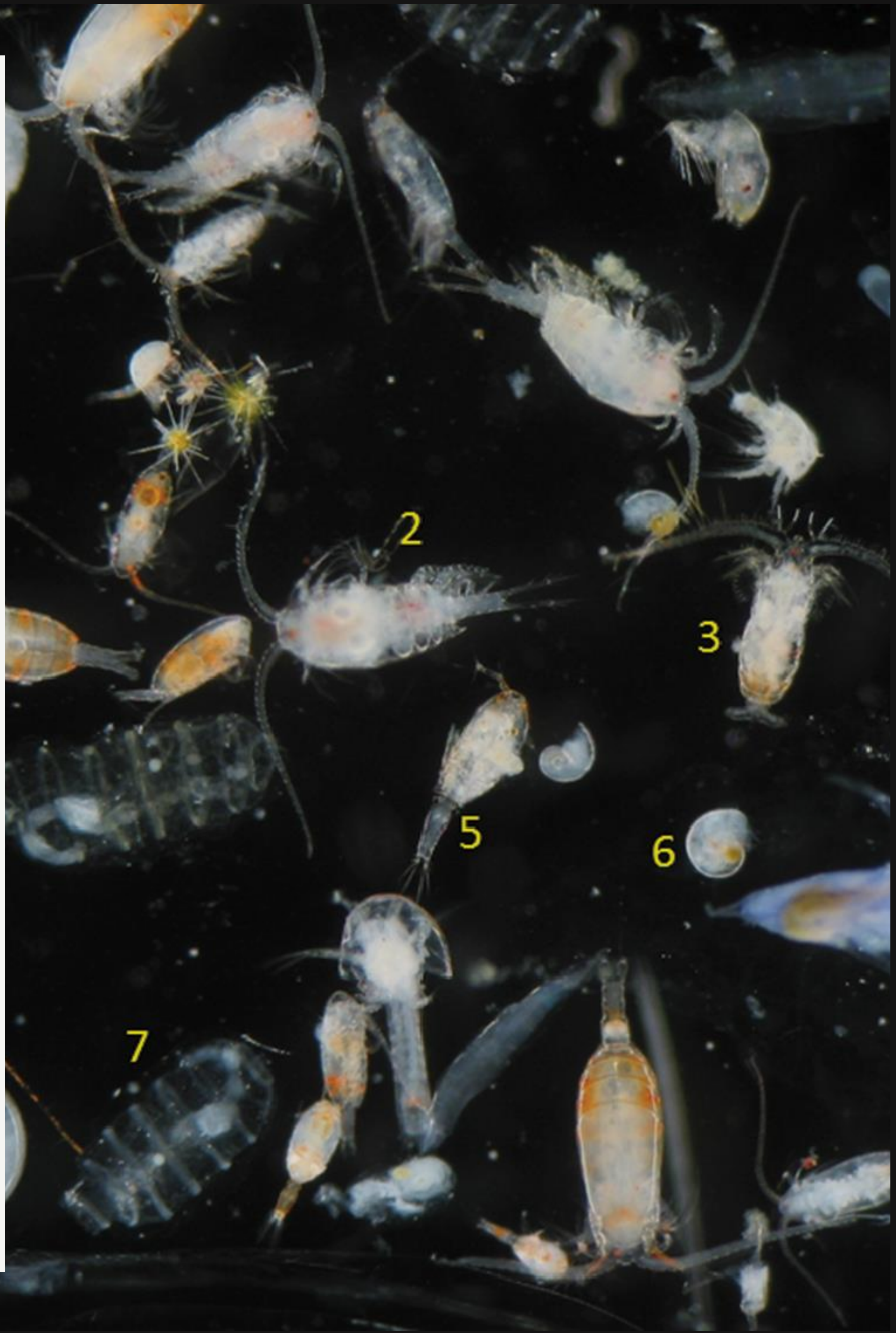


Ecological Data Series

Is there a trend???

1) Regress data series on time (or distance)

2) Use the **“sign test”**



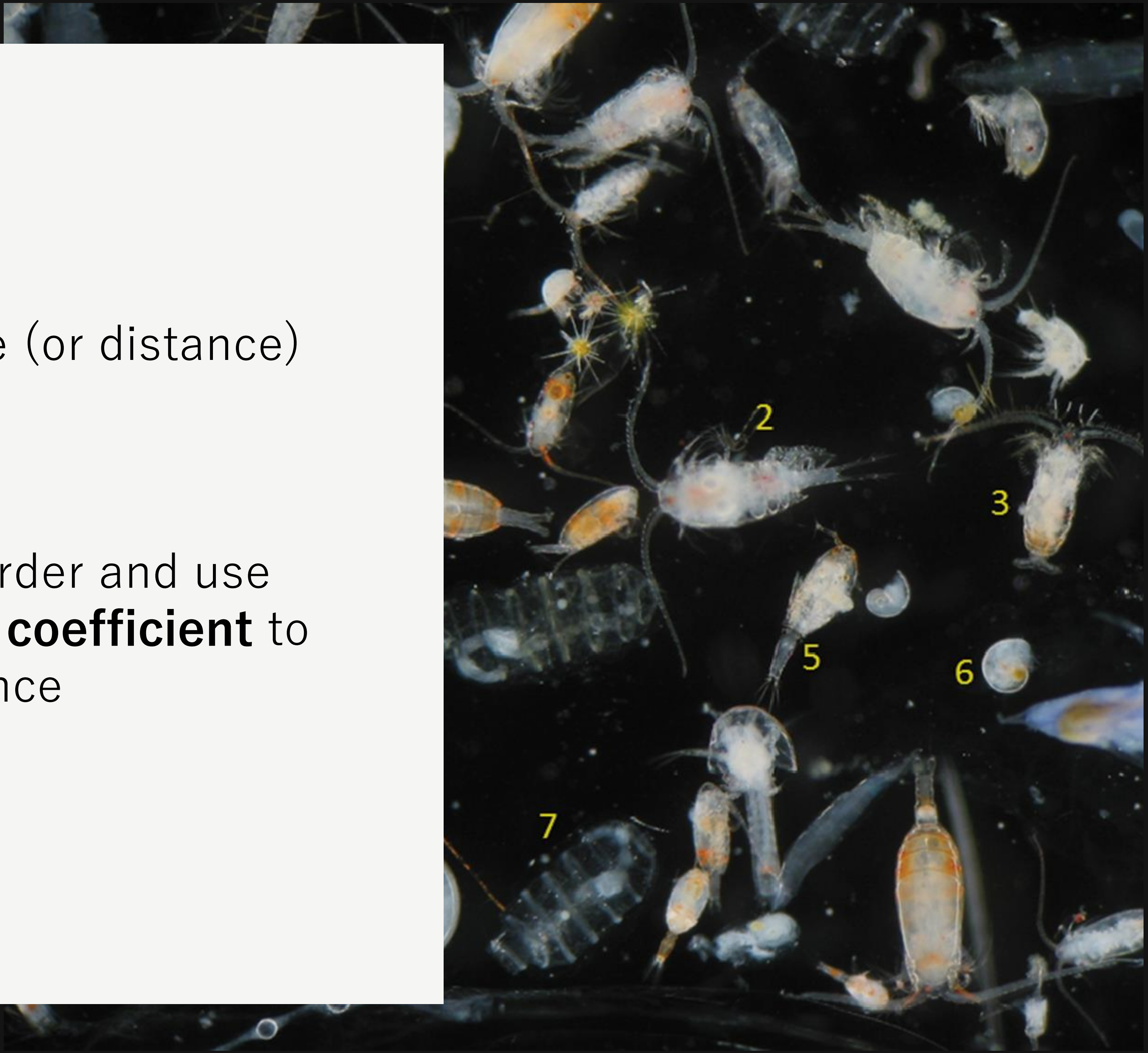
Ecological Data Series

Is there a trend???

1) Regress data series on time (or distance)

2) Use the “sign test”

3) Rank values in increasing order and use **Kendall’s rank correlation coefficient** to assess degree of resemblance



Ecological Data Series

Is there a trend???

1) Regress data series on time (or distance)

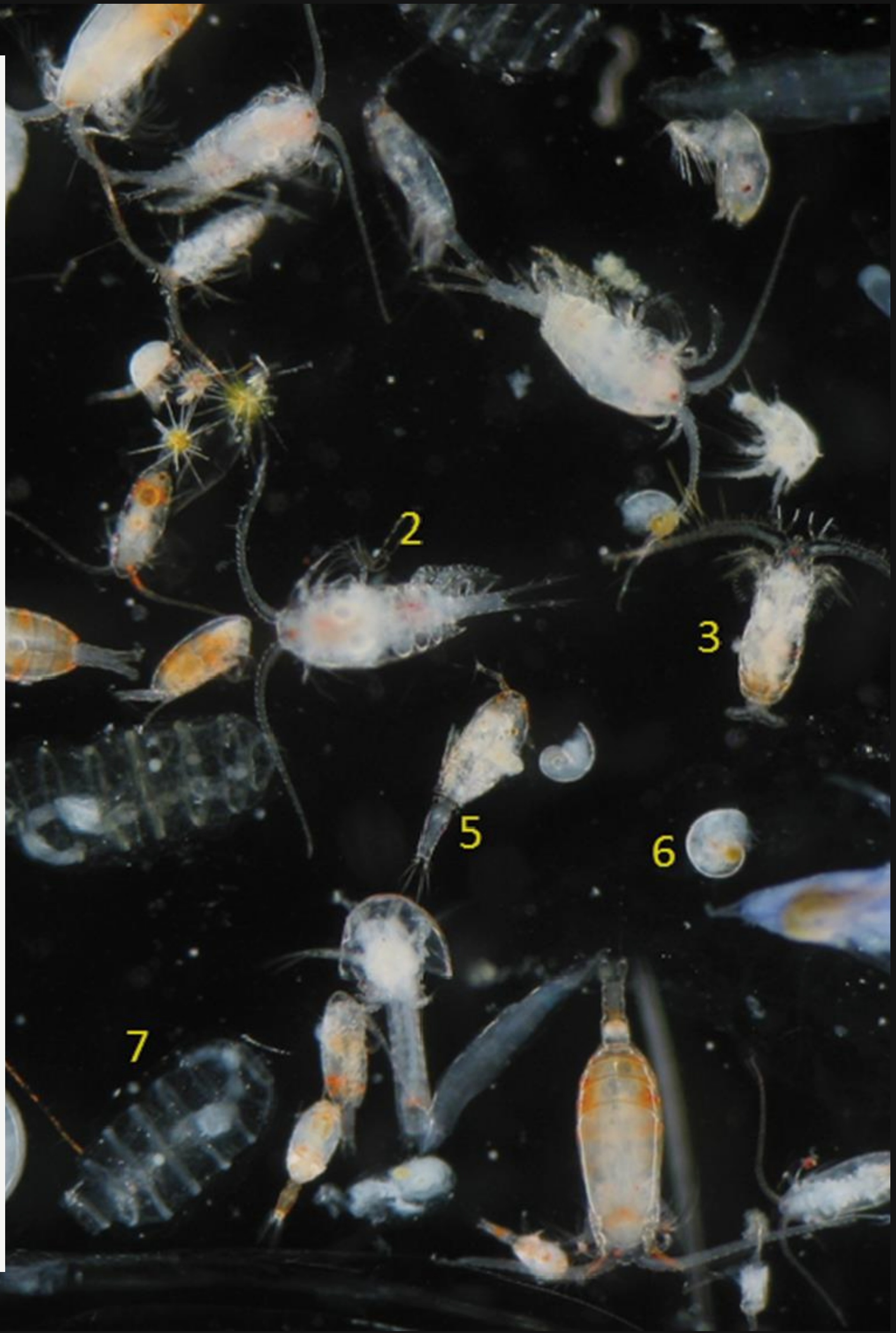
2) Use the “sign test”

3) Rank values in increasing order and use Kendall’s rank correlation coefficient to assess degree of resemblance

4) Use **non-parametric** approach to compare observed and random runs of signs (+/-)

Ecological Data Series

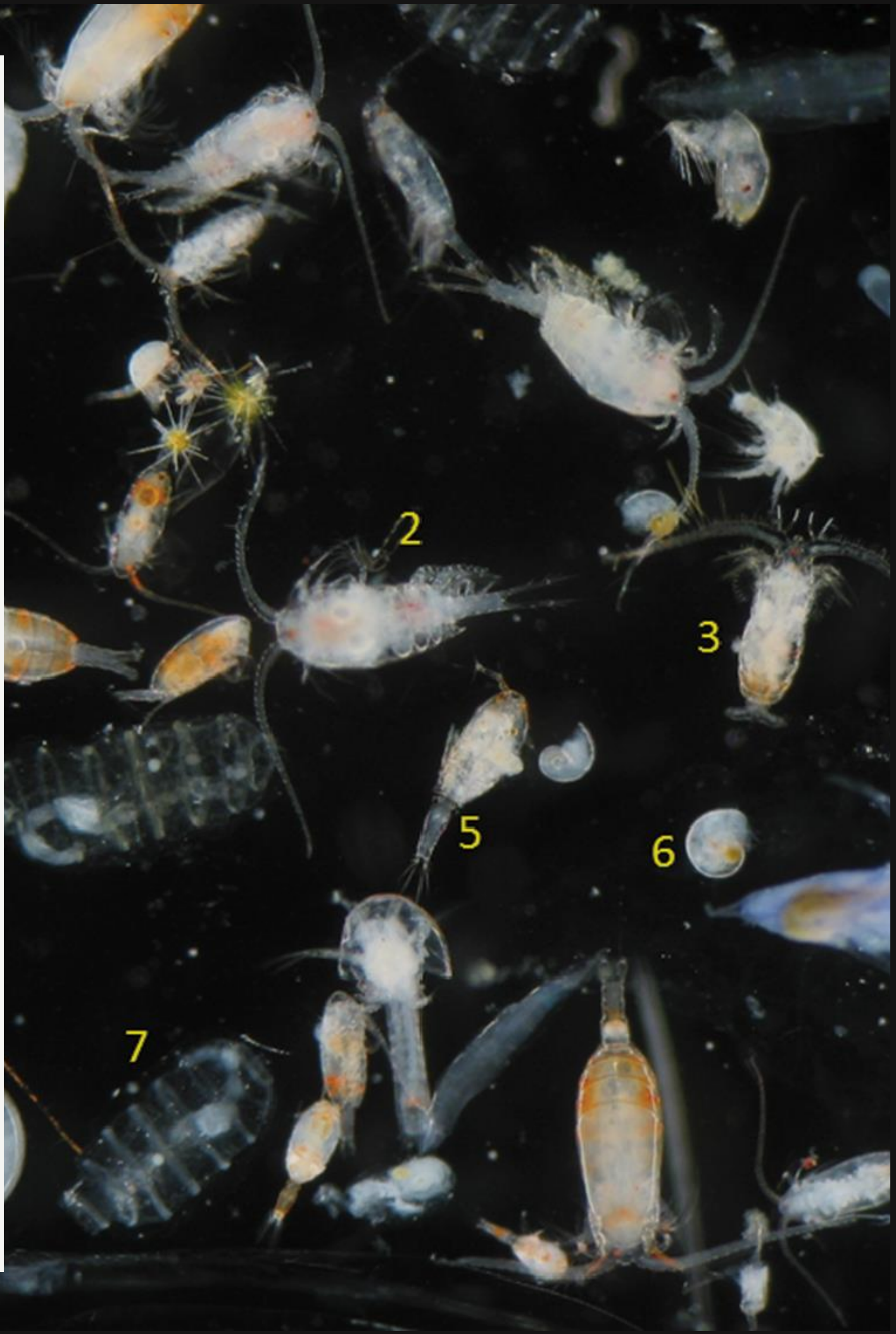
Ok, there’s a trend, so how do we estimate it???



Ecological Data Series

Ok, there's a trend, so how do we estimate it???

1) **Analytical method:** Regression of data series through time

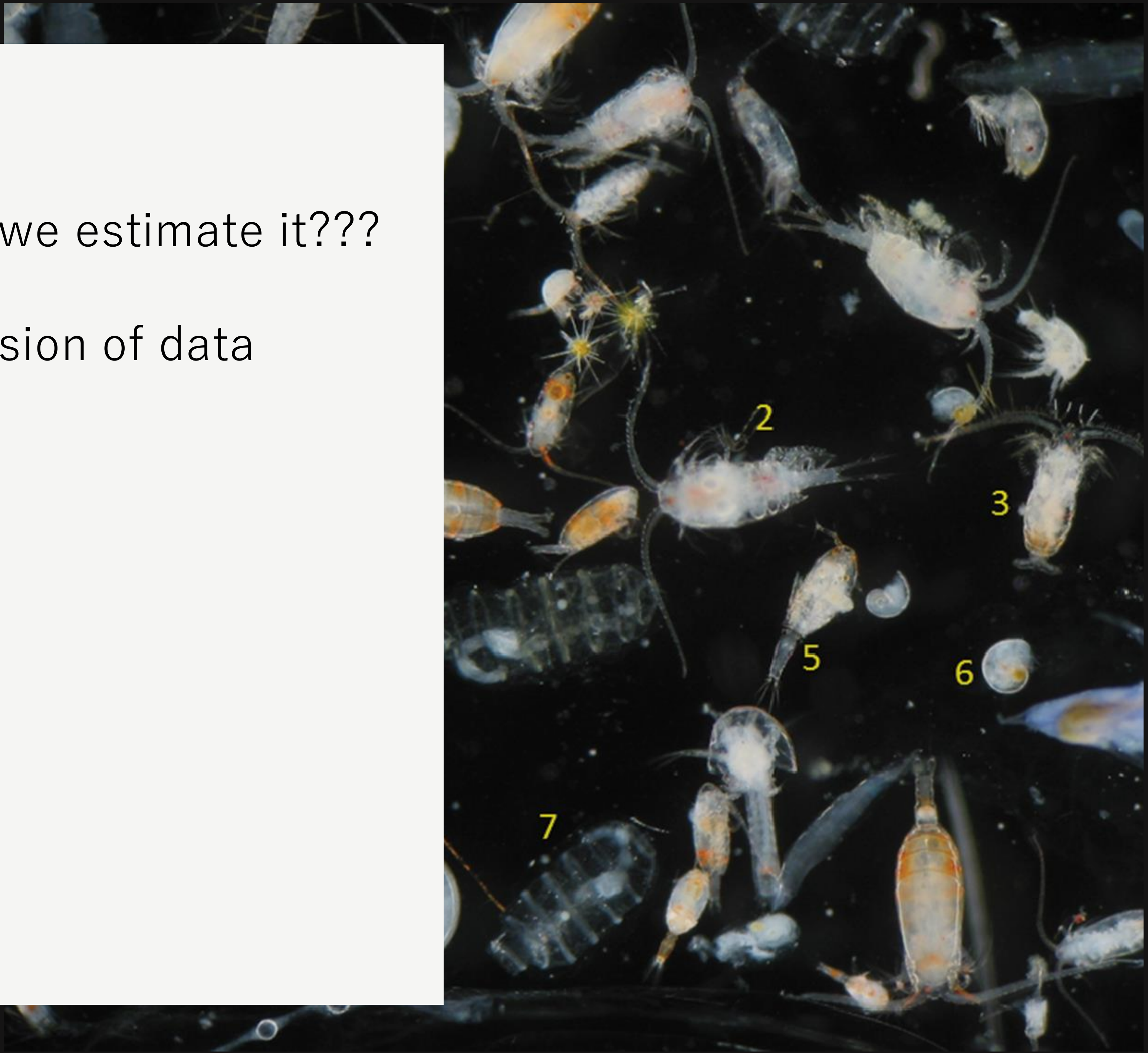


Ecological Data Series

Ok, there's a trend, so how do we estimate it???

1) **Analytical method:** Regression of data series through time

2) **Moving average** window



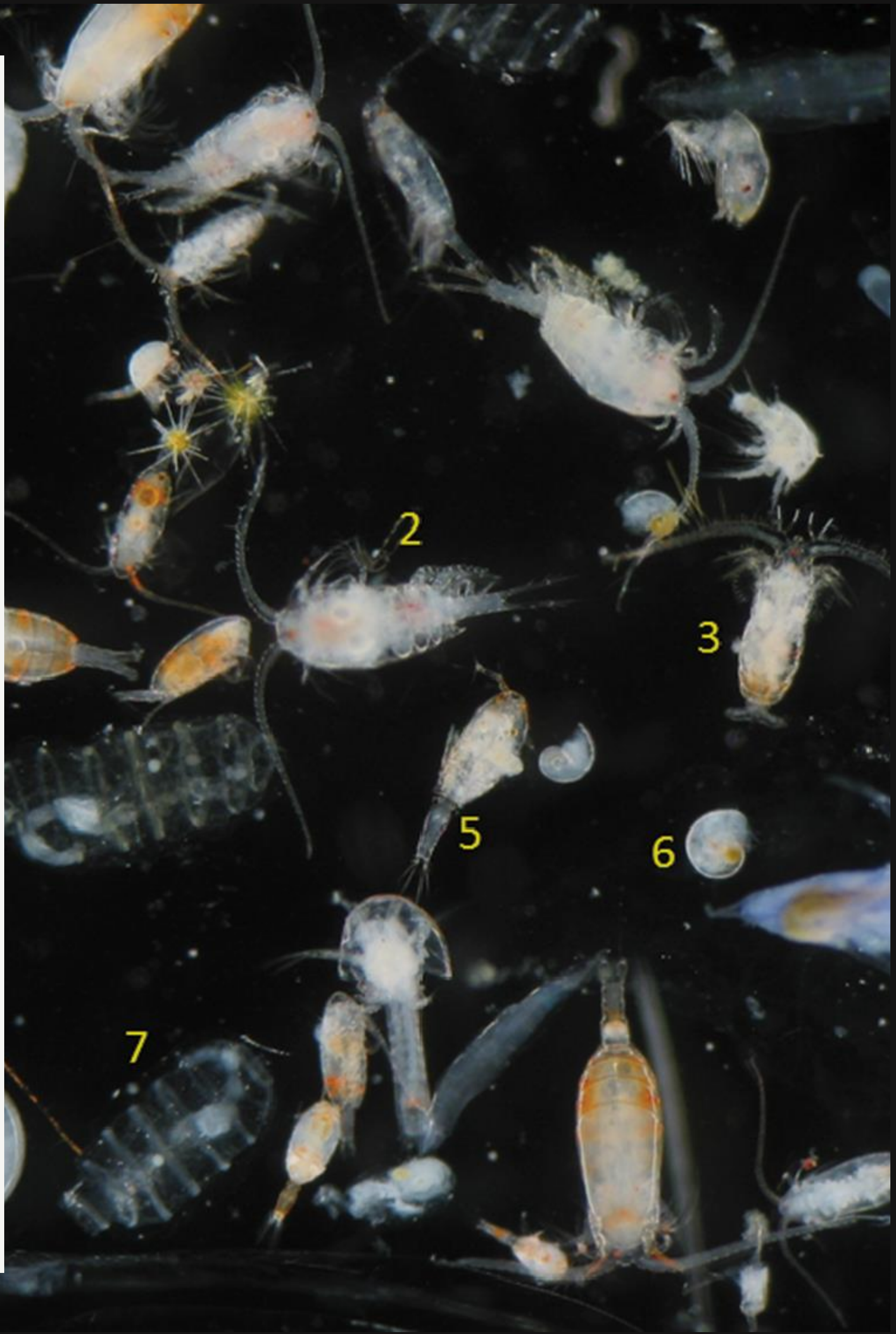
Ecological Data Series

Ok, there's a trend, so how do we estimate it???

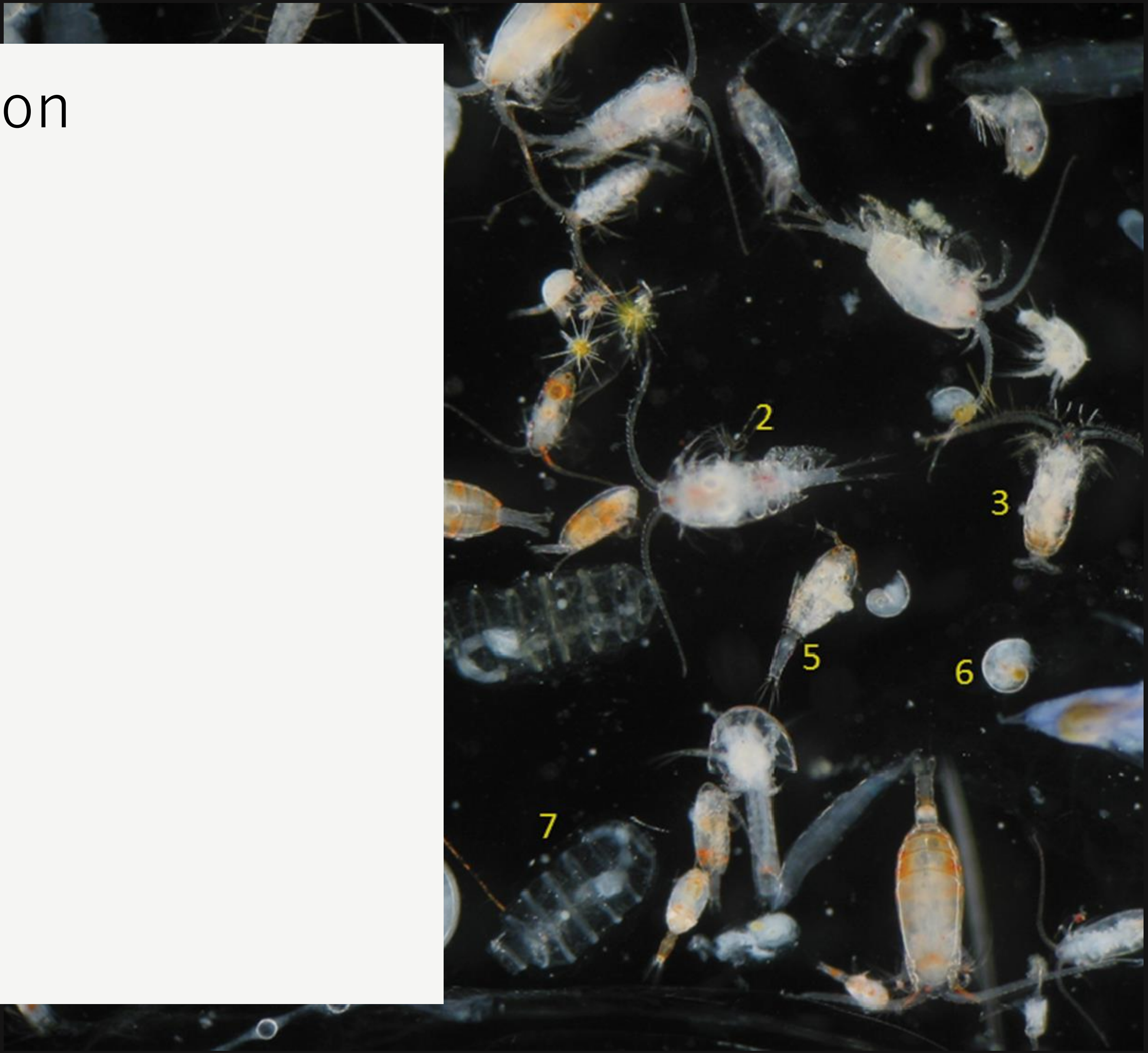
1) **Analytical method:** Regression of data series through time

2) **Moving average** window

3) **Autoregressive or "random walk"** type approaches



Temporal Autocorrelation

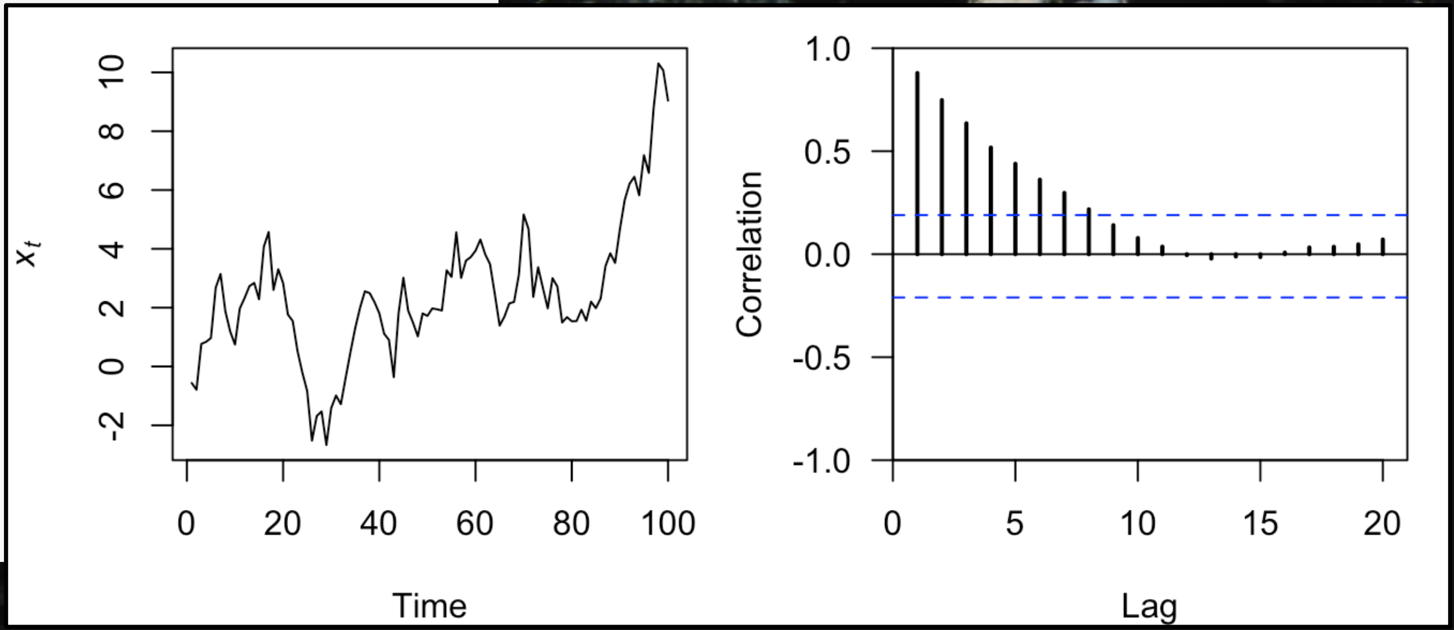


Temporal Autocorrelation

“The most natural axis along which processes may be studied is time, because temporal phenomena develop in an irreversible way, and independently of any decision made by the observer.”

Temporal Autocorrelation

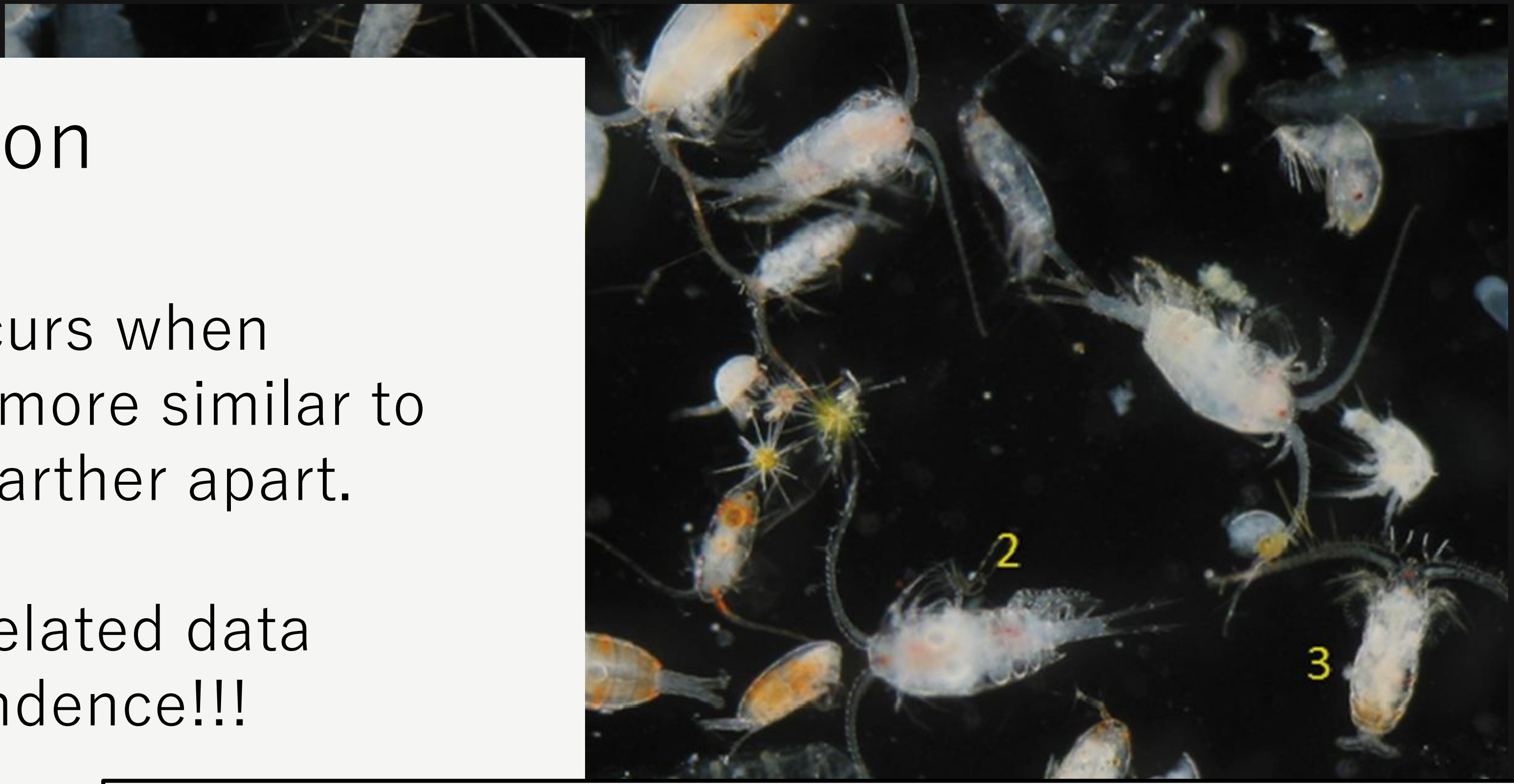
Temporal autocorrelation occurs when observations close in time are more similar to each other than observations farther apart.

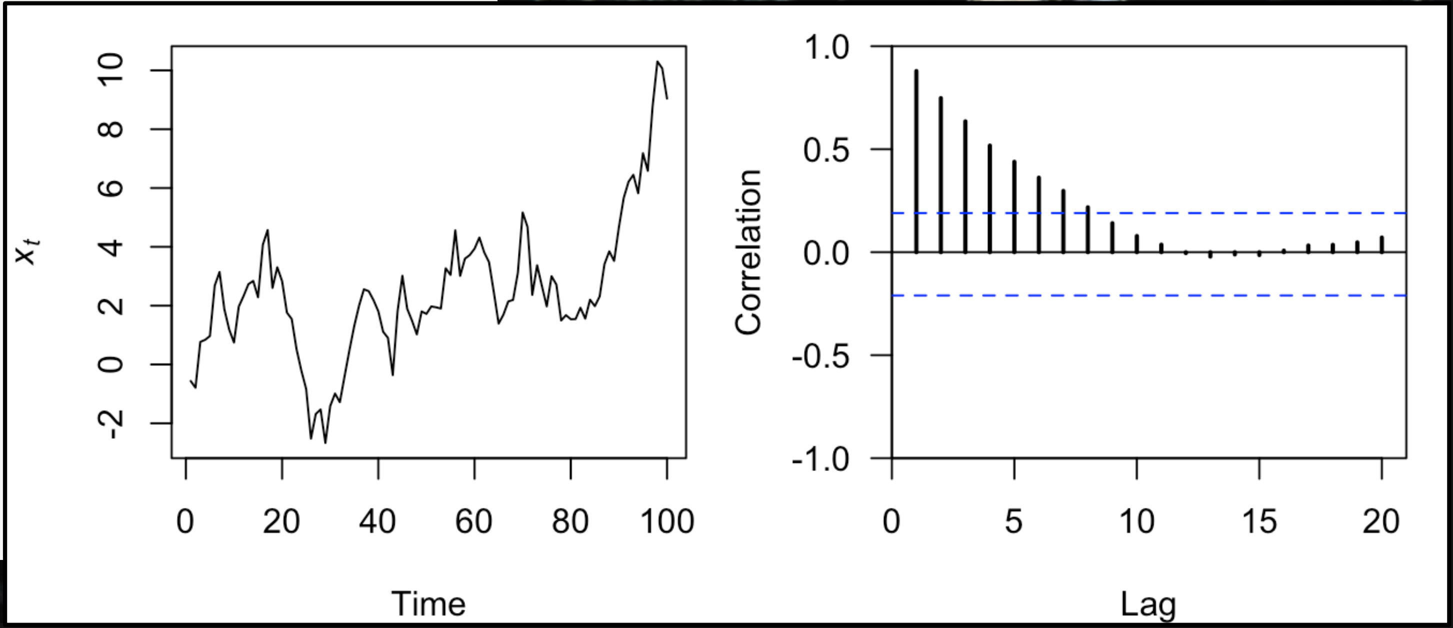


Temporal Autocorrelation

Temporal autocorrelation occurs when observations close in time are more similar to each other than observations farther apart.

Problem: Temporally autocorrelated data violate assumptions of independence!!!



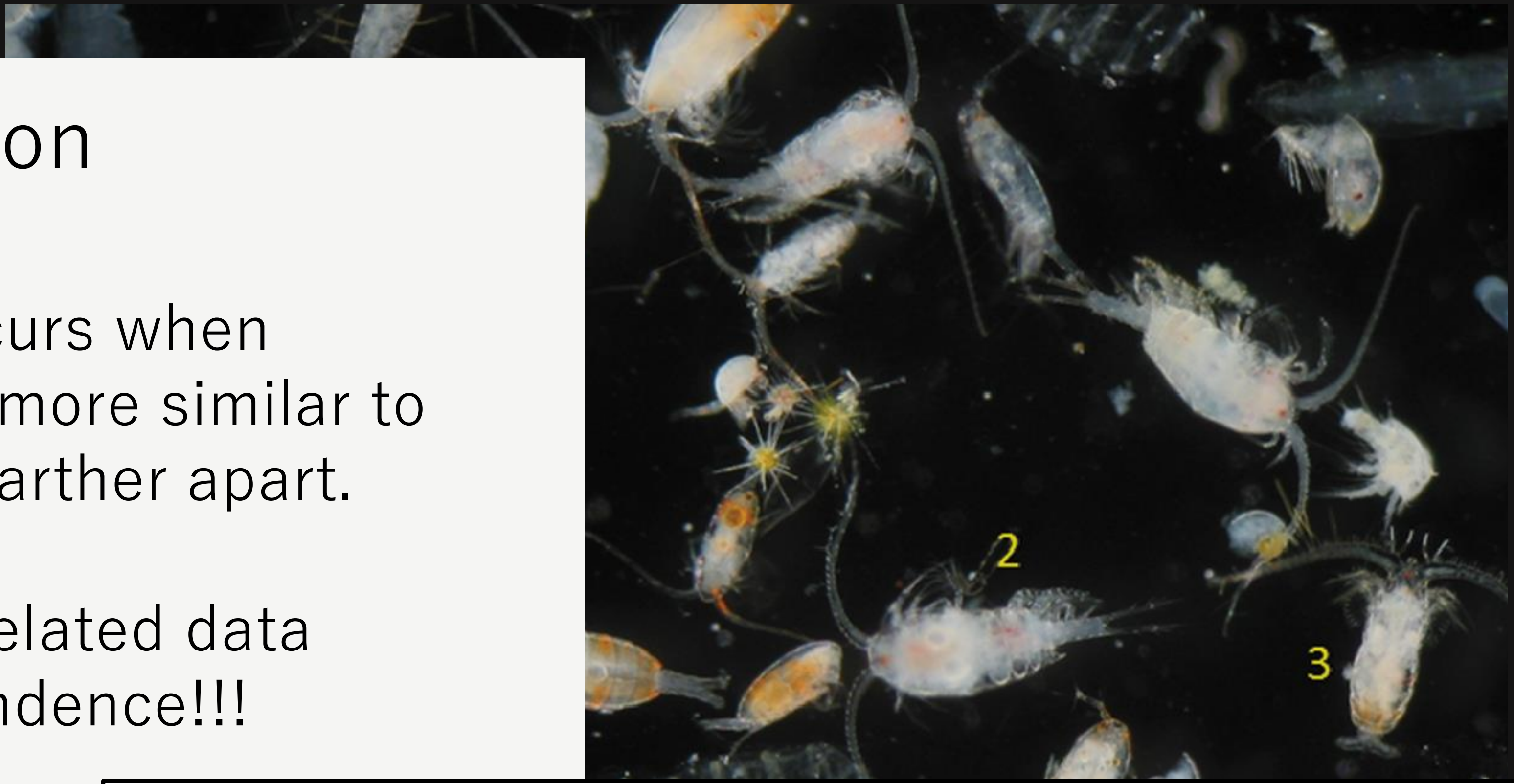


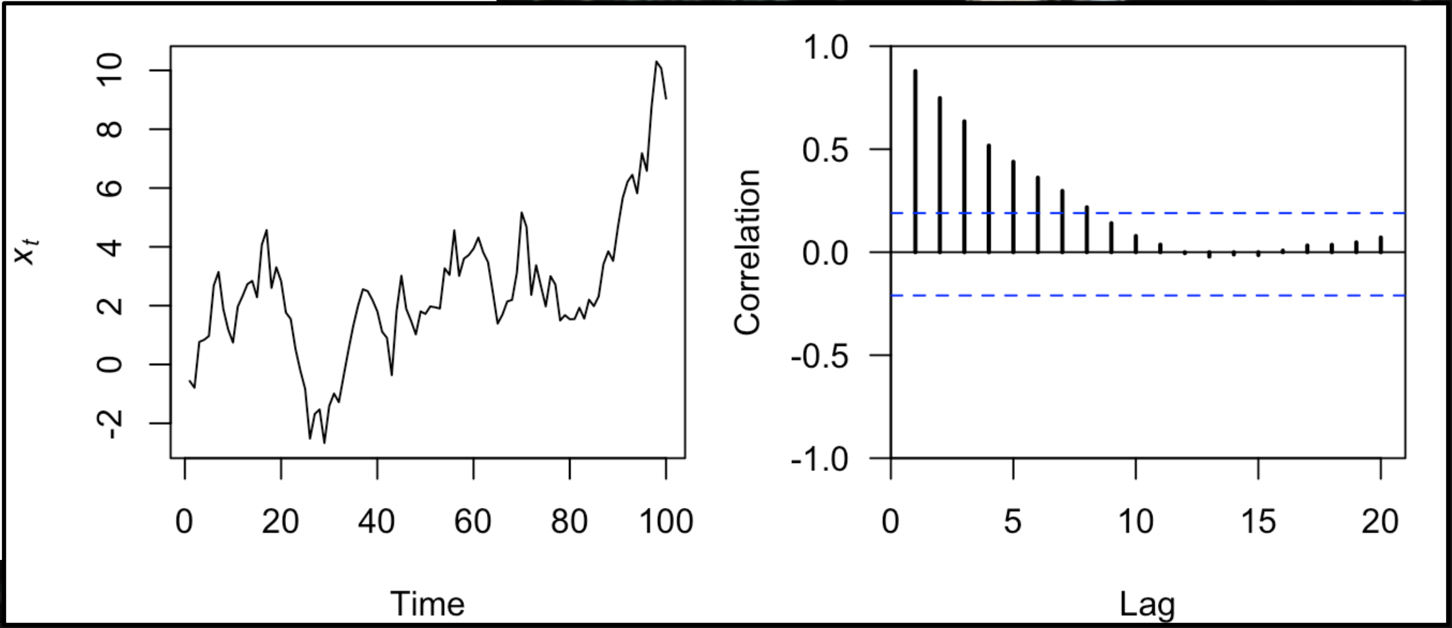
Temporal Autocorrelation

Temporal autocorrelation occurs when observations close in time are more similar to each other than observations farther apart.

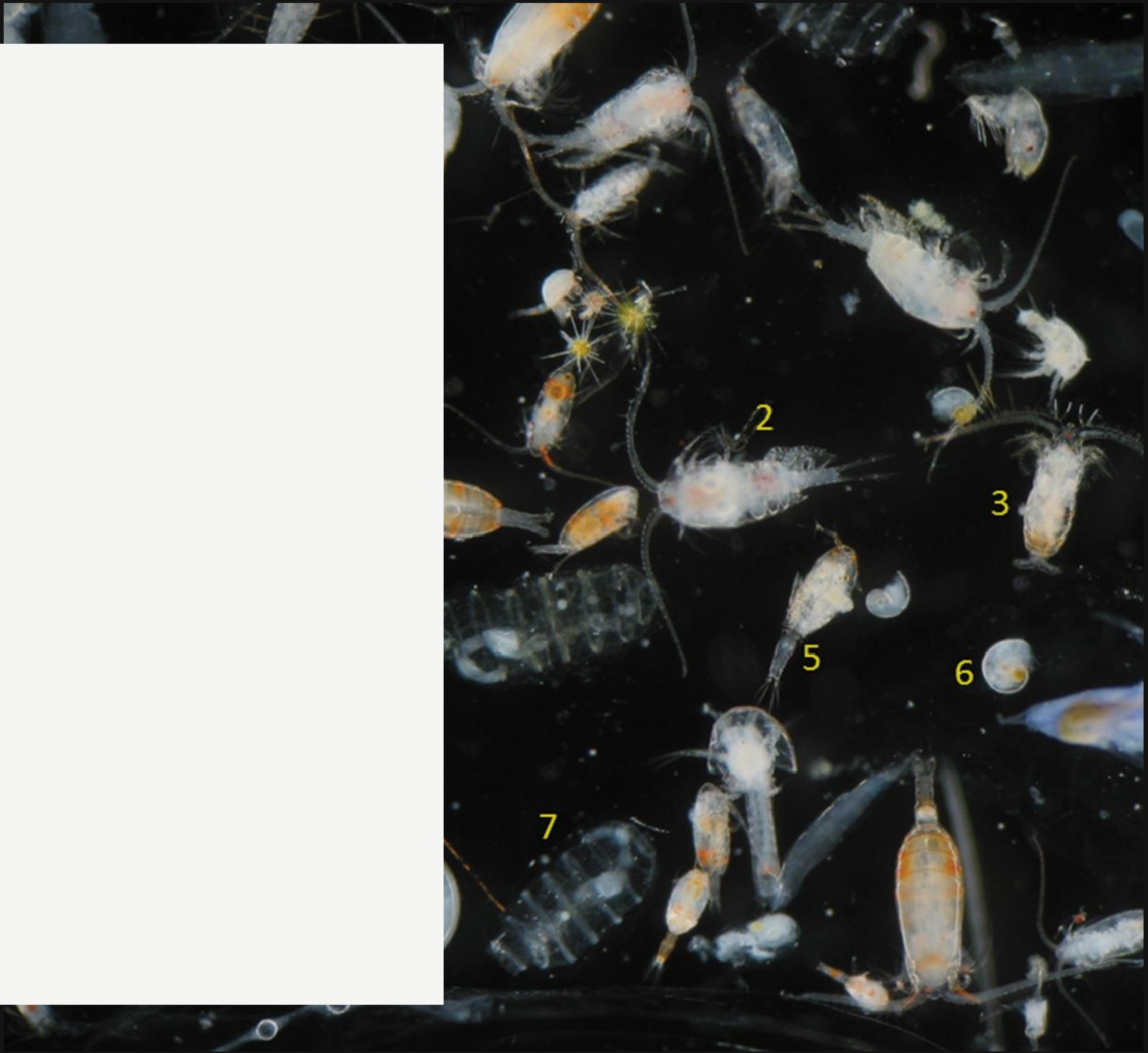
Problem: Temporally autocorrelated data violate assumptions of independence!!!

- Increases Type I Error
- Biased parameter estimates
- Underestimation of S.E.





Time Series Models

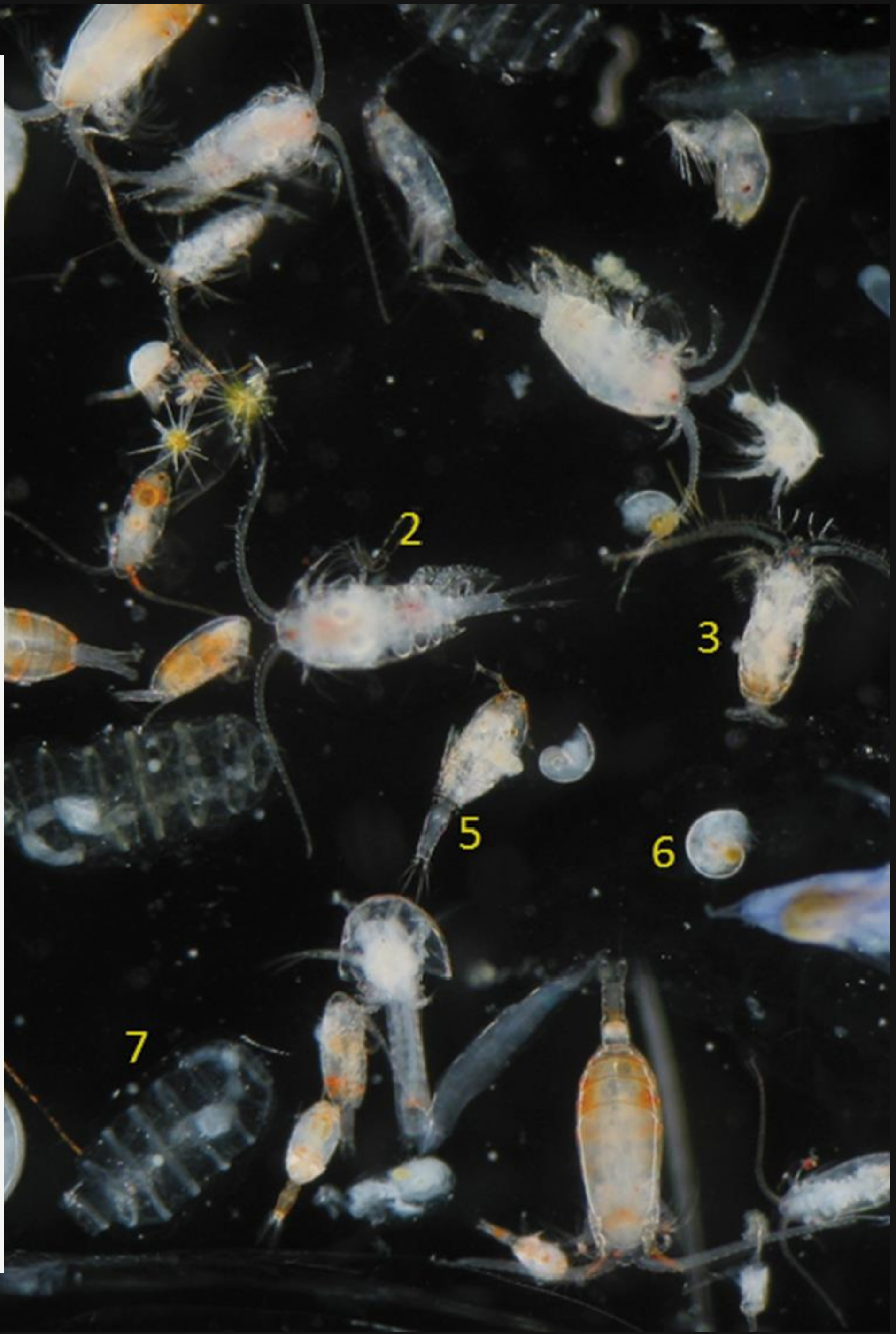


Time Series Models

Stochastic linear models model time series such that each data point is related to the previous data point(s) plus some independent, random "shock" (a).

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \ldots + \phi_q \tilde{y}_{t-q} + a_t$$

Also sometimes referred to as "**random walk.**"



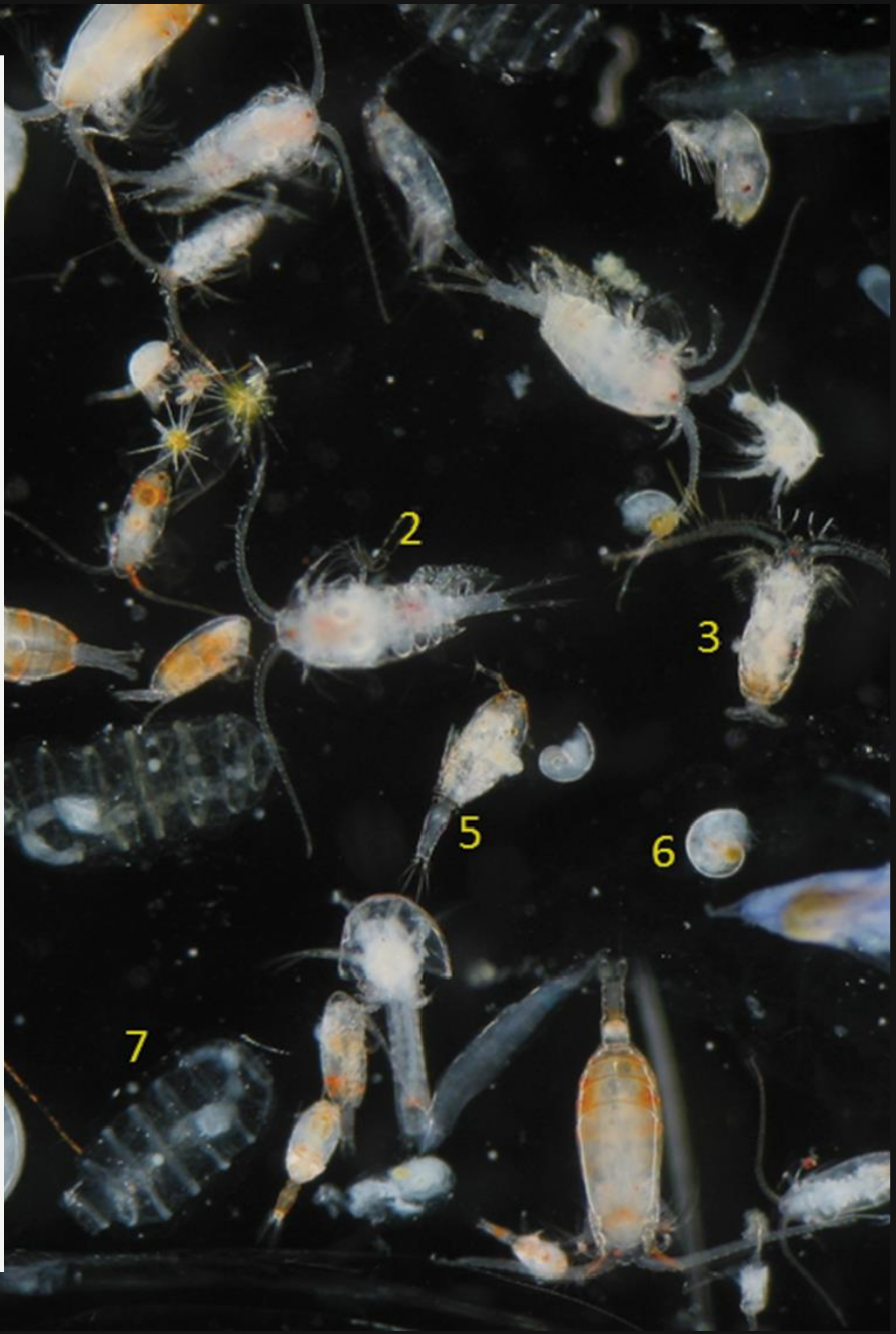
Time Series Models

Autoregressive (AR) models predict the current value in a time series based on past values.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

where:

- X_t is the value of the series at time t,
- $\phi_1, \phi_2, \ldots, \phi_p$ are parameters (coefficients) to be estimated,
- ϵ_t is white noise (random error term).



Time Series Models

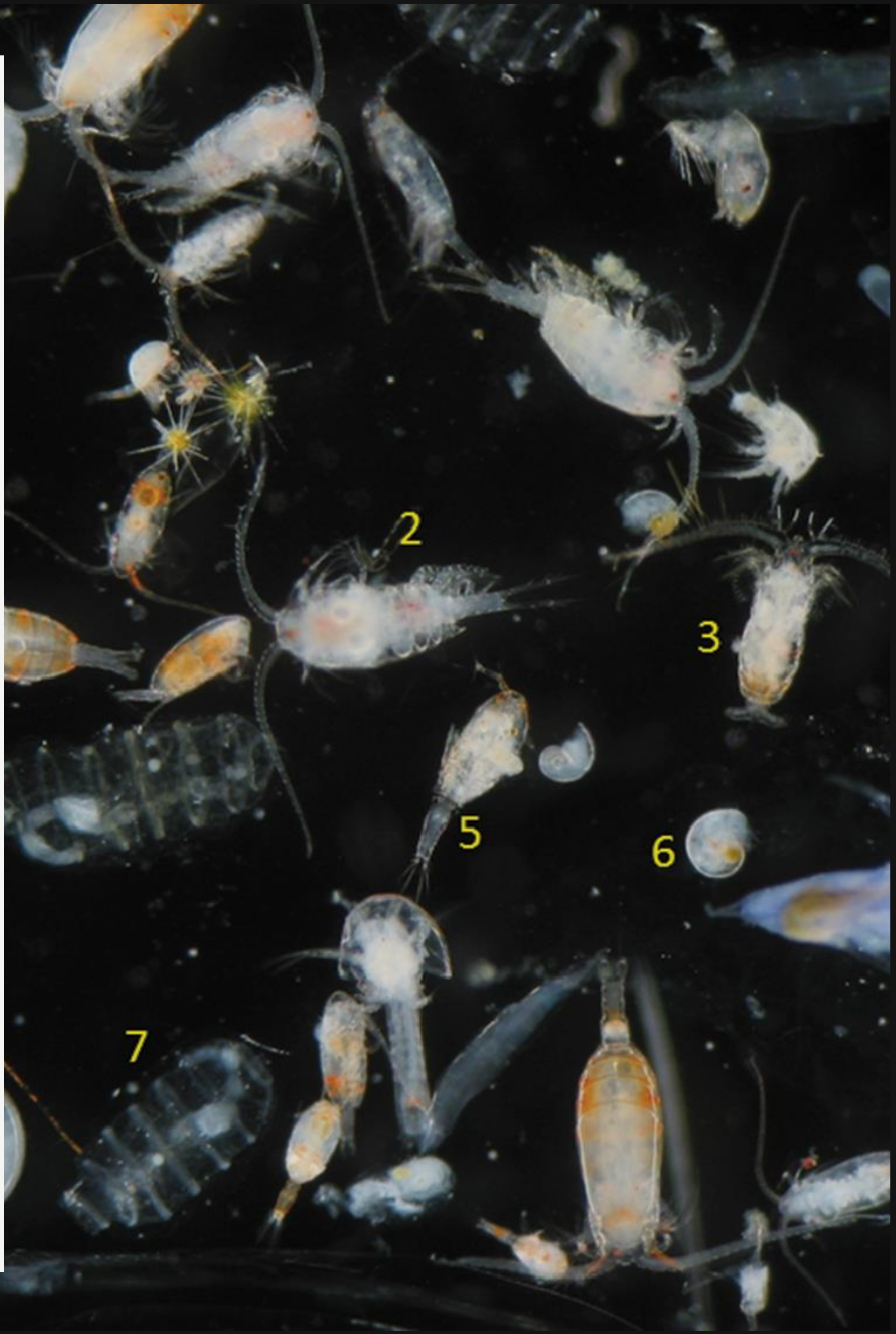
Autoregressive (AR) models predict the current value in a time series based on past values.

Moving Average (MA) models model the series as a linear function of past errors (i.e., shocks).

$$X_t = \mu + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

where:

- μ is the mean of the series,
- $\theta_1, \theta_2, \ldots, \theta_q$ are parameters to be estimated,
- $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q}$ are the previous error terms (random shocks).

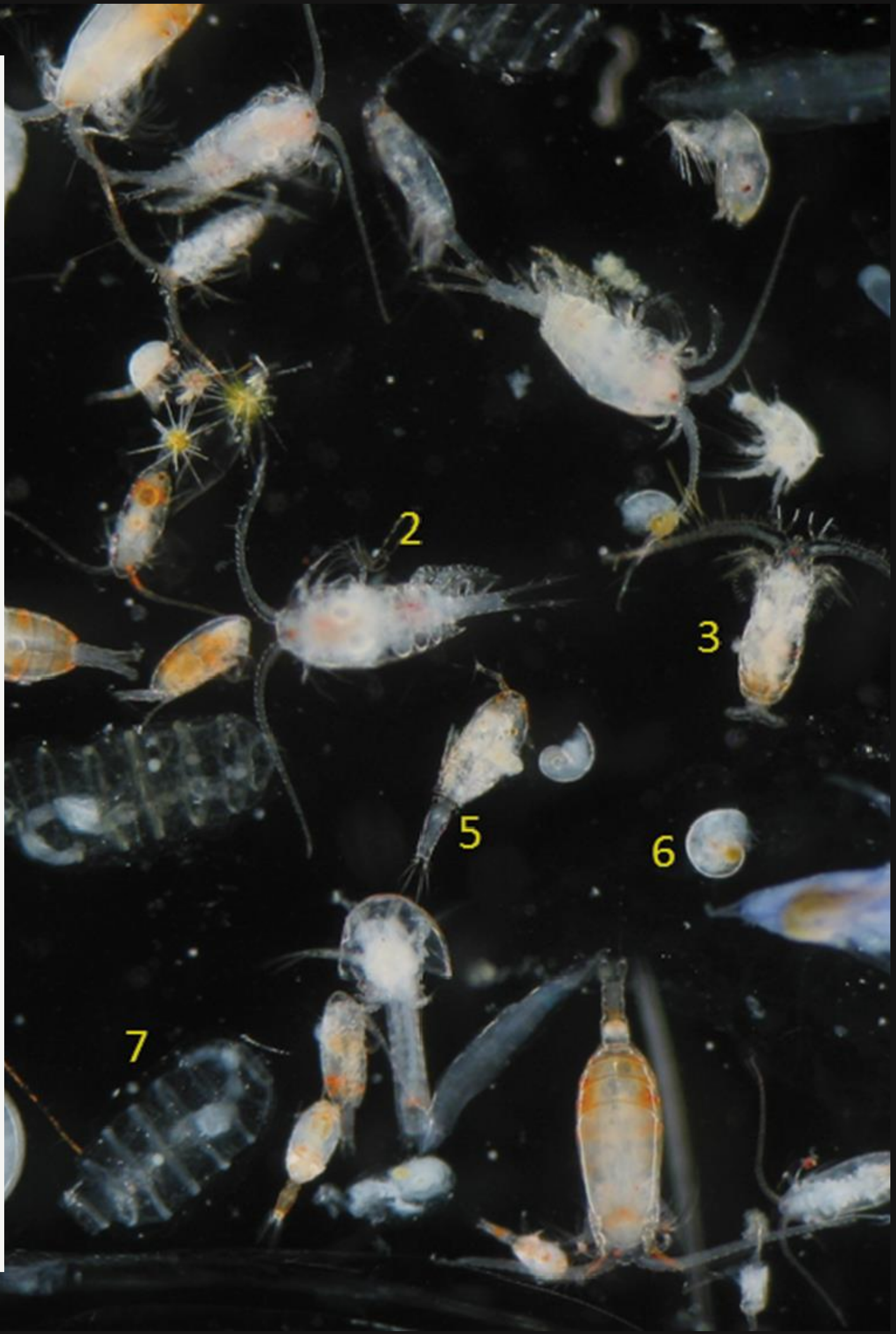


Time Series Models

Autoregressive (AR) models predict the current value in a time series based on past values.

Moving Average (MA) models model the series as a linear function of past errors (i.e., shocks).

Autoregressive Moving Average (ARMA) models combine the AR and MA components to capture the autoregressive and moving average structures in the data.



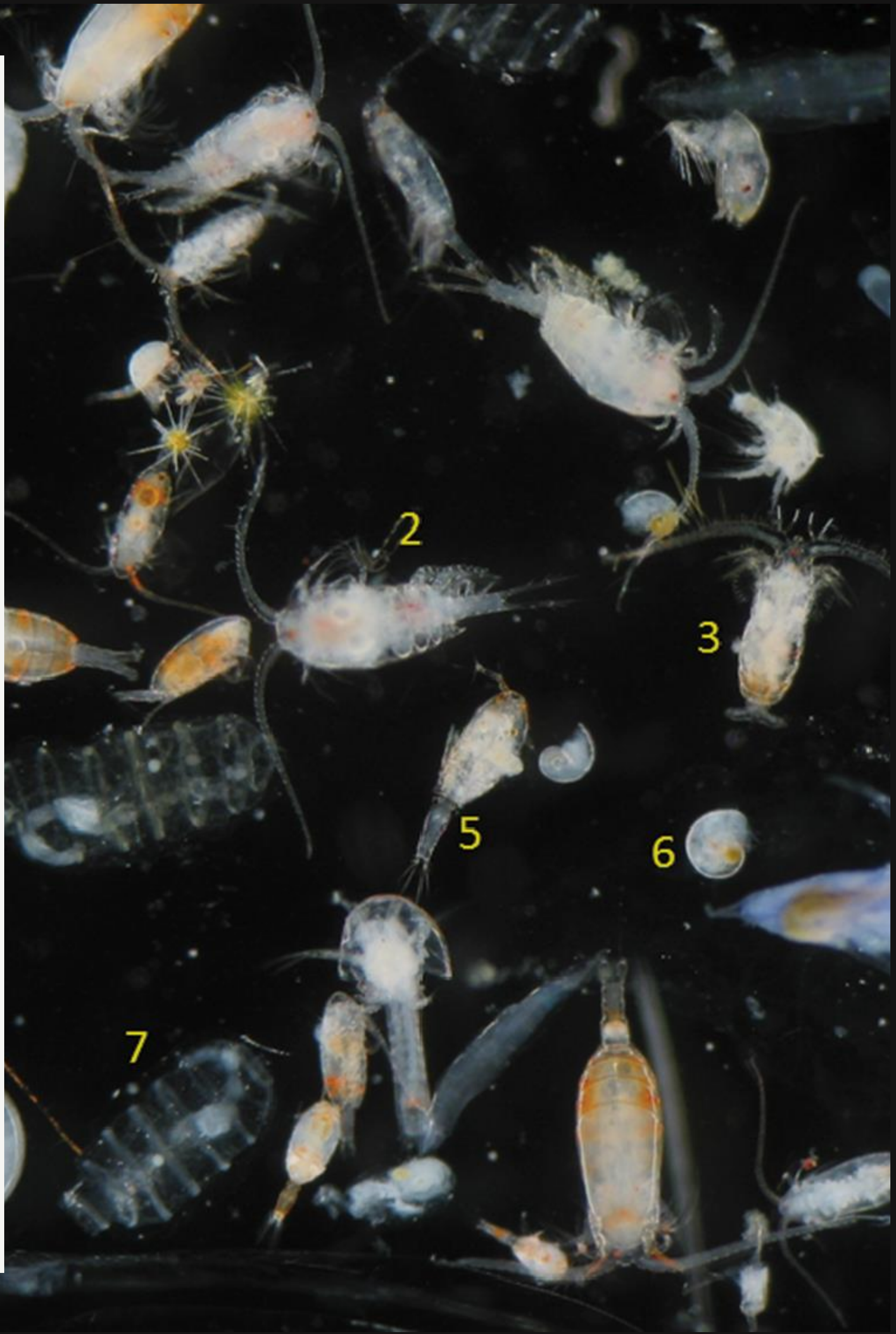
Time Series Models

Autoregressive (AR) models predict the current value in a time series based on past values.

Moving Average (MA) models model the series as a linear function of past errors (i.e., shocks).

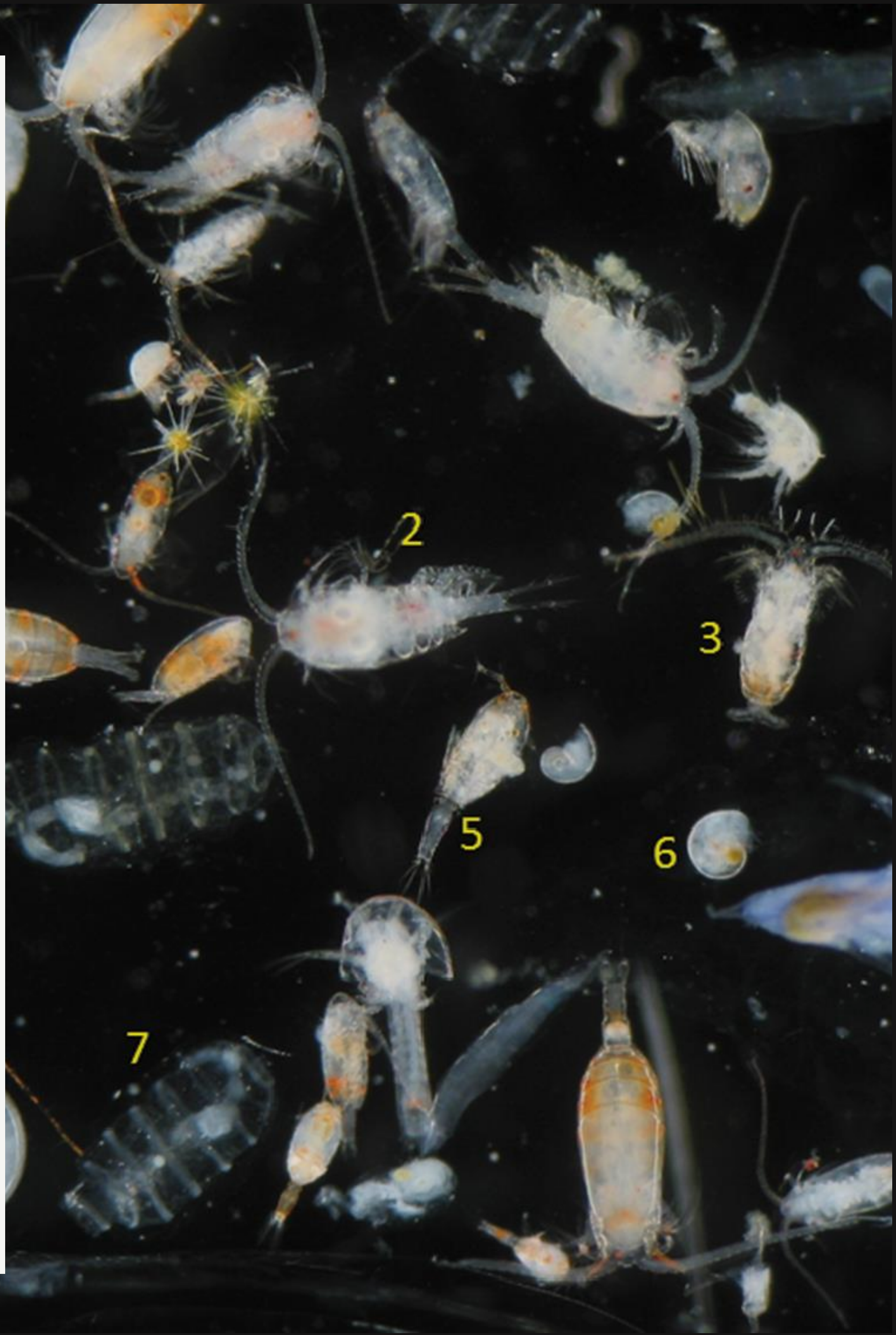
Autoregressive Moving Average (ARMA) models combine the AR and MA components to capture the autoregressive and moving average structures in the data.

Autoregressive Integrated Moving Average Models (ARIMA) differences the data to make it stationary before applying the ARMA model.



Time Series Models

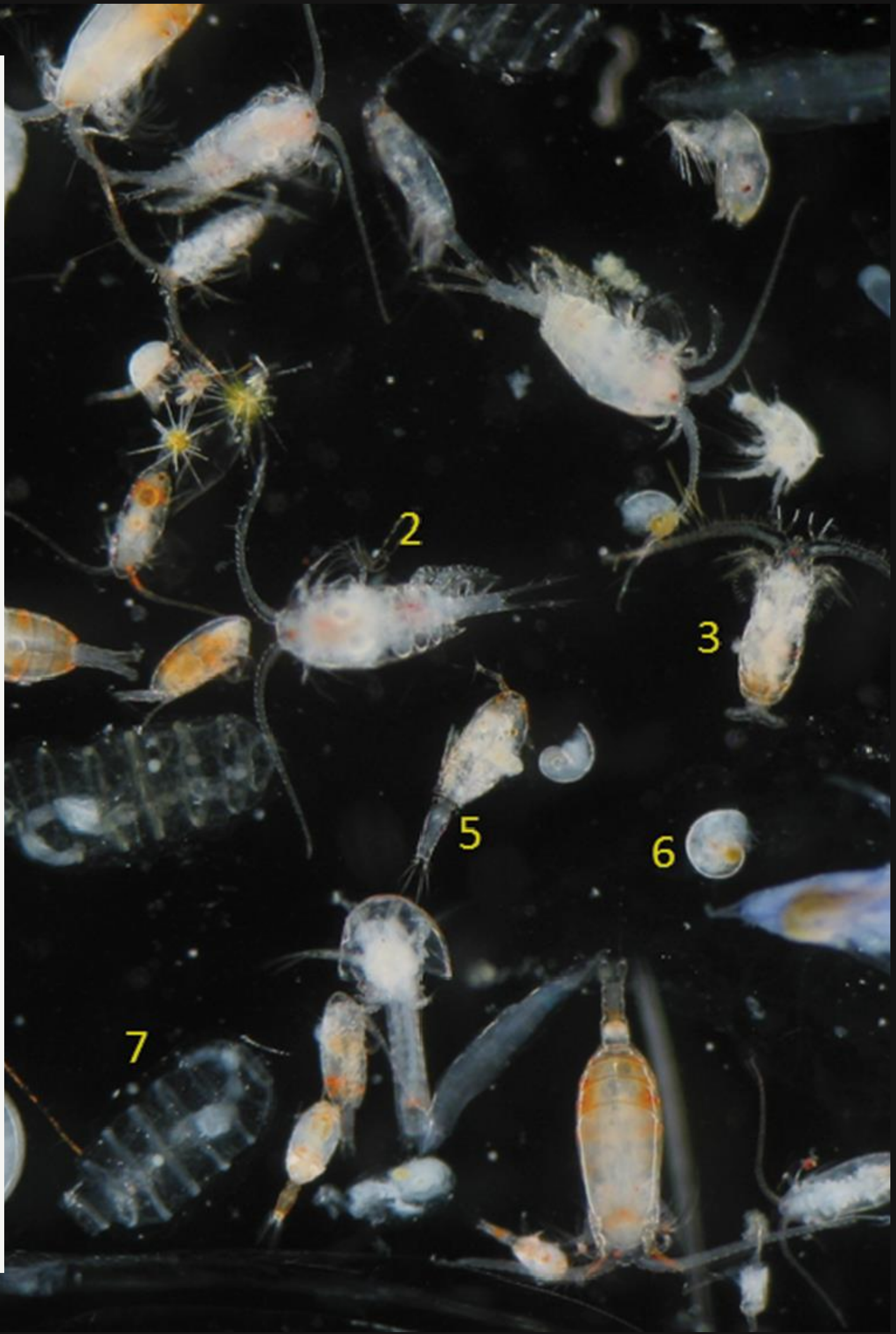
The benefit of an **ARIMA** model is that it can handle non-stationary (i.e., not detrended) data!



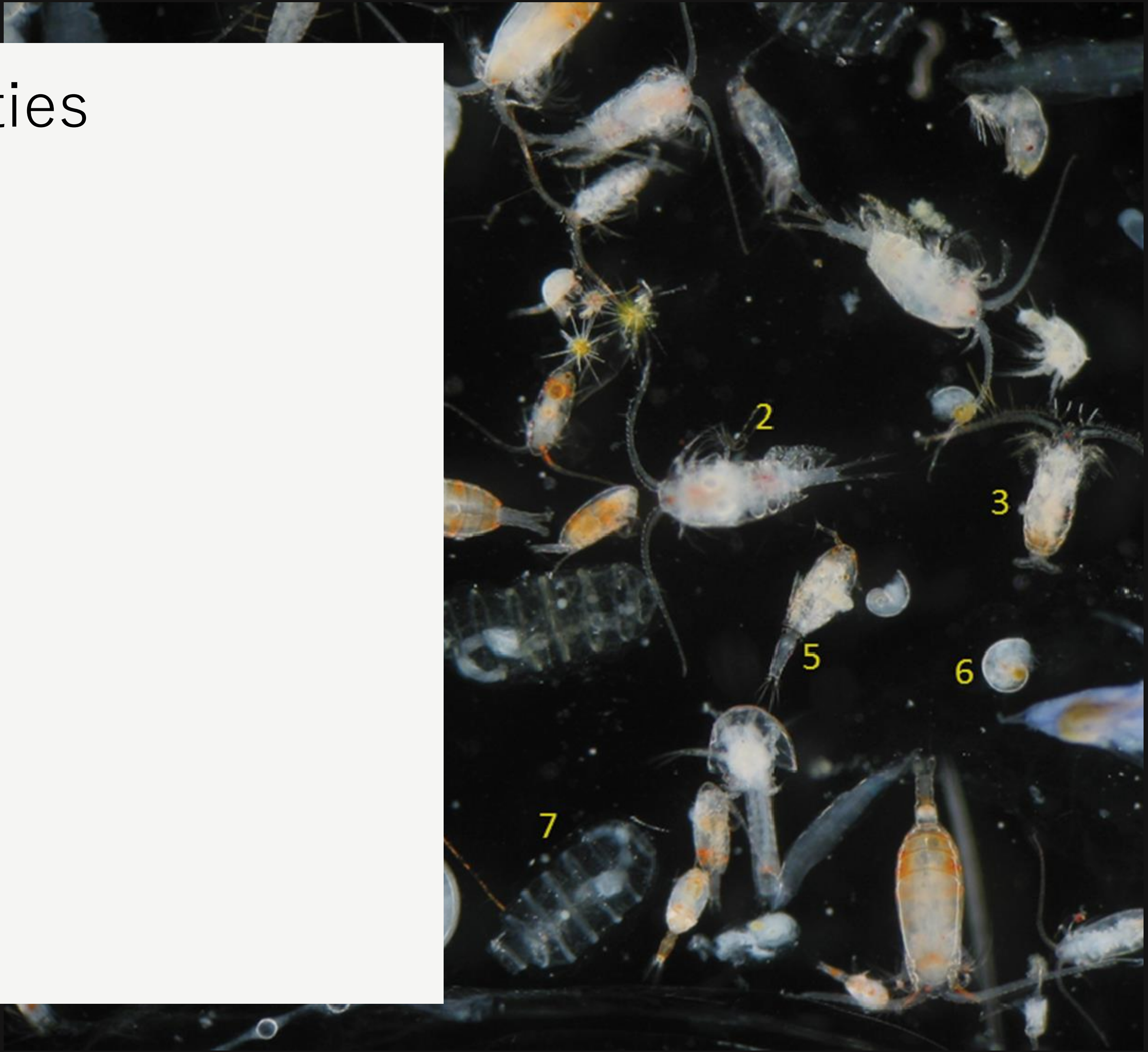
Time Series Models

The benefit of an **ARIMA** model is that it can handle non-stationary (i.e., not detrended) data!

Once the parameters are estimated and the residuals are checked for independence, the model can be used to **forecast** past the last observation.

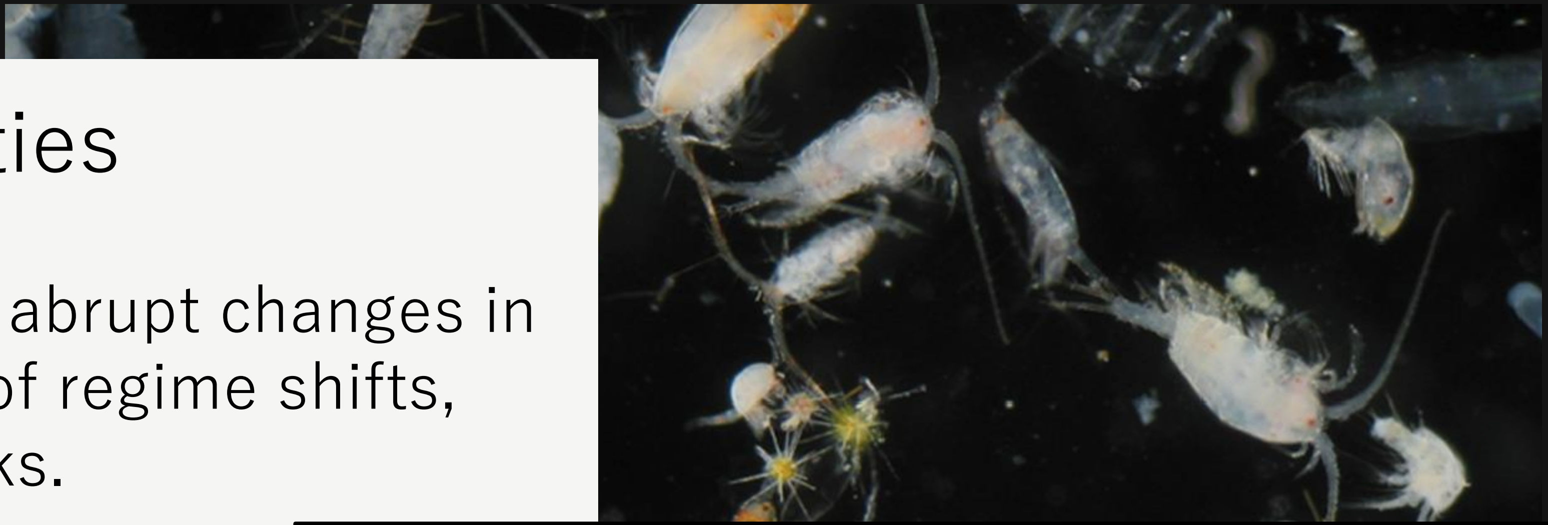


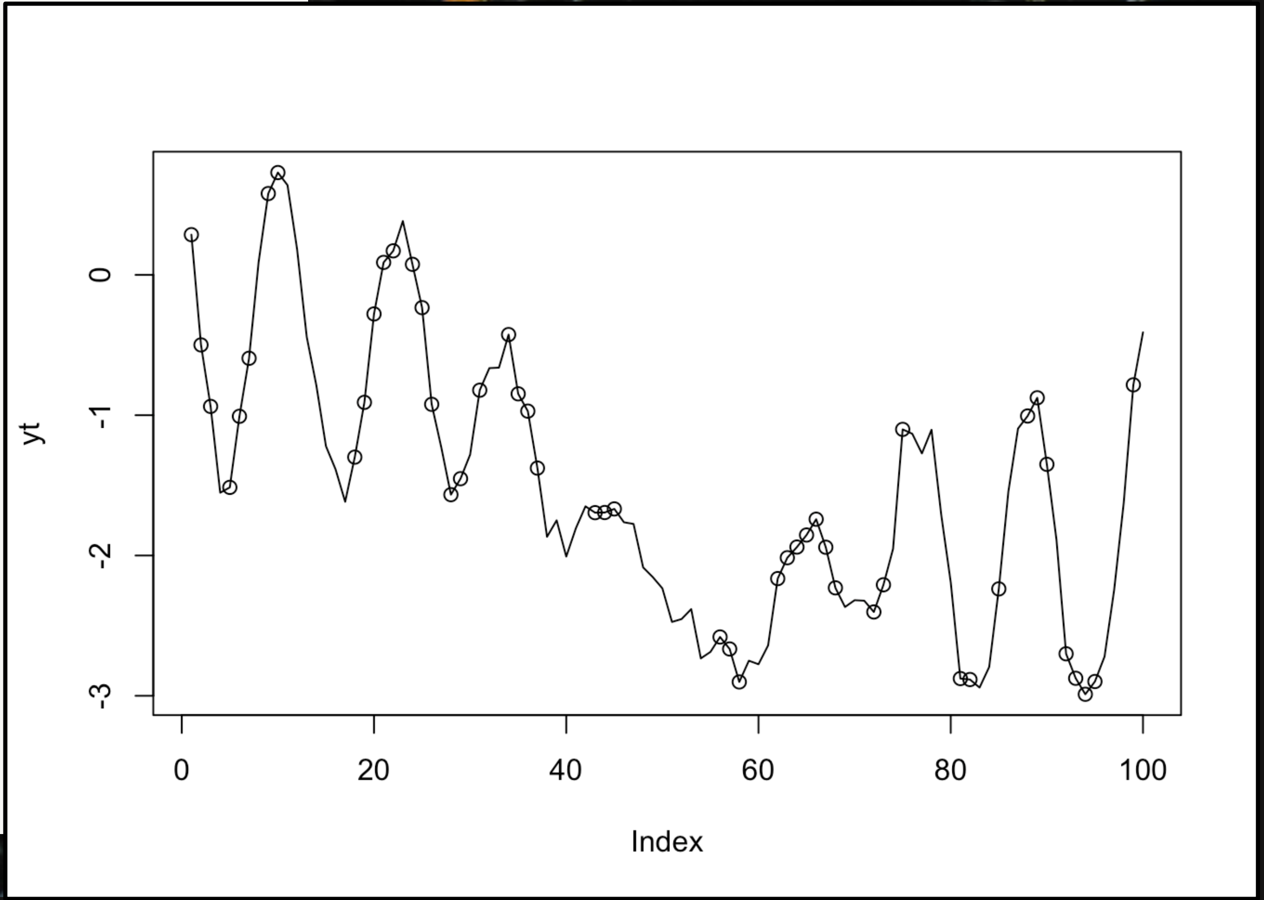
Identifying Discontinuities



Identifying Discontinuities

Discontinuities are sudden or abrupt changes in a time series, often indicative of regime shifts, disturbances, or external shocks.

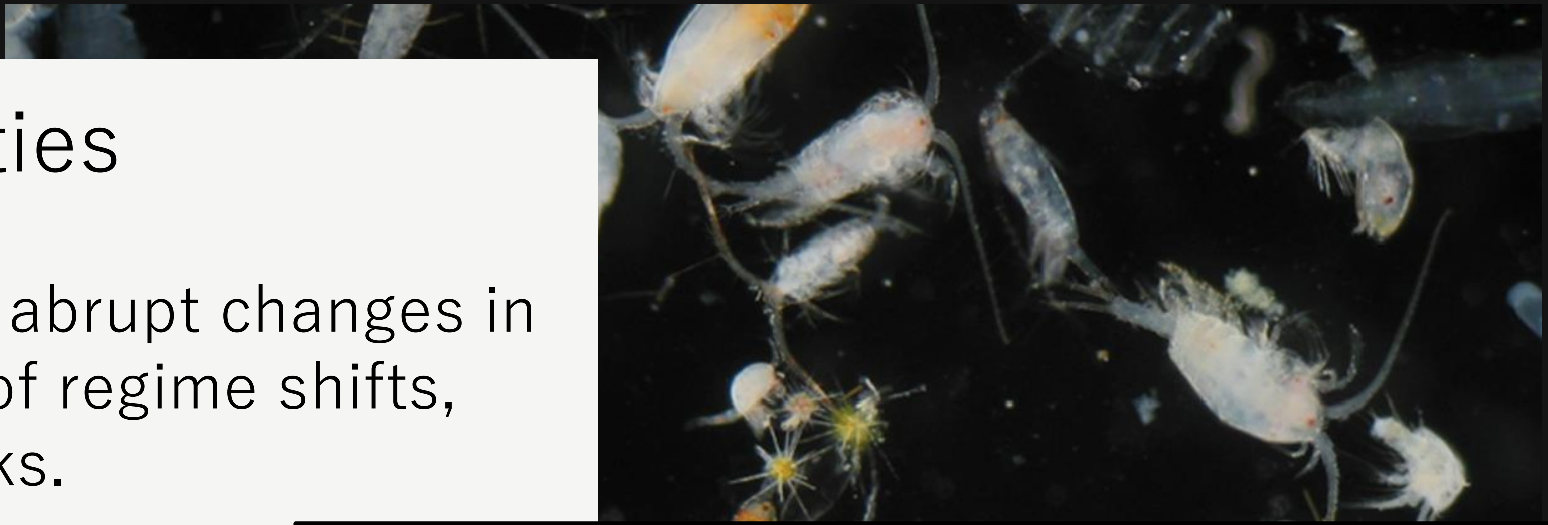


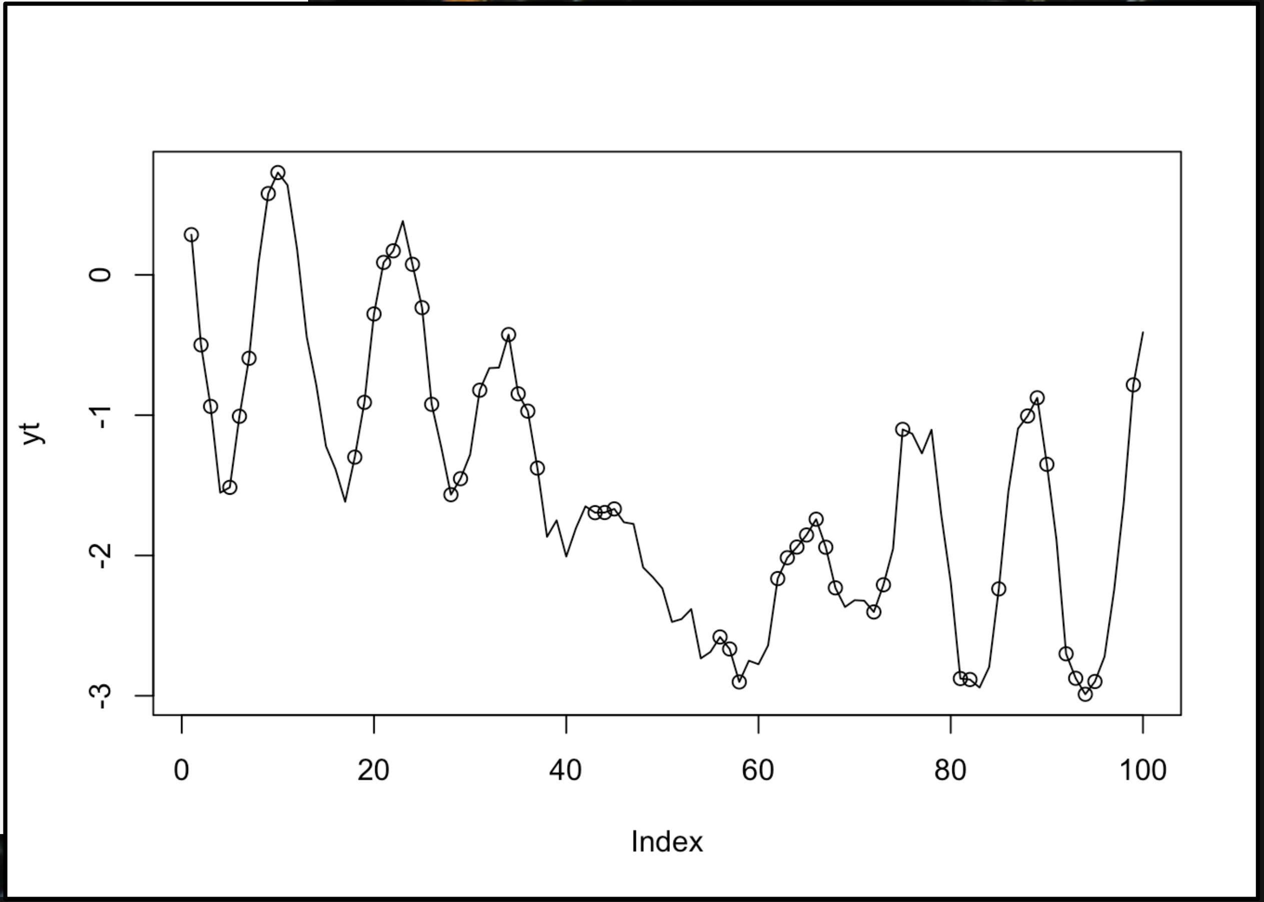


Identifying Discontinuities

Discontinuities are sudden or abrupt changes in a time series, often indicative of regime shifts, disturbances, or external shocks.

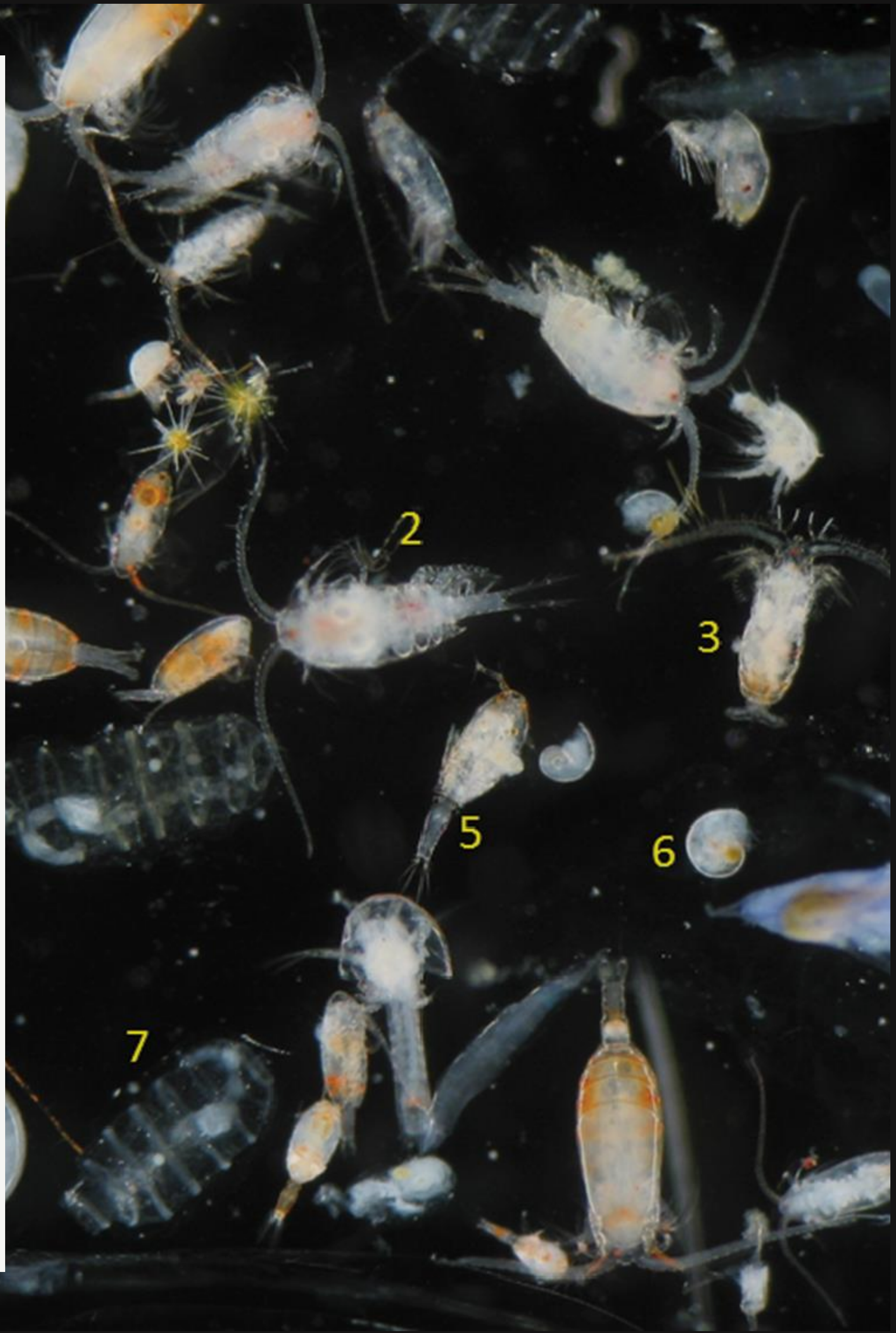
For example:
- Diet shifts
- Climate shifts
- Species migrations





Identifying Discontinuities

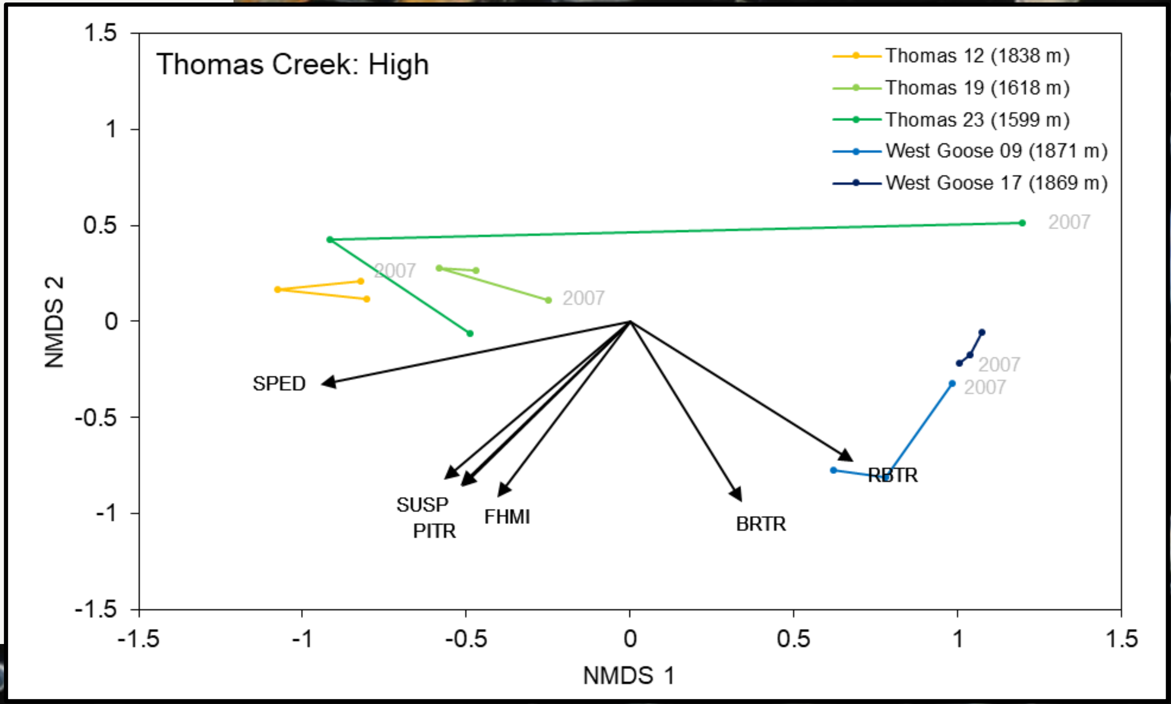
Ways of identifying discontinuities in multivariate data:



Identifying Discontinuities

Ways of identifying discontinuities in multivariate data:

1) **Ordination in reduced space** – large jumps in two-dimensional space are interpreted as discontinuities.

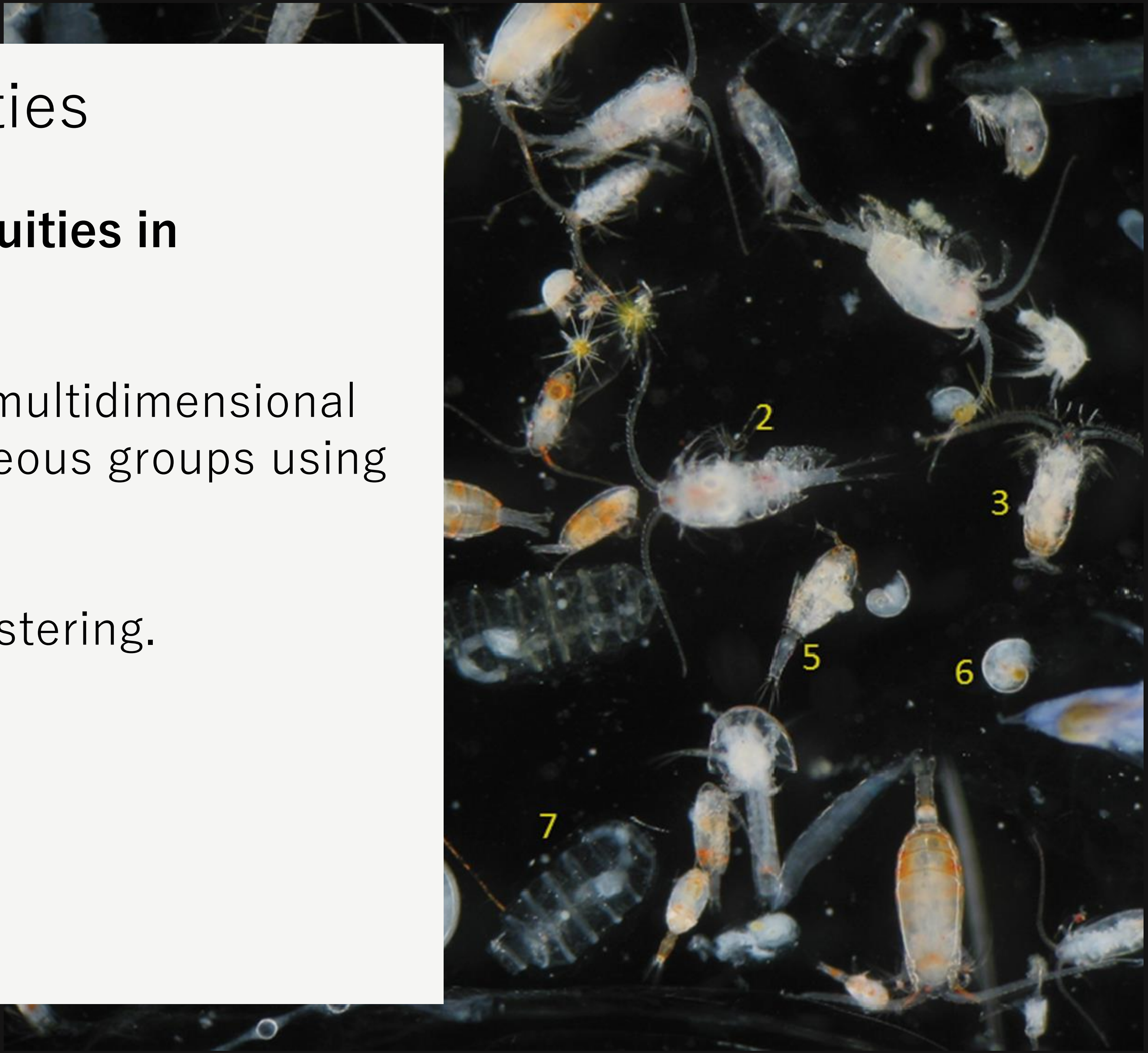


Identifying Discontinuities

Ways of identifying discontinuities in multivariate data:

2) Segmenting data series – multidimensional data partitioned into homogeneous groups using objective clustering criterion.

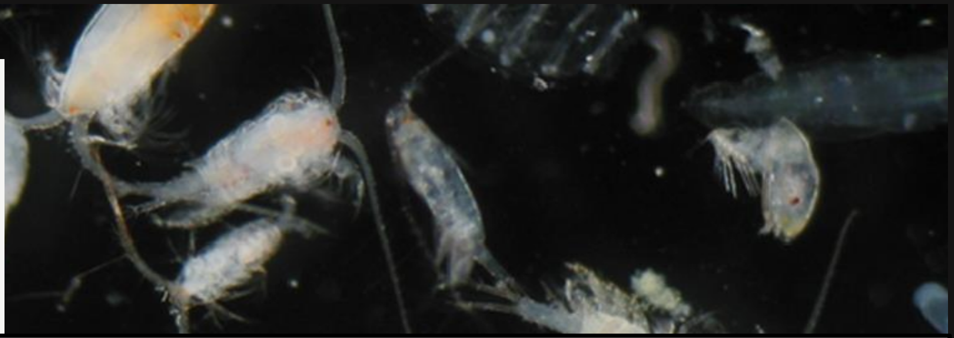
Same concept as K-means clustering.

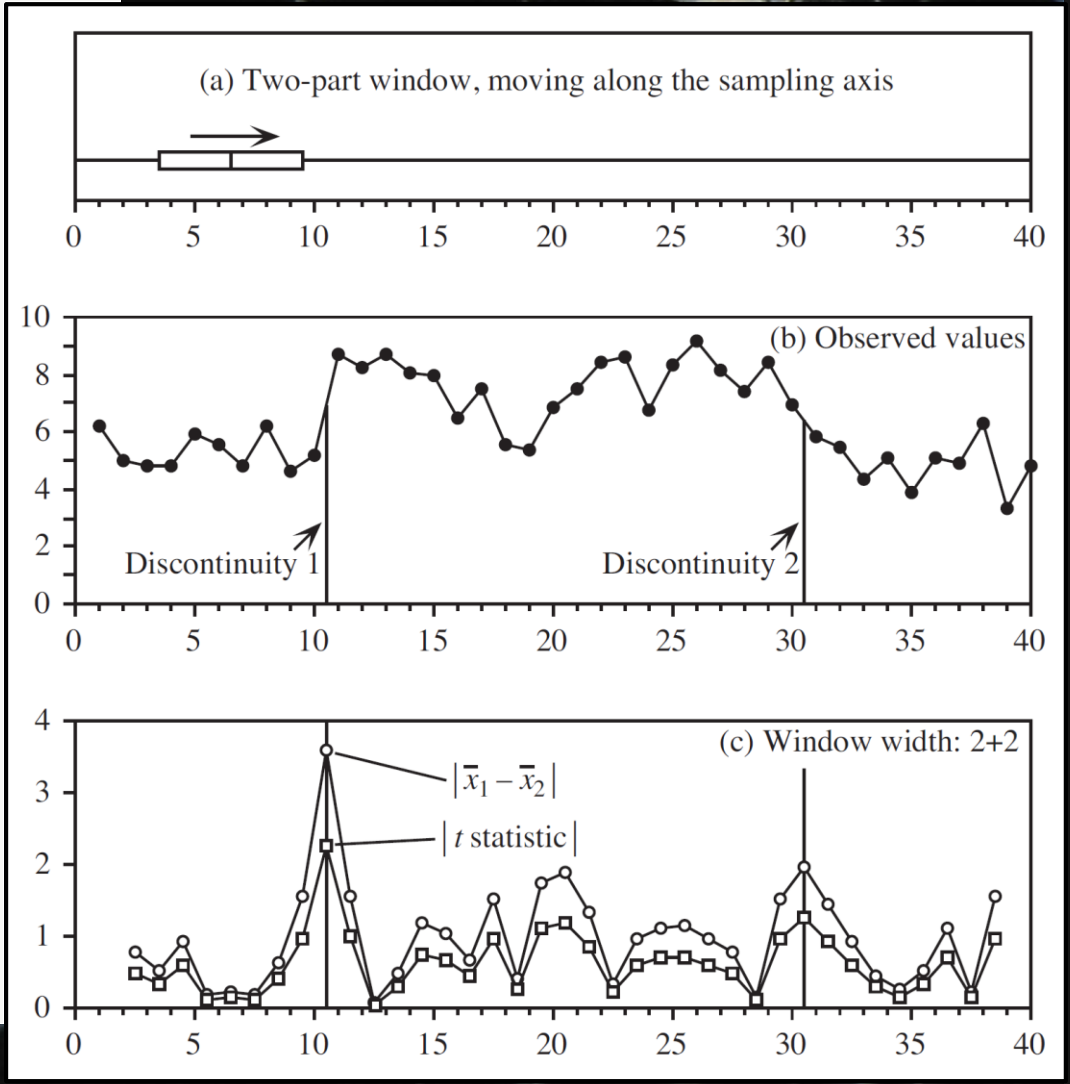


Identifying Discontinuities

Ways of identifying discontinuities in multivariate data:

3) Webster's "moving window" method – Mahalanobis distances are compared between the left and right hand side of the window.

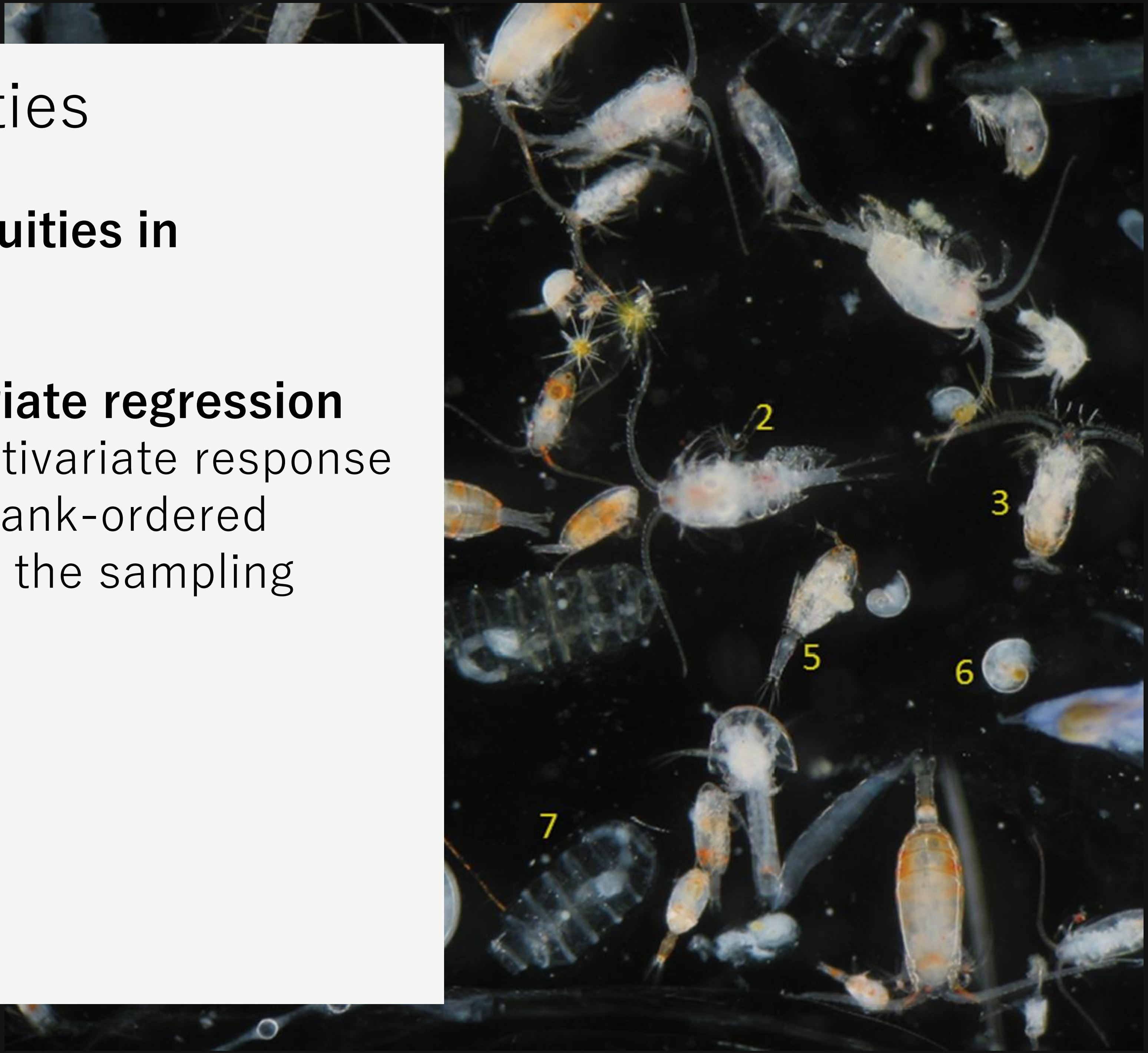




Identifying Discontinuities

Ways of identifying discontinuities in multivariate data:

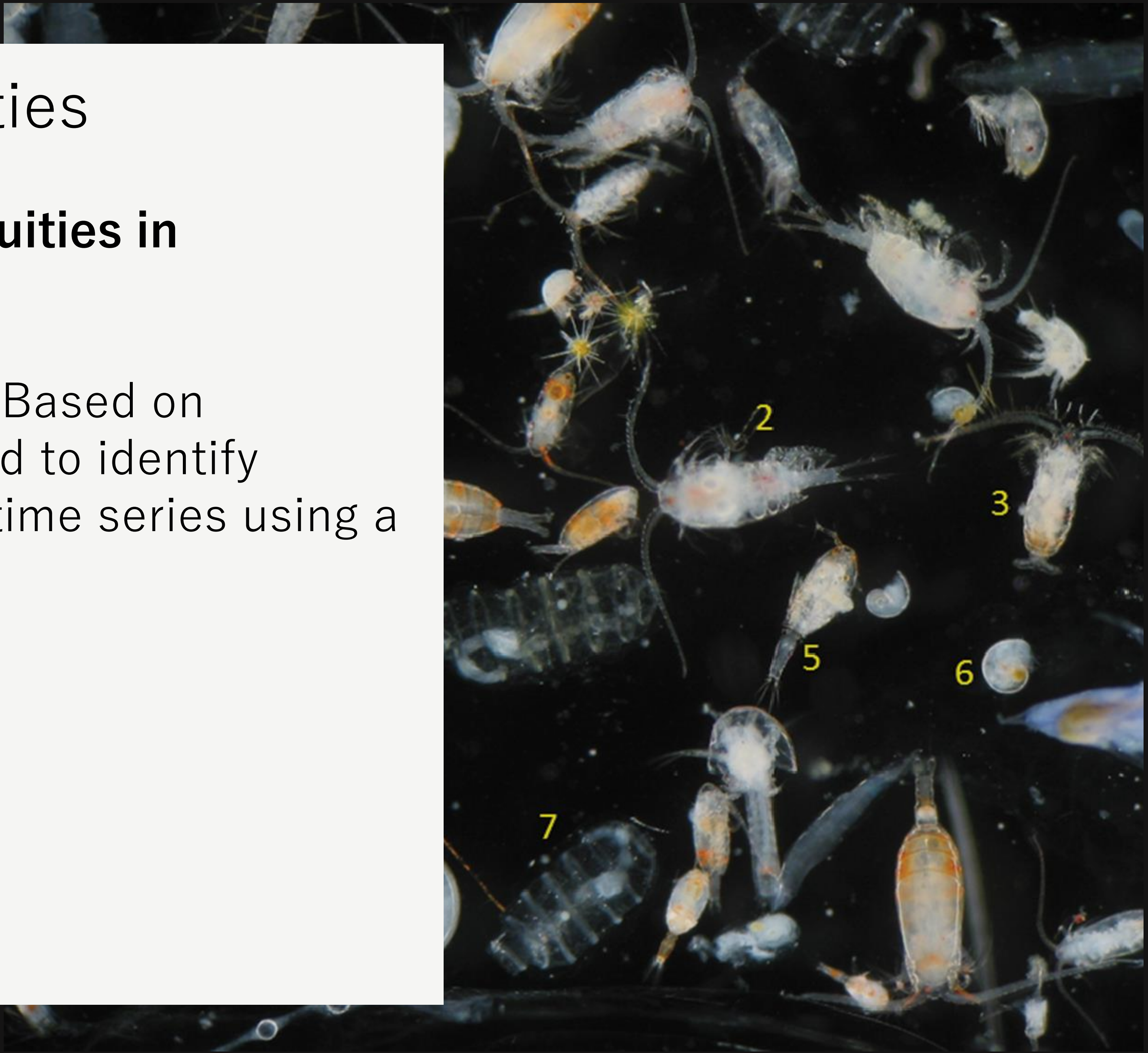
4) Time-constrained multivariate regression tree analysis – Analyzes a multivariate response matrix using a quantitative or rank-ordered predictor variable representing the sampling sequence through time.



Identifying Discontinuities

Ways of identifying discontinuities in multivariate data:

5) Chronological clustering – Based on hierarchical clustering and used to identify discontinuities in multivariate time series using a "contiguity constraint."



A great resource!

Applied Time Series Analysis for Fisheries and Environmental Sciences
E.E. Holmes, M.D. Scheuerell, and E.J. Ward

https://atsa-es.github.io/atsa-labs/

A great resource!

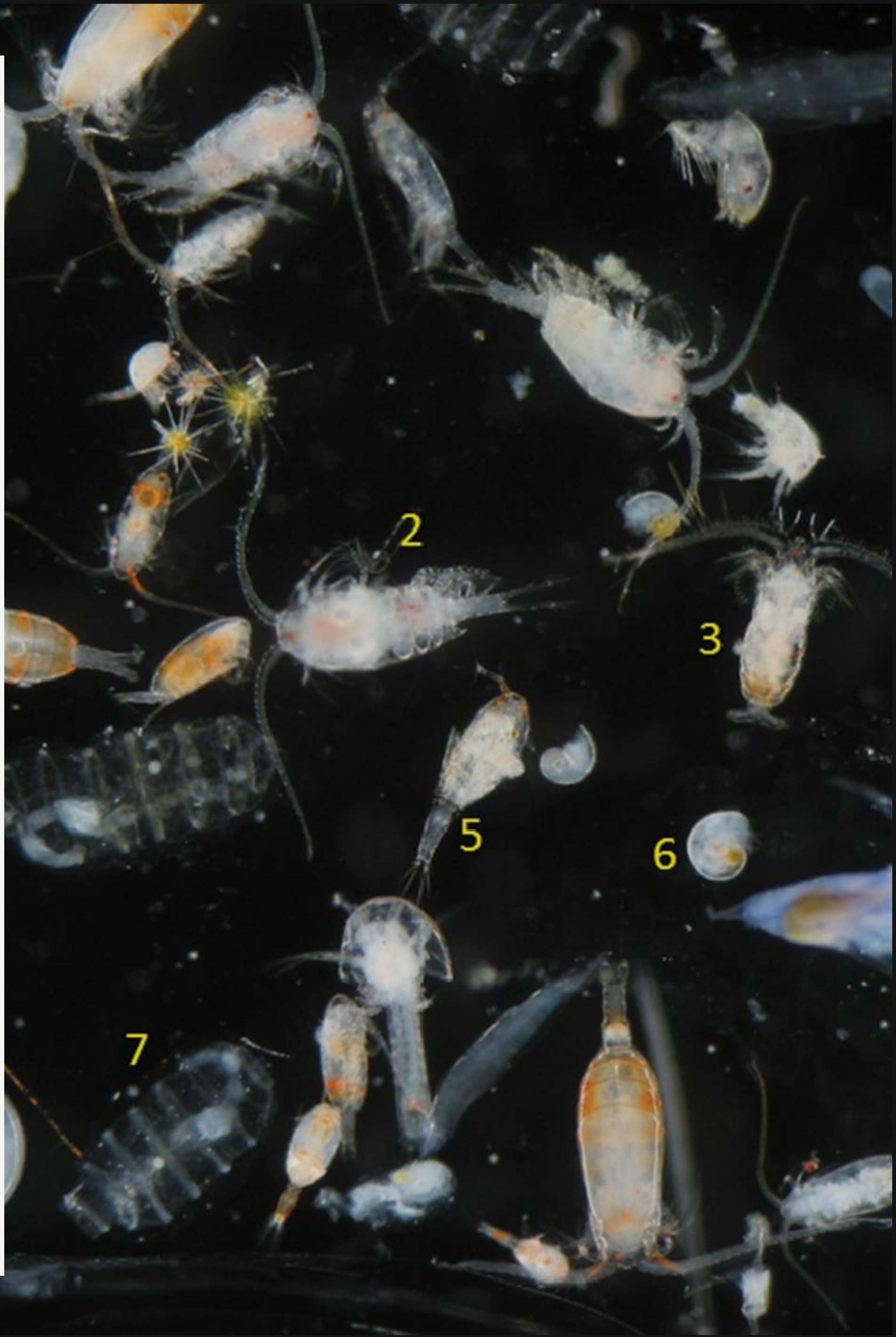
Applied Time Series Analysis for Fisheries and Environmental Sciences
E.E. Holmes, M.D. Scheuerell, and E.J. Ward

https://atsa-es.github.io/atsa-labs/

Also see `**glmmTMB**` package for a univariate option

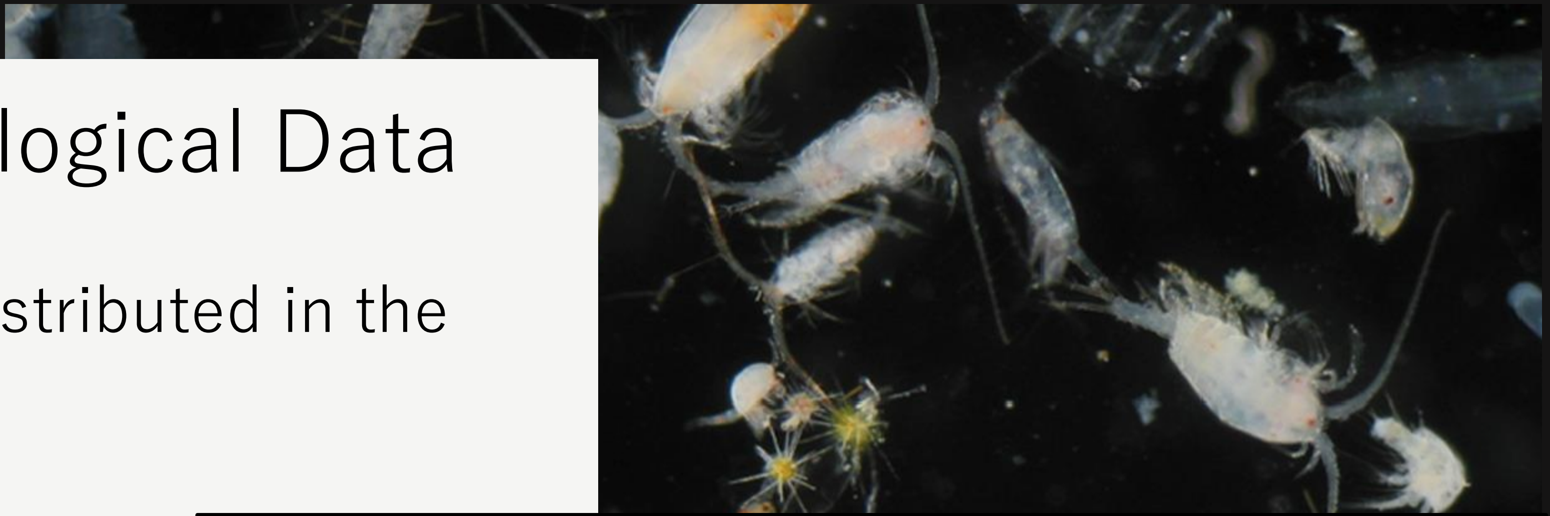
https://glmmtmb.github.io/glmmTMB/articles/covstruct.html

Spatial Patterns in Ecological Data



Spatial Patterns in Ecological Data

Organisms are not randomly distributed in the natural environment!

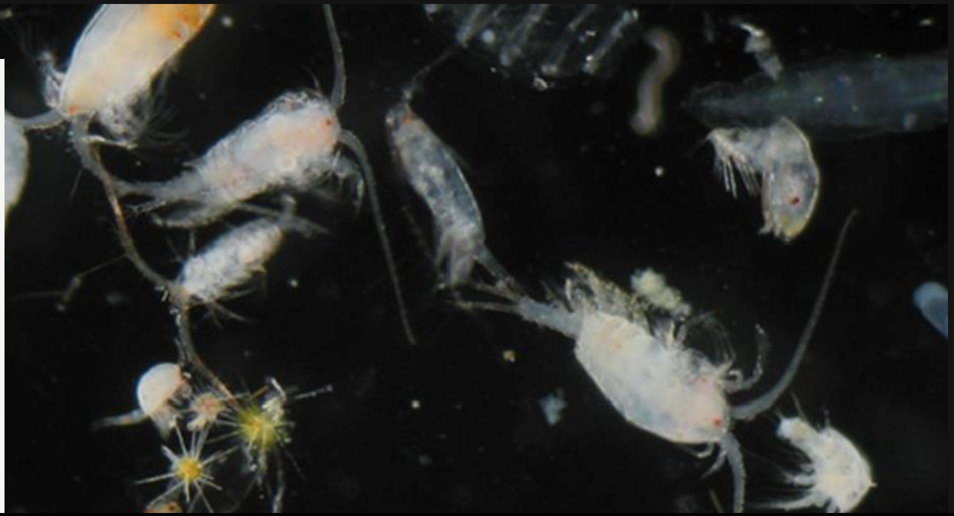


Spatial Patterns in Ecological Data

Organisms are not randomly distributed in the natural environment! **Processes like:**

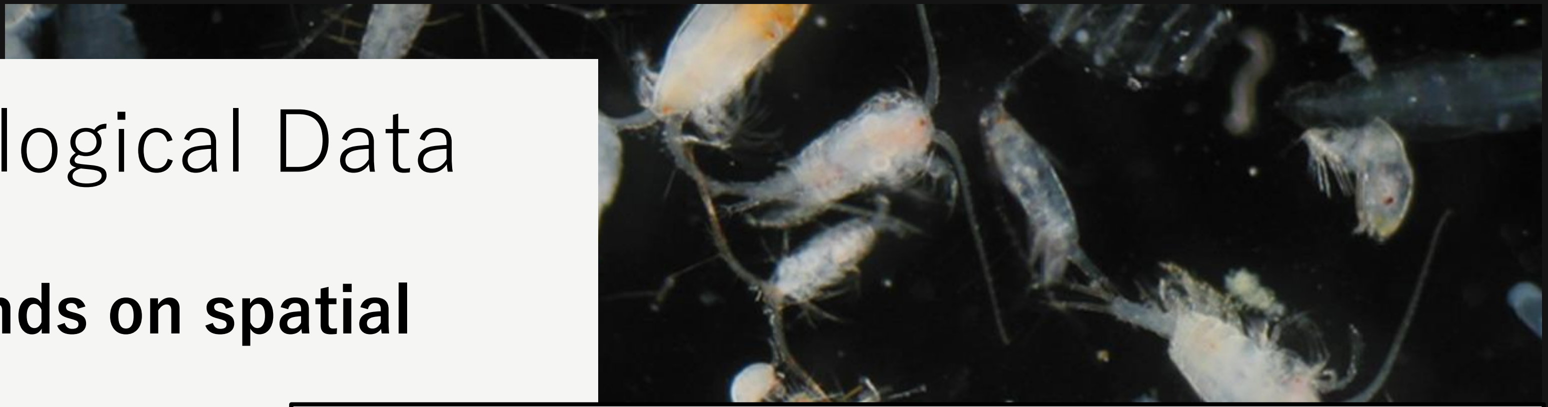
- Growth
- Dispersal
- Reproduction
- Mortality

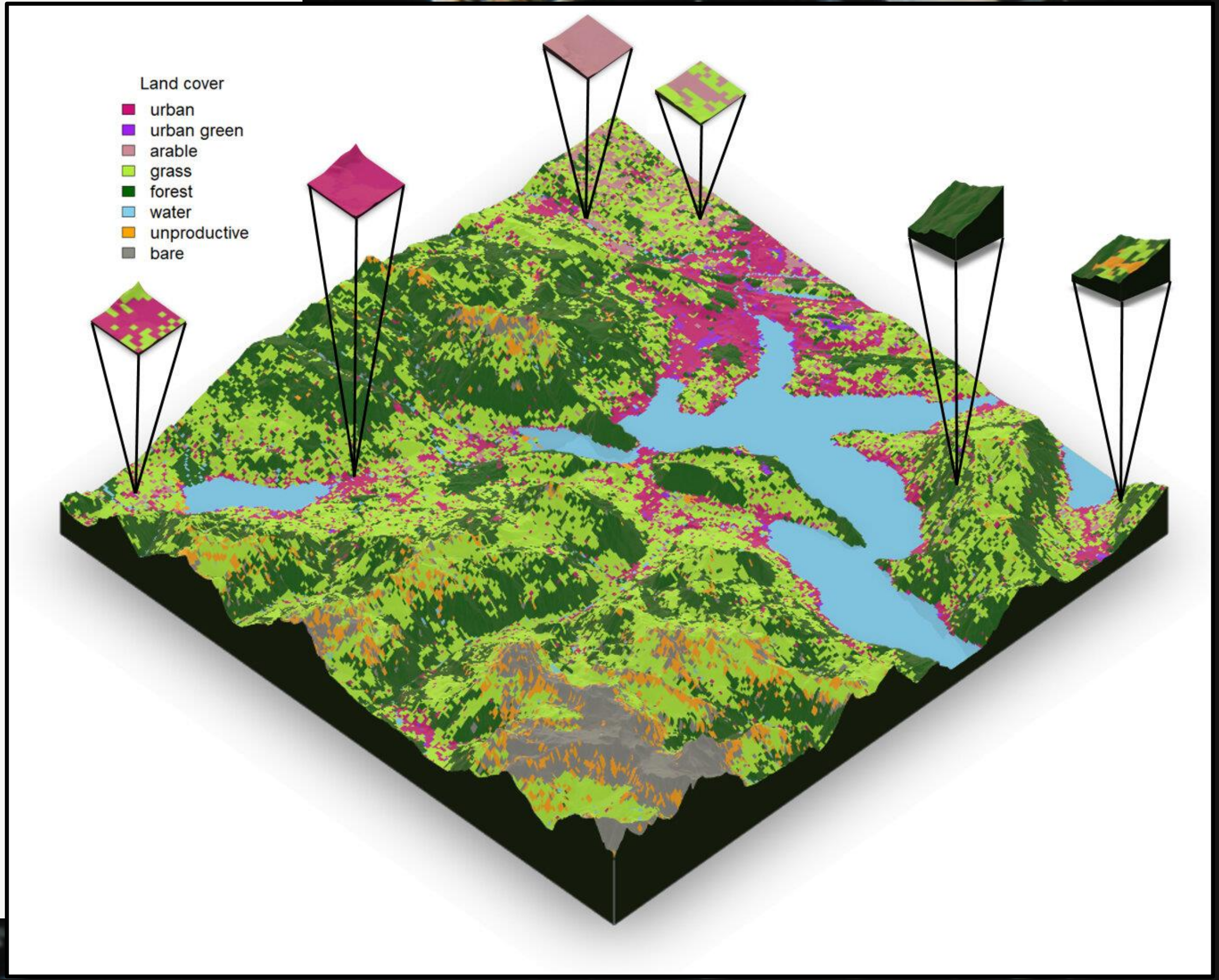
All generate spatial correlation!



Spatial Patterns in Ecological Data

Ecosystem functioning depends on spatial heterogeneity!





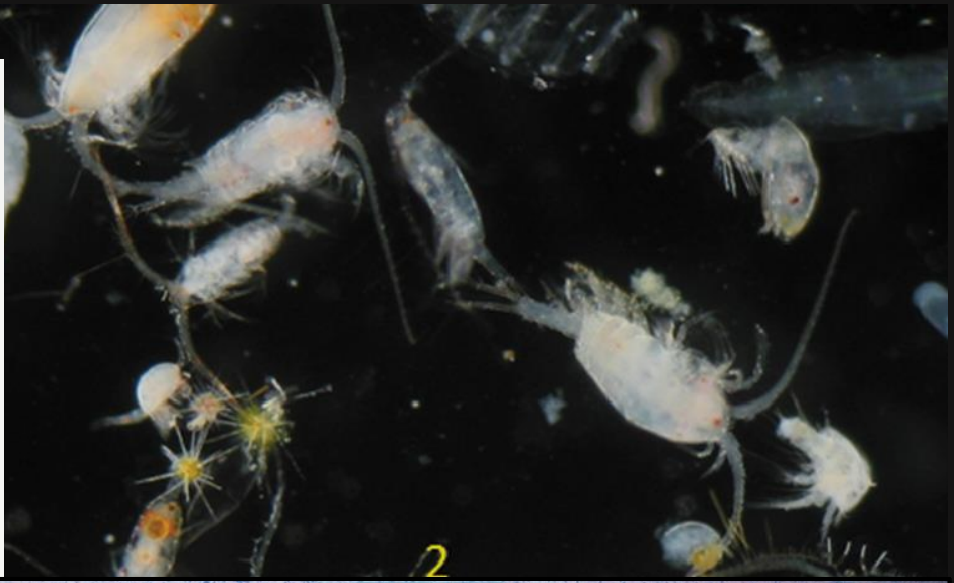
Spatial Autocorrelation

Spatial structures produce relationships between values observed at neighboring points, resulting in a **lack of independence** of values of the observed variable.

Spatial Autocorrelation

Spatial structures produce relationships between values observed at neighboring points, resulting in a **lack of independence** of values of the observed variable.

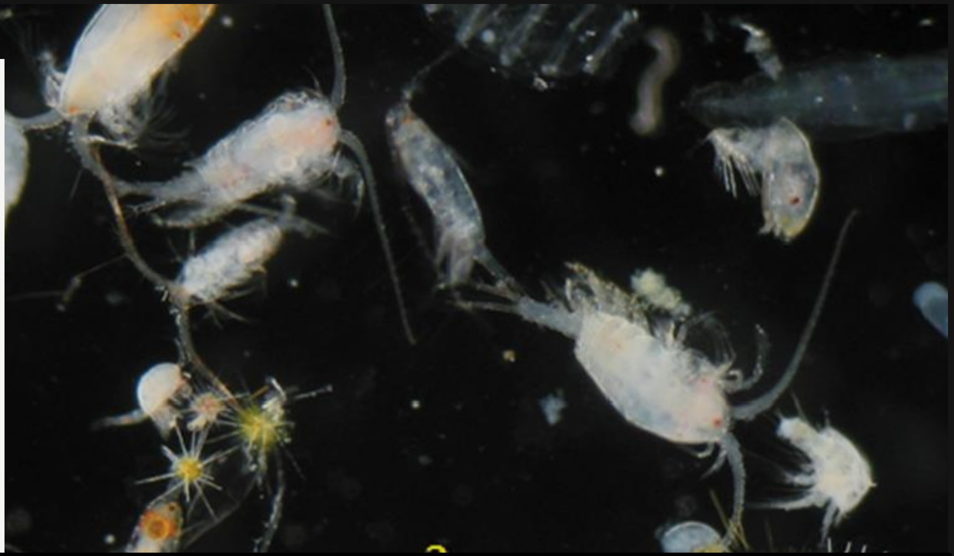
Positive spatial autocorrelation:
Observations that are closer together tend to display values that are *more* similar than observations that are further apart



Spatial Autocorrelation

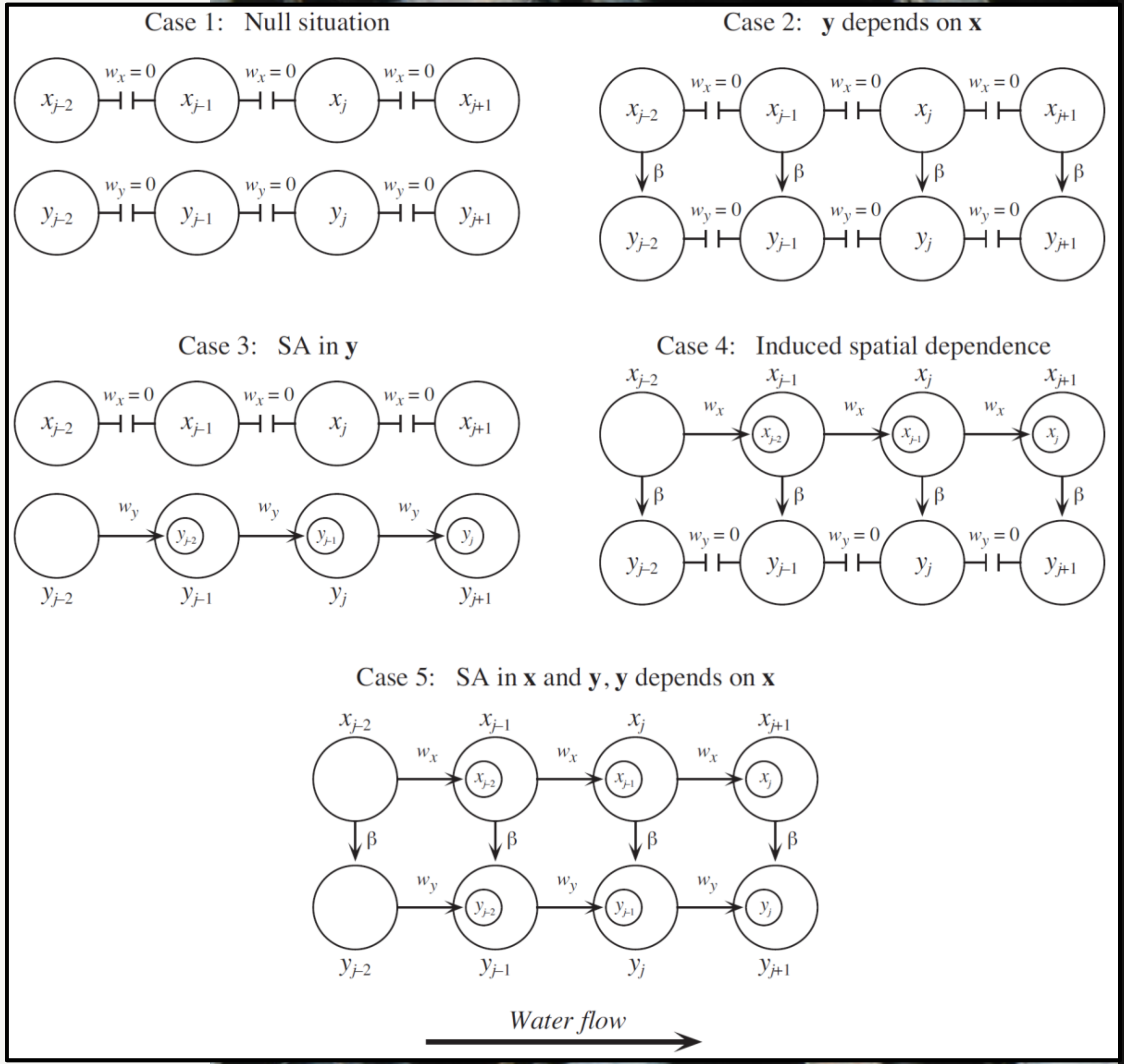
Spatial structures produce relationships between values observed at neighboring points, resulting in a **lack of independence** of values of the observed variable.

Negative spatial autocorrelation:
Observations that are closer together tend to display values that are *less* similar than observations that are further apart



Spatial Autocorrelation

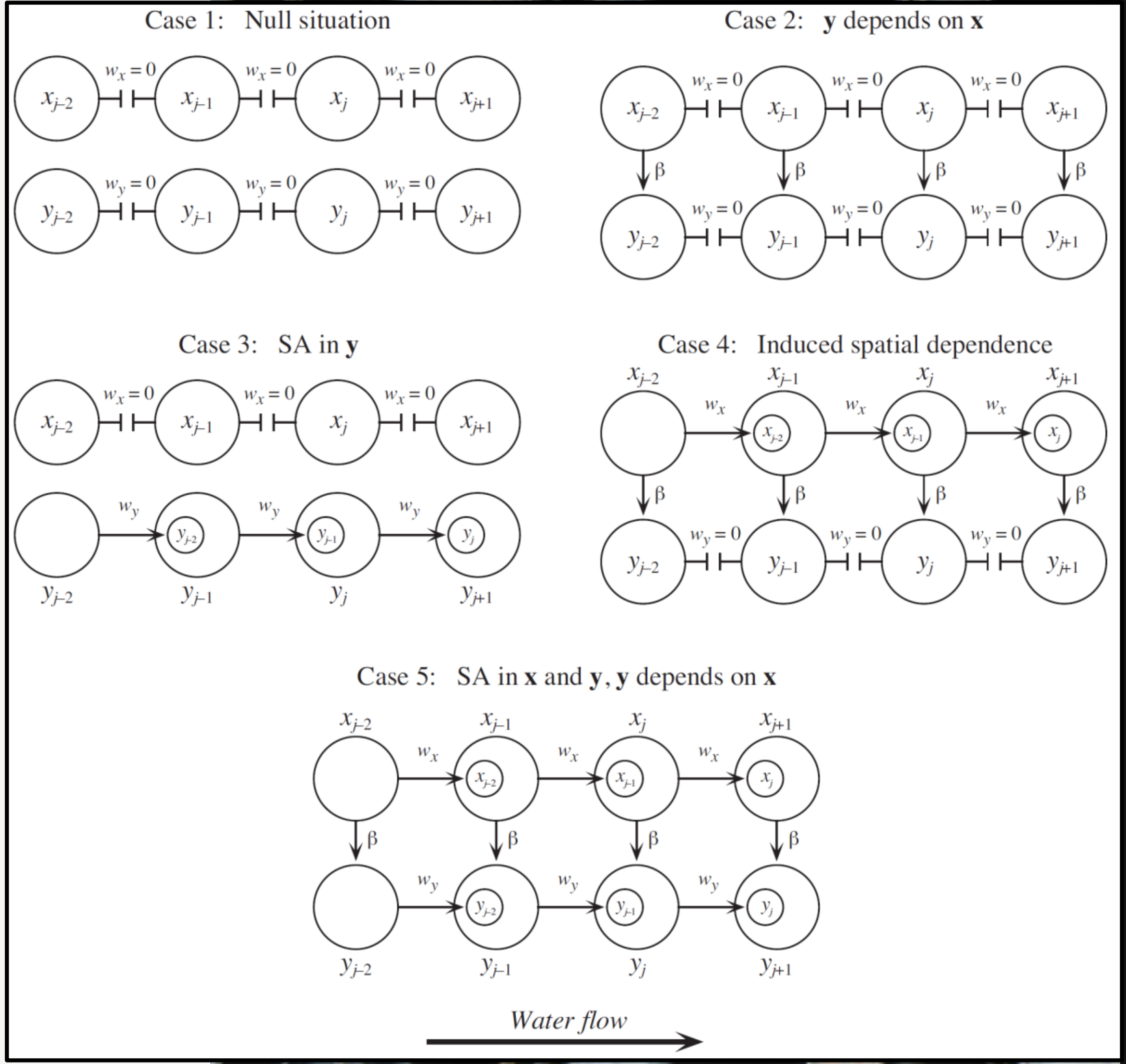
Case 1 – **Null:** No relationship between response variable (**y**) and explanatory variable (**x**)



Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

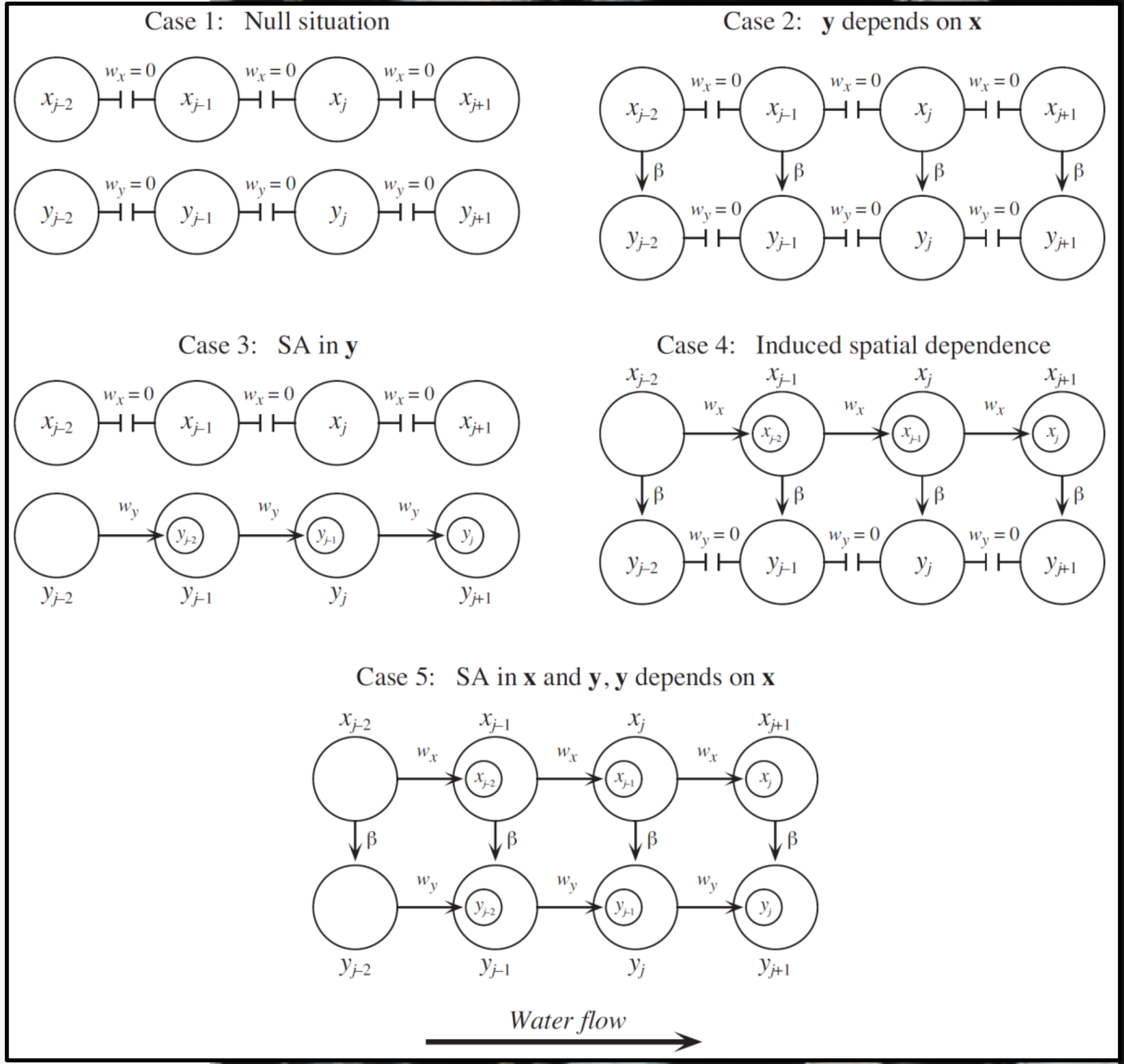
Case 2: Response variable (**y**) depends on explanatory variable (**x**) but there is no spatial dependence



Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

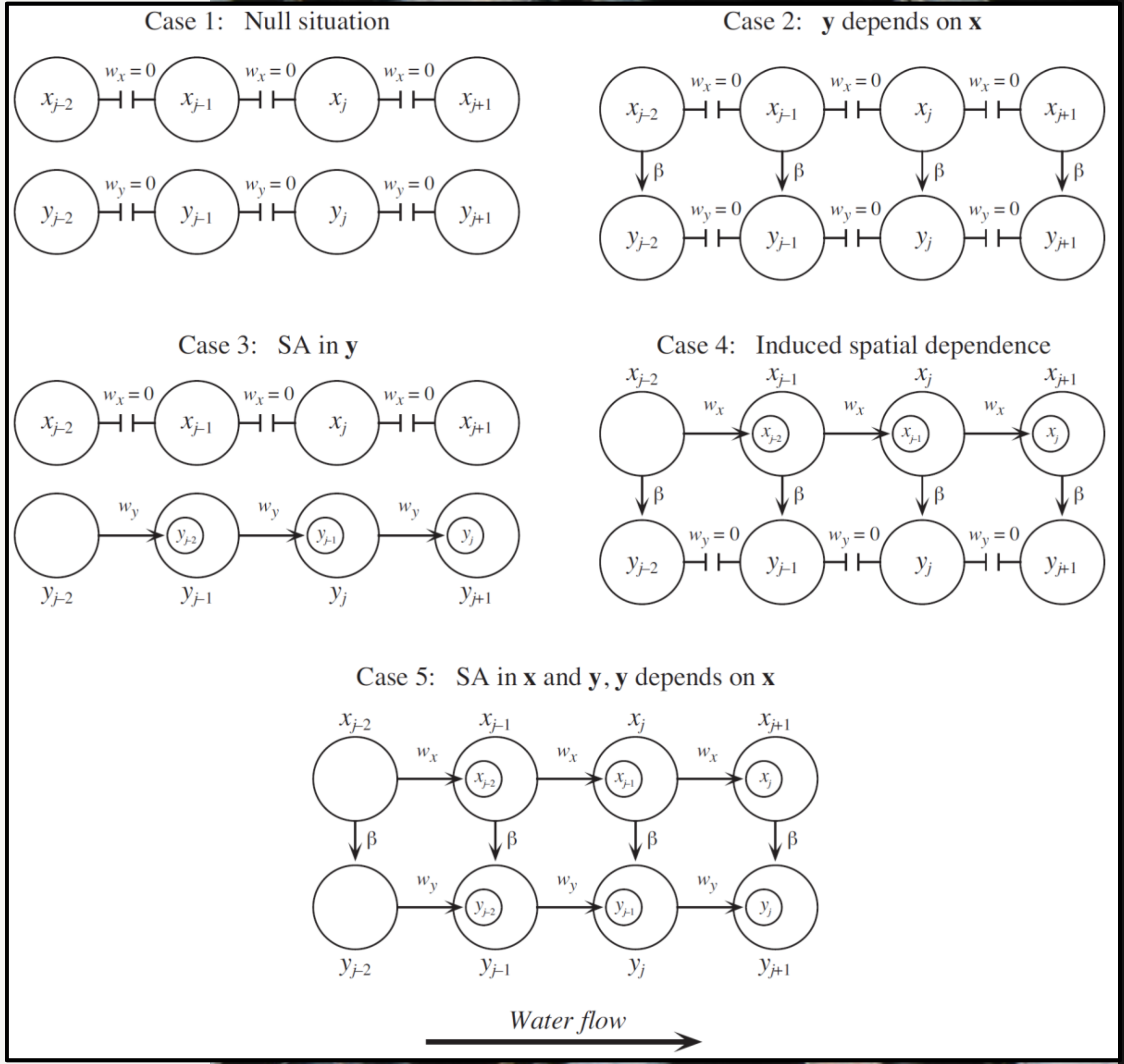
Case 3 – Spatial autocorrelation in response variable y: Response variable (**y**) is spatially autocorrelated but does not functionally depend on explanatory variable (**x**)



Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

Case 4 – Induced spatial dependence:
Response variable (**y**) is functionally dependent on explanatory variable (**x**), which is spatially autocorrelated

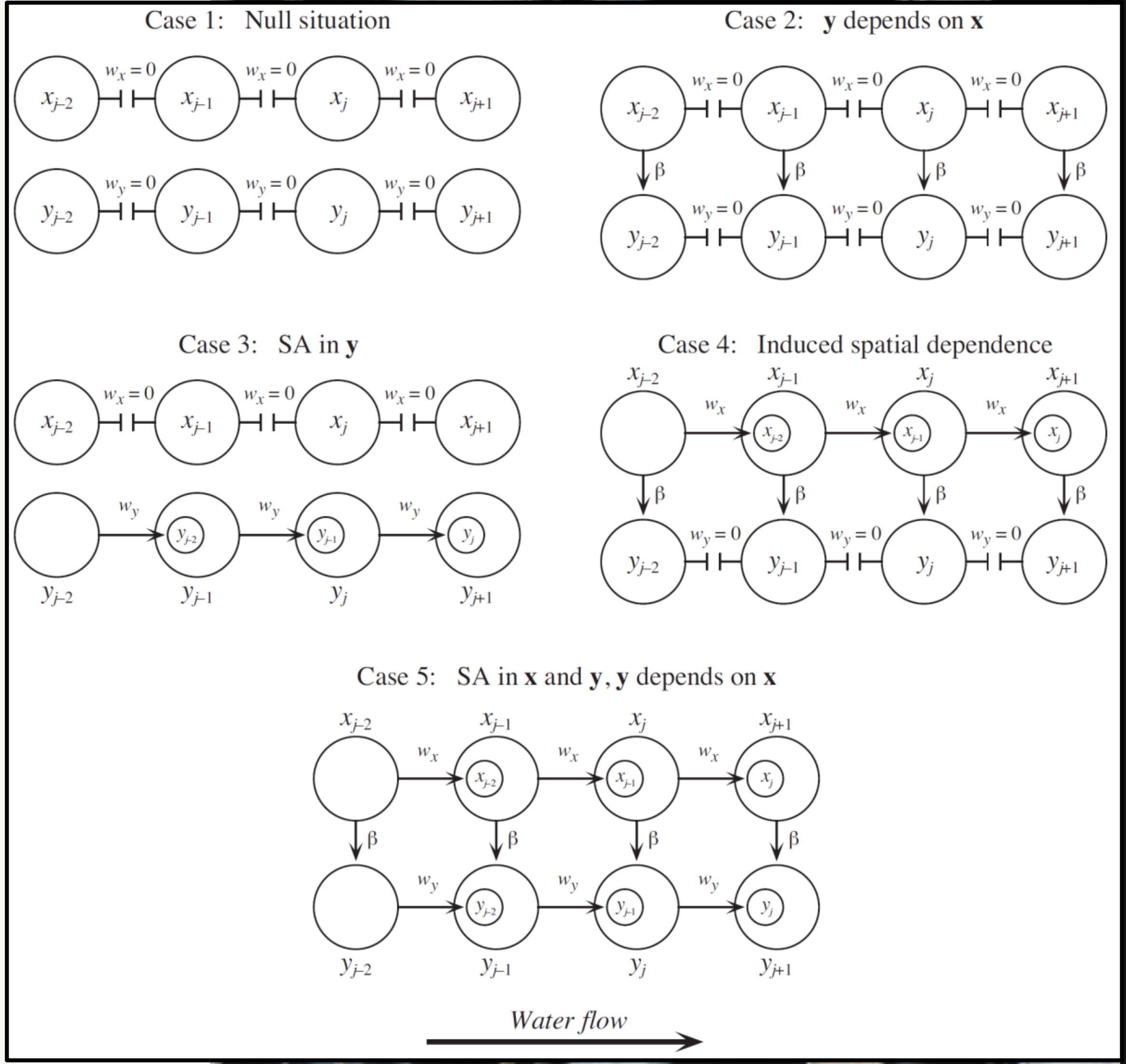


Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

Case 4 – Induced spatial dependence:
Response variable (**y**) is functionally dependent on explanatory variable (**x**), which is spatially autocorrelated

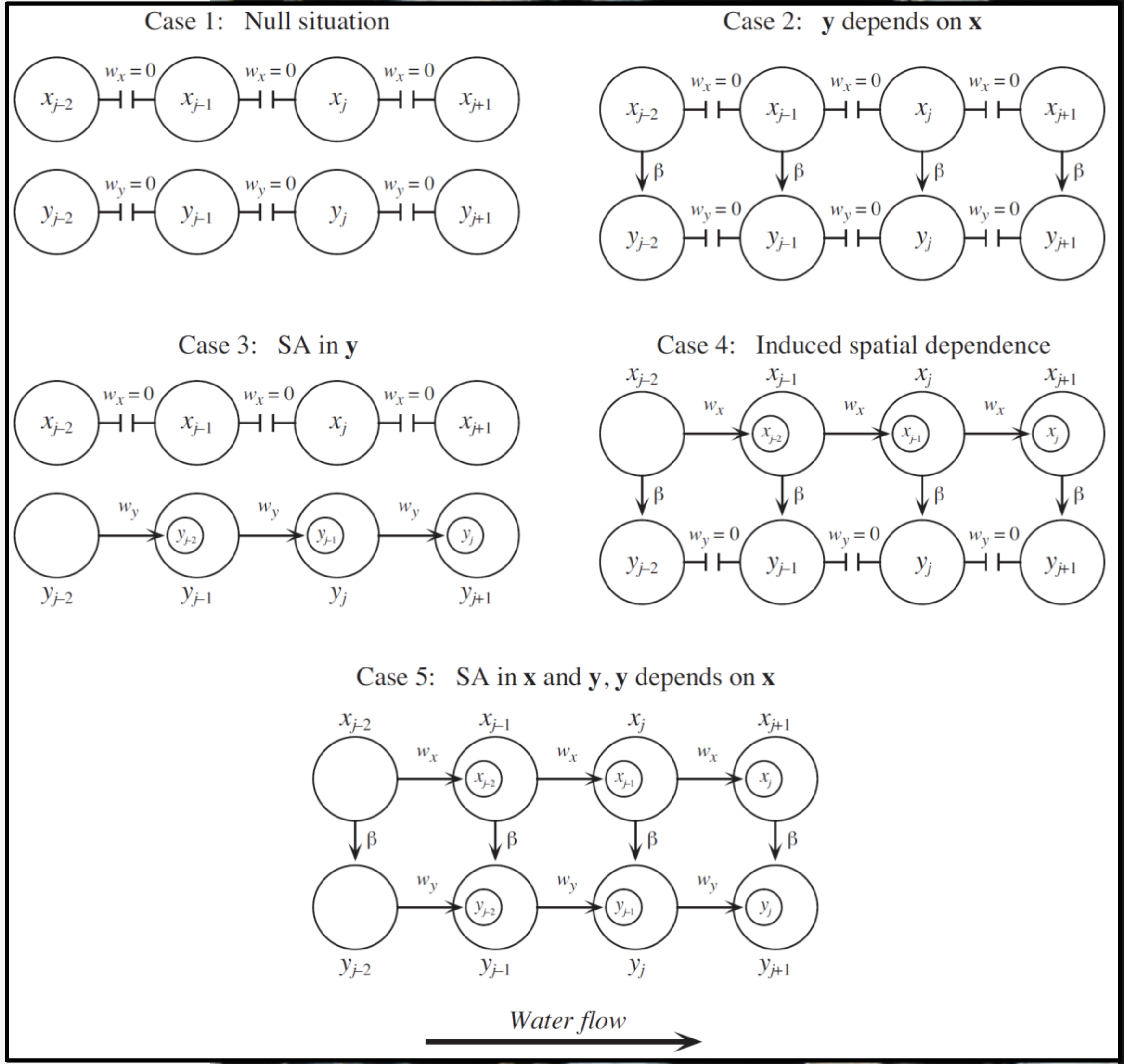
If all important, spatially-structured explanatory variables are included in the analysis, the model will correctly account for the spatial structure induced in y.



Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

Case 5 – Spatial autocorrelation:
Response variable (**y**) is functionally dependent on explanatory variable (**x**), and both are spatially autocorrelated

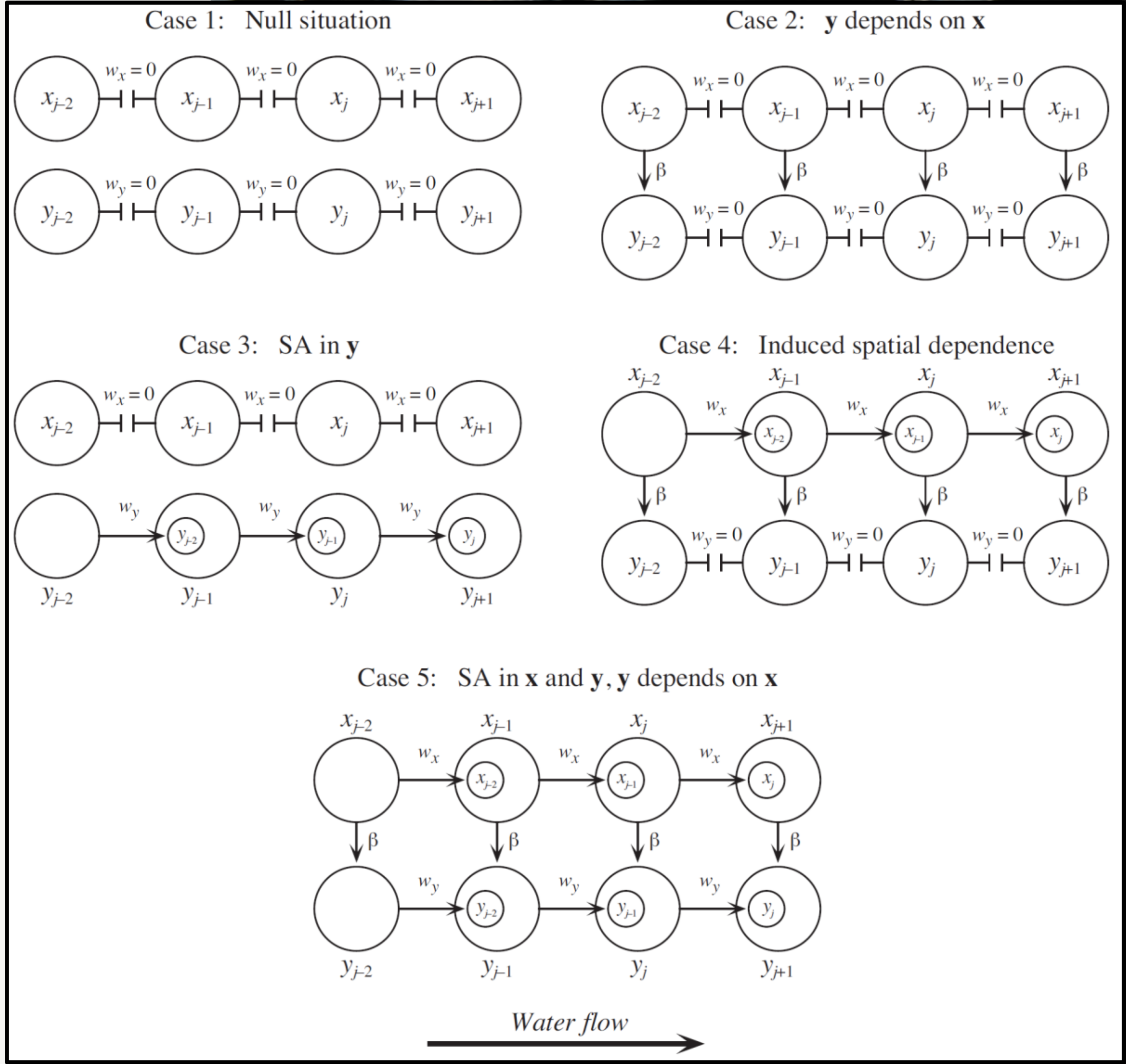


Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

Case 5 – Spatial autocorrelation:
Response variable (**y**) is functionally dependent on explanatory variable (**x**), and both are spatially autocorrelated

Variation partitioning can be used to determine whether or not the entire spatial structure in the data can be explained by the environmental variables, or if there remains an unexplained portion of spatial variation.

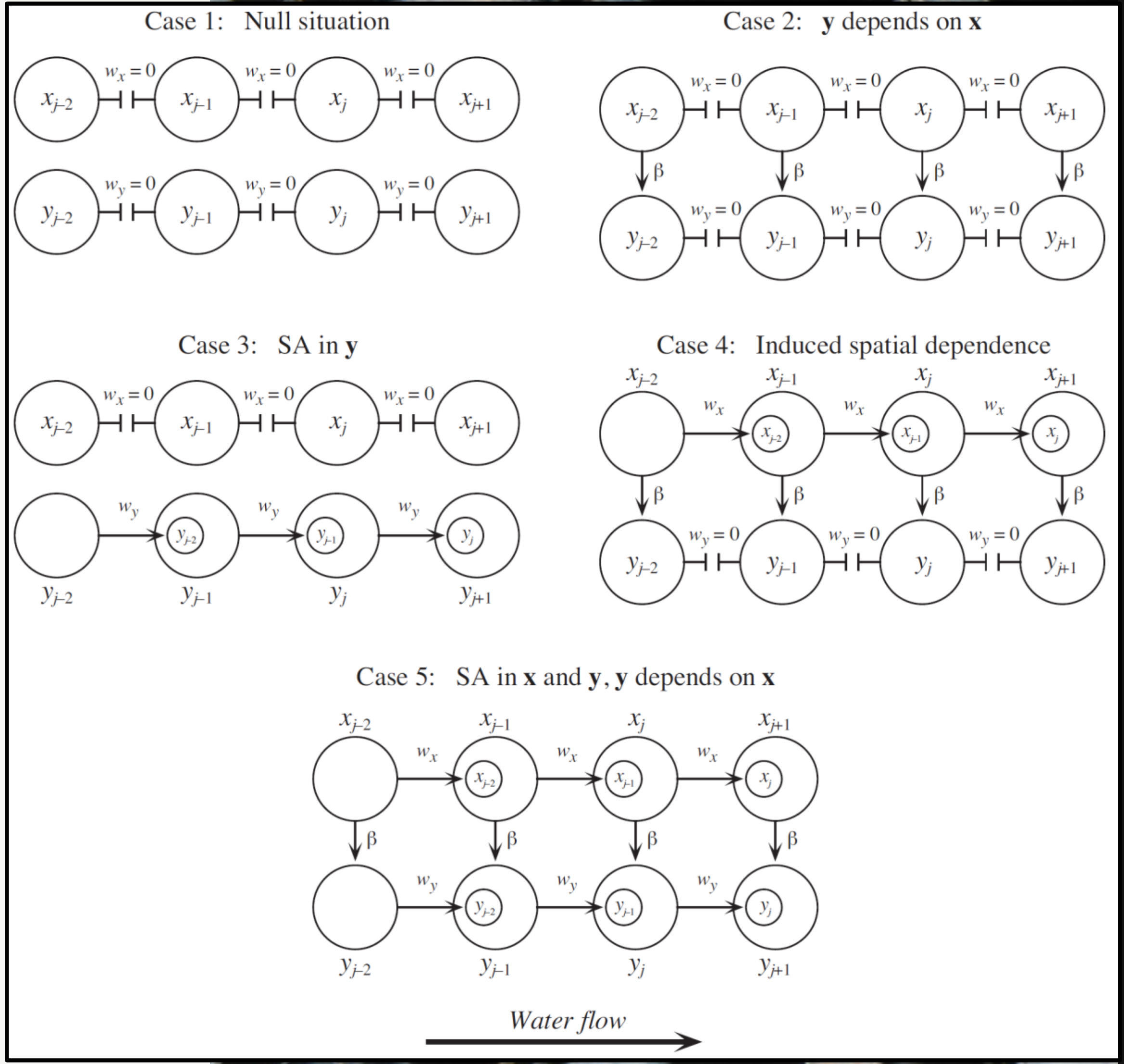


Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

Influence on inference and tests of significance:

Spatially autocorrelated values are not stochastically independent of one another!



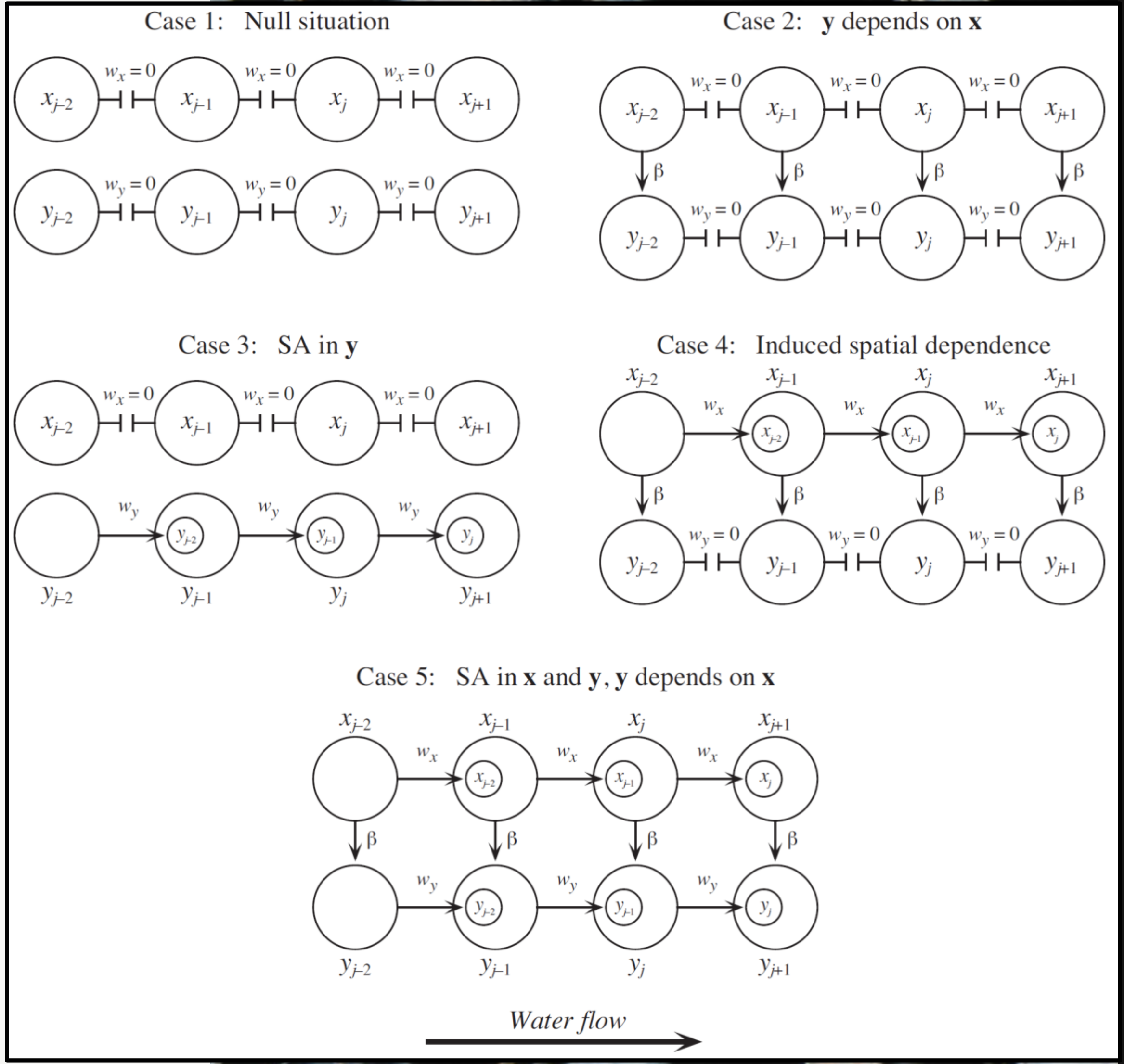
Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

Influence on inference and tests of significance:

Spatially autocorrelated values are not stochastically independent of one another!

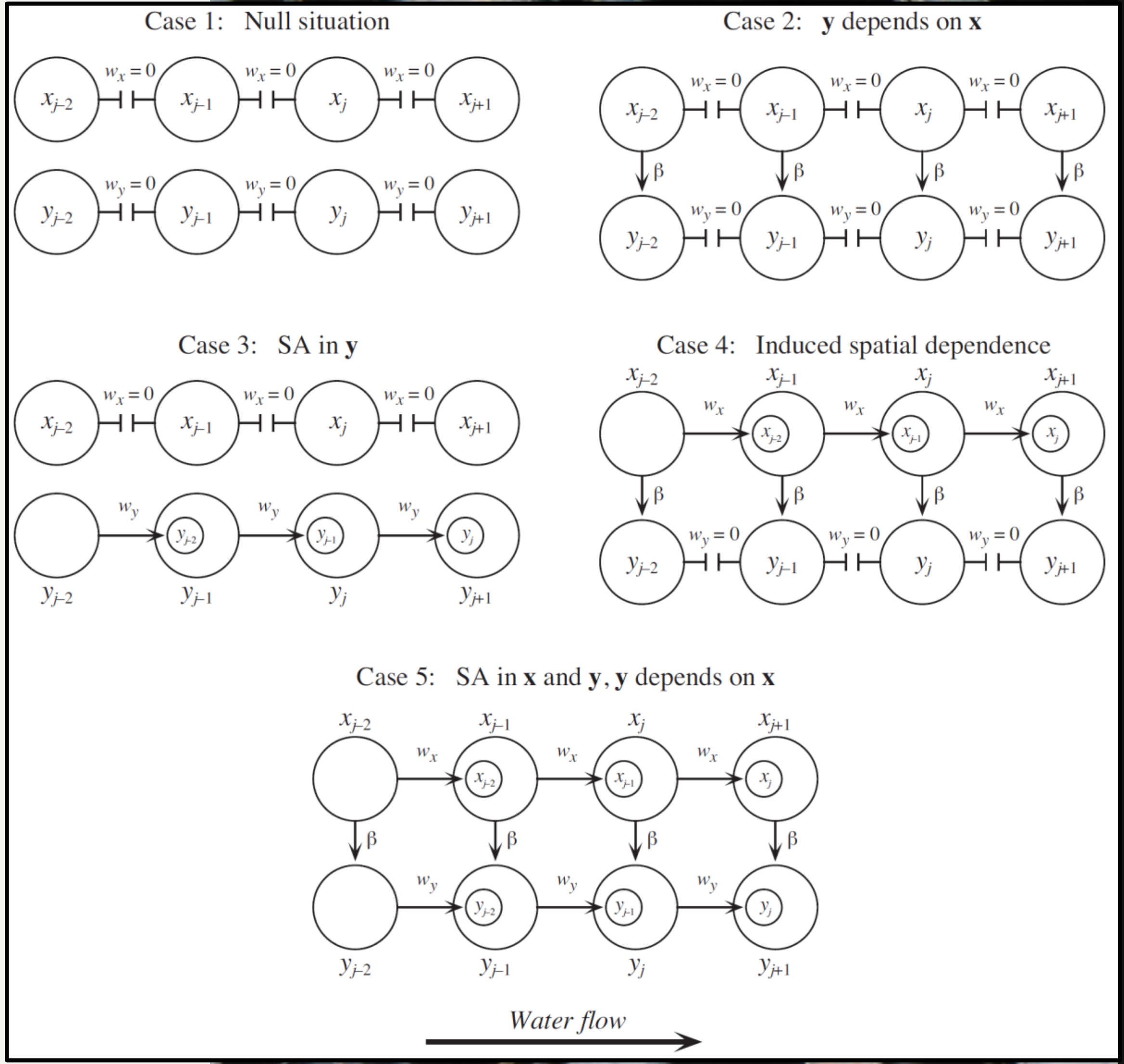
For both parametric and non-parametric tests, ***the computed test statistics will exhibit Type 1 error (false significance) when positive spatial autocorrelation is present.***



Legendre & Legendre Fig. 1.5

Spatial Autocorrelation

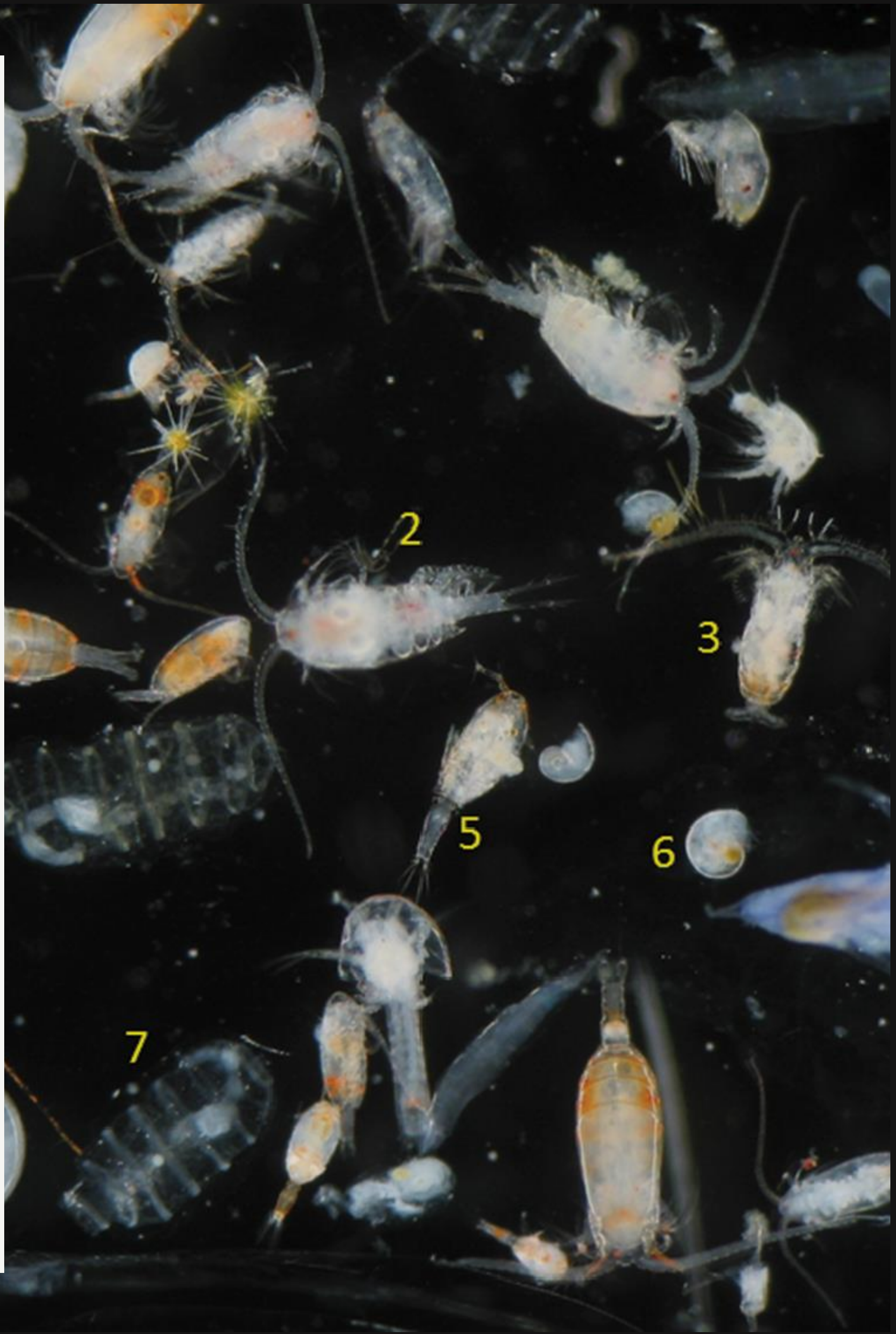
Because clustering and ordination do not rely on tests of statistical significance, they are not affected by the presence of spatial autocorrelation.



Legendre & Legendre Fig. 1.5

Semi-Variograms

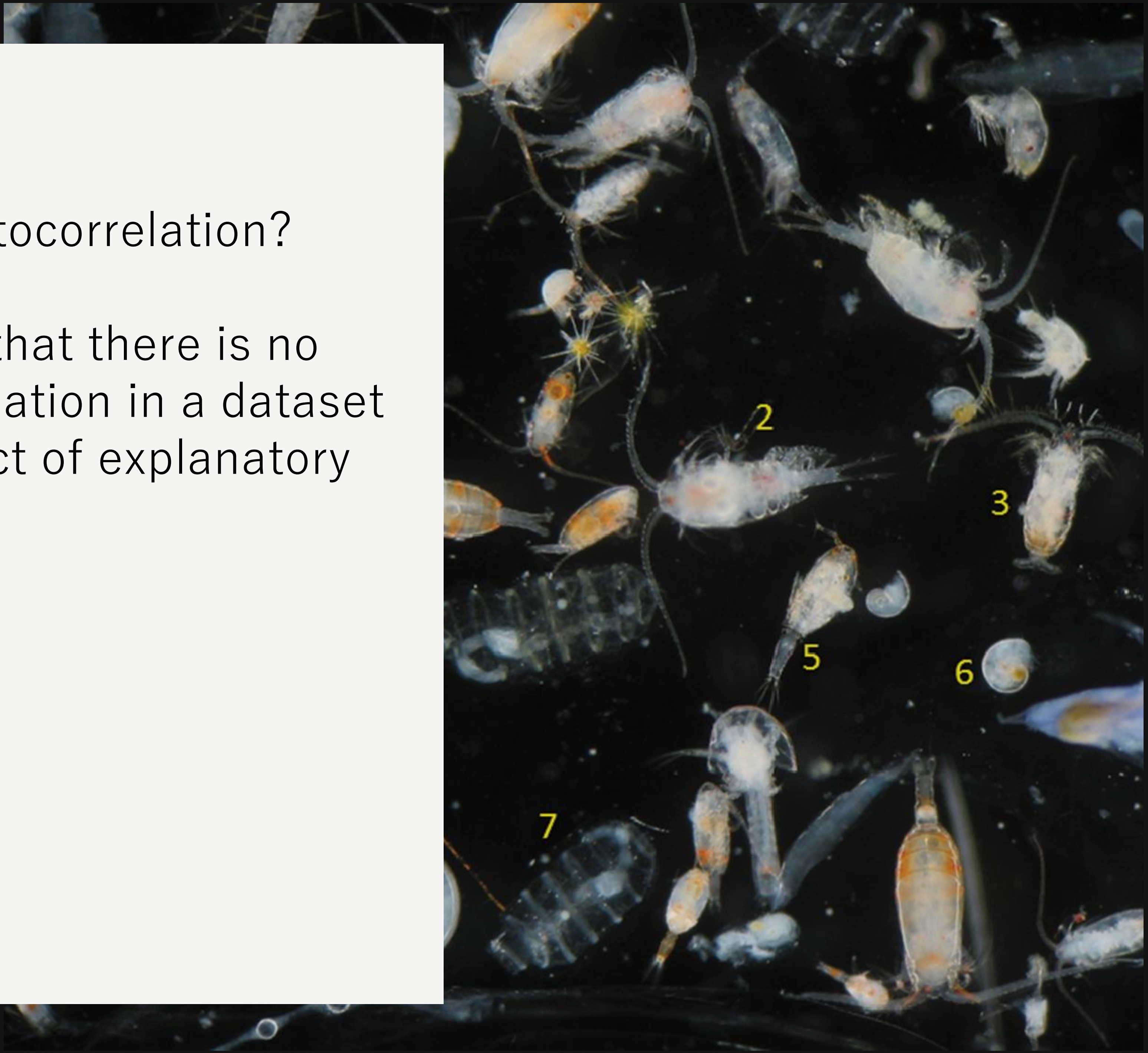
How do we quantify spatial autocorrelation?



Semi-Variograms

How do we quantify spatial autocorrelation?

- Testing the null hypothesis that there is no significant spatial autocorrelation in a dataset after controlling for the effect of explanatory variables.

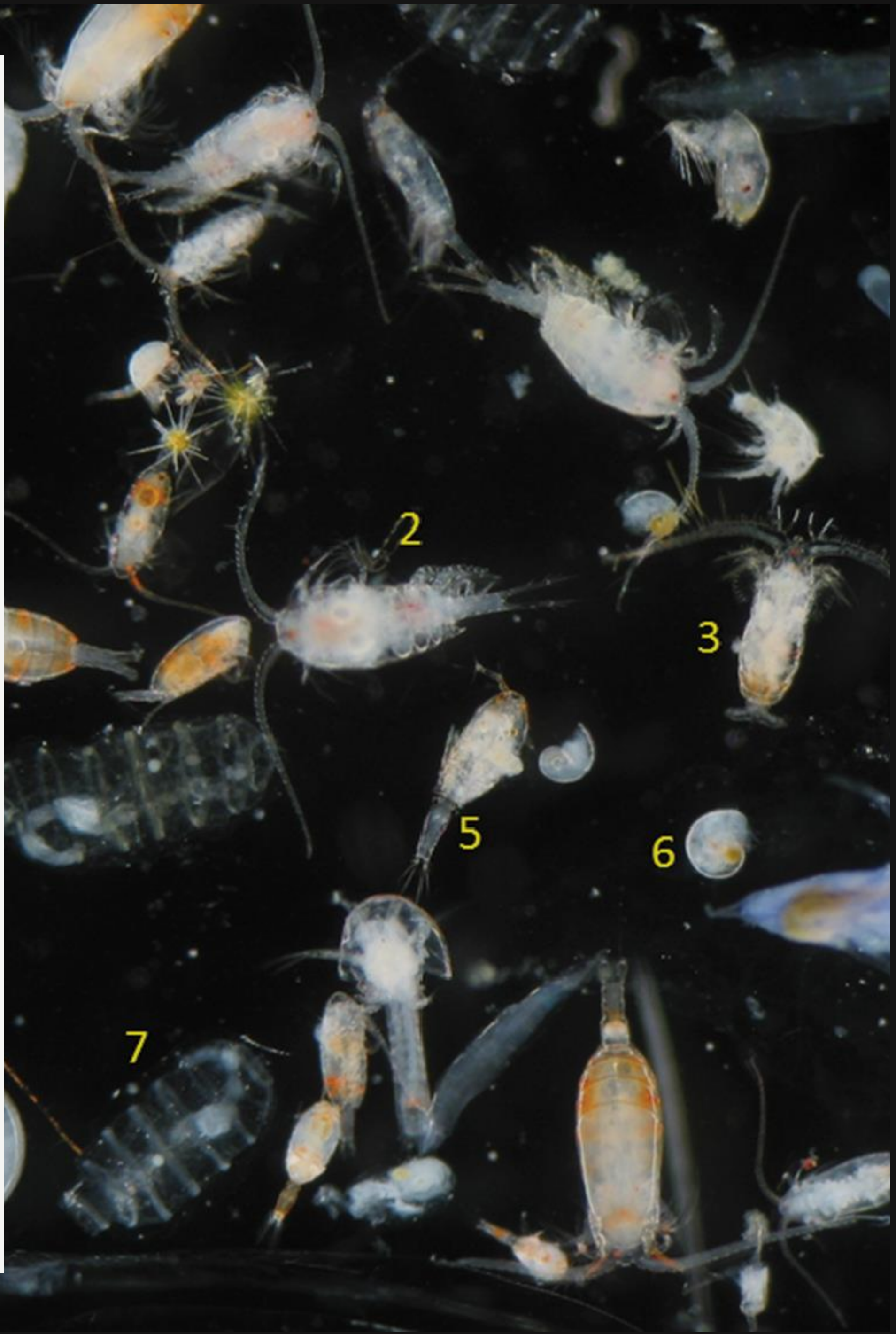


Semi-Variograms

How do we quantify spatial autocorrelation?

- Testing the null hypothesis that there is no significant spatial autocorrelation in a dataset after controlling for the effect of explanatory variables.

- Rejecting the null hypothesis to show significant spatial autocorrelation and to include it in conceptual or statistical models.

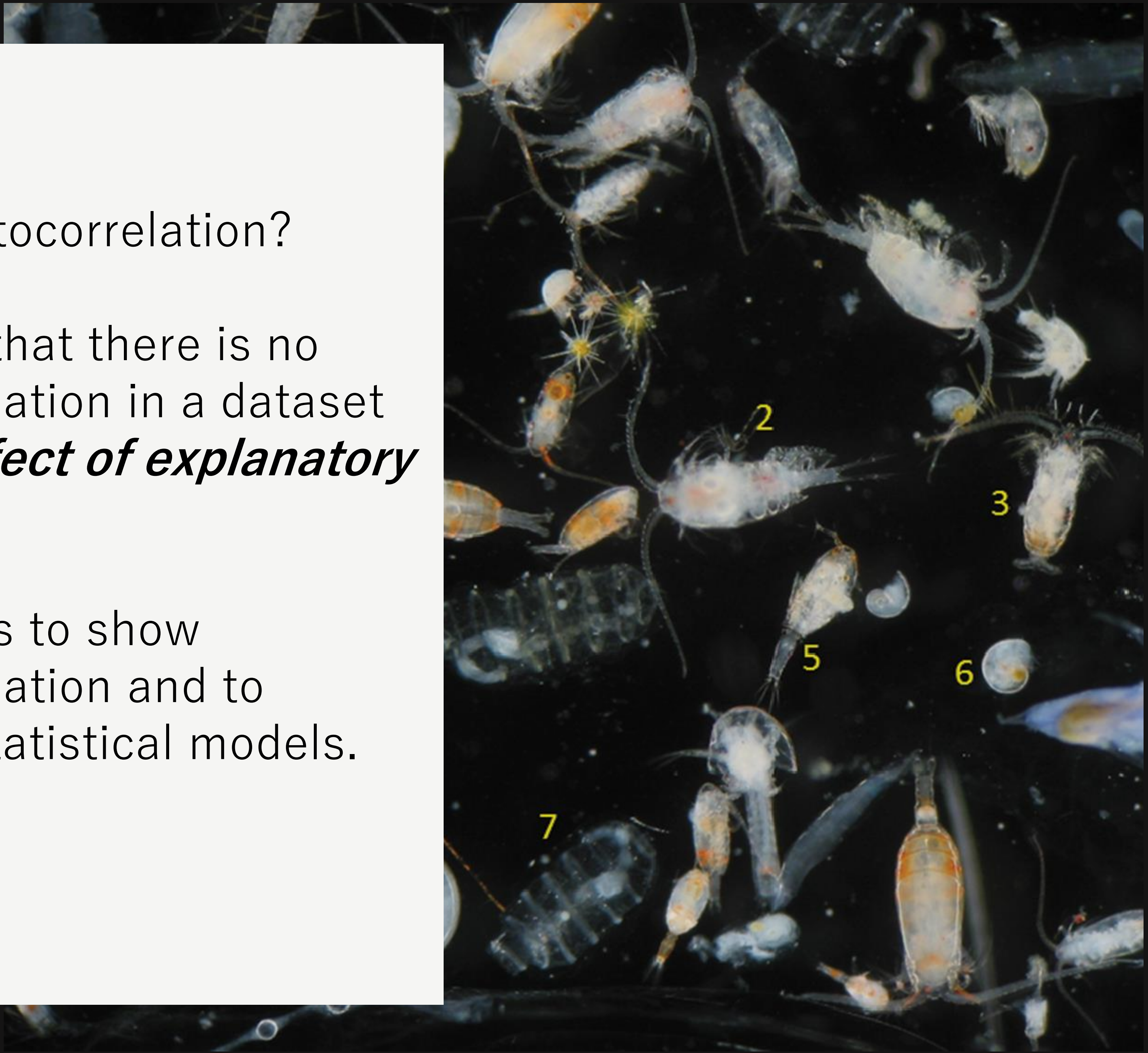


Semi-Variograms

How do we quantify spatial autocorrelation?

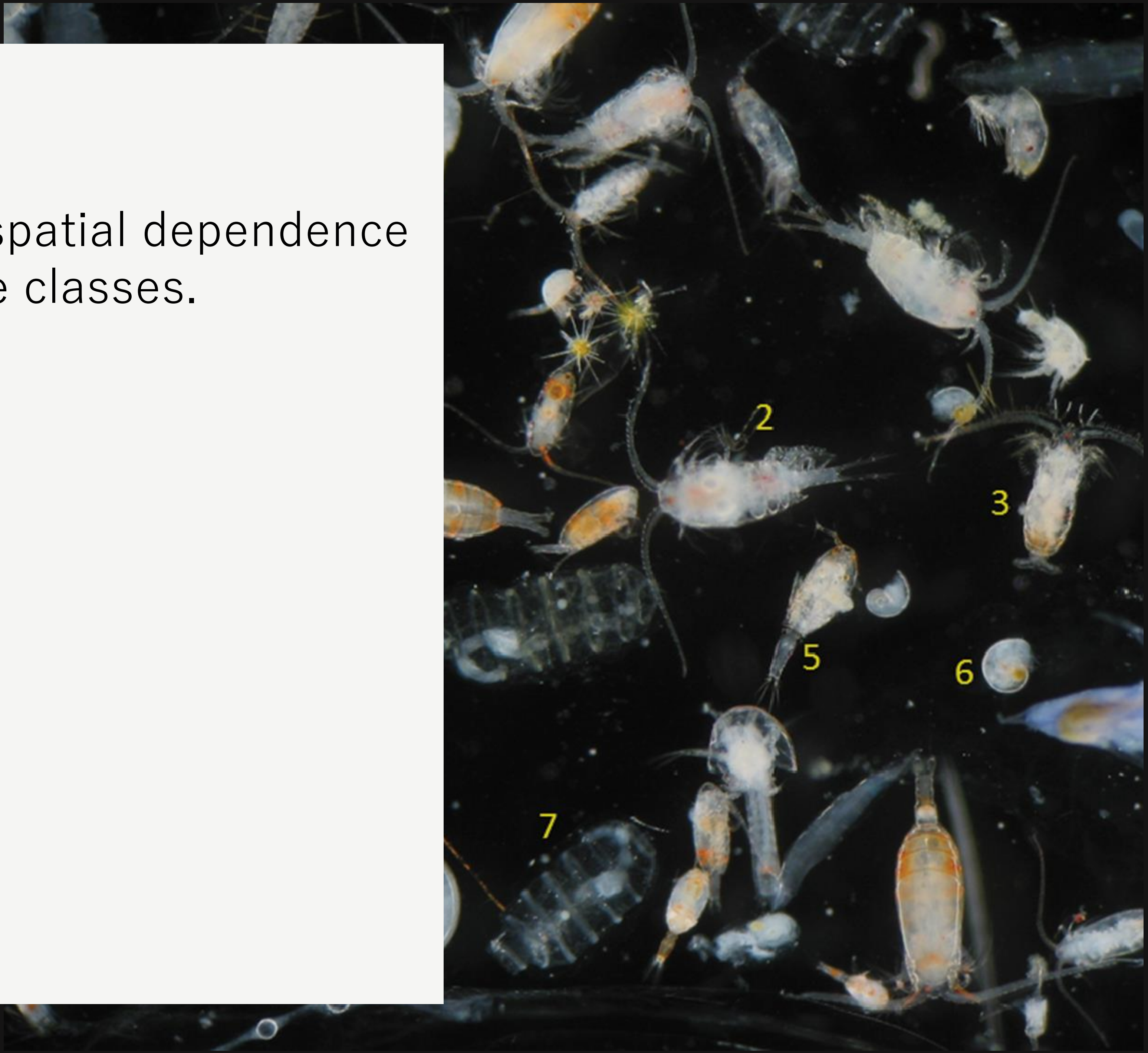
- Testing the null hypothesis that there is no significant spatial autocorrelation in a dataset ***after controlling for the effect of explanatory variables***.

- Rejecting the null hypothesis to show significant spatial autocorrelation and to include it in conceptual or statistical models.



Semi-Variograms

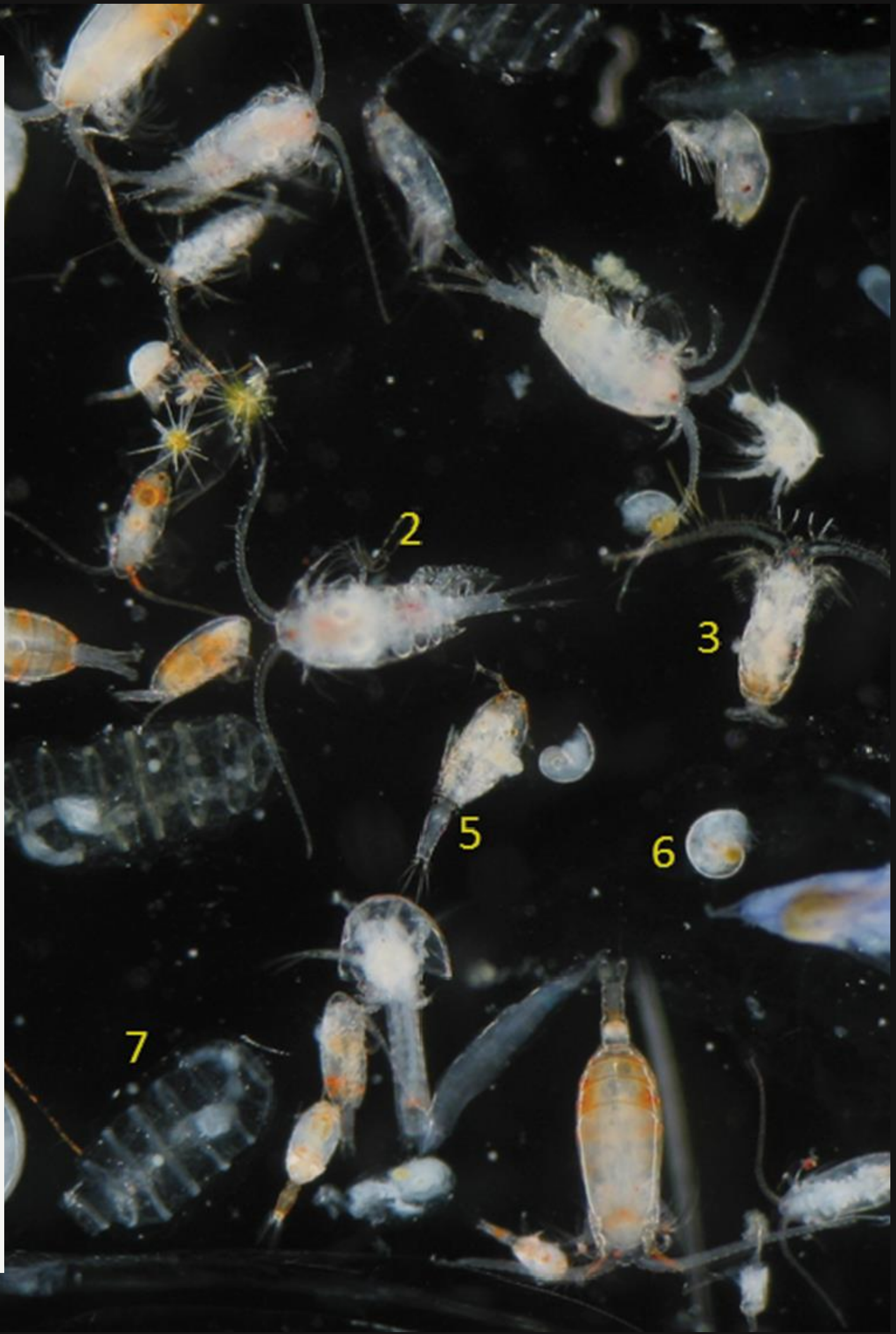
Structure functions quantify spatial dependence and partition it among distance classes.



Semi-Variograms

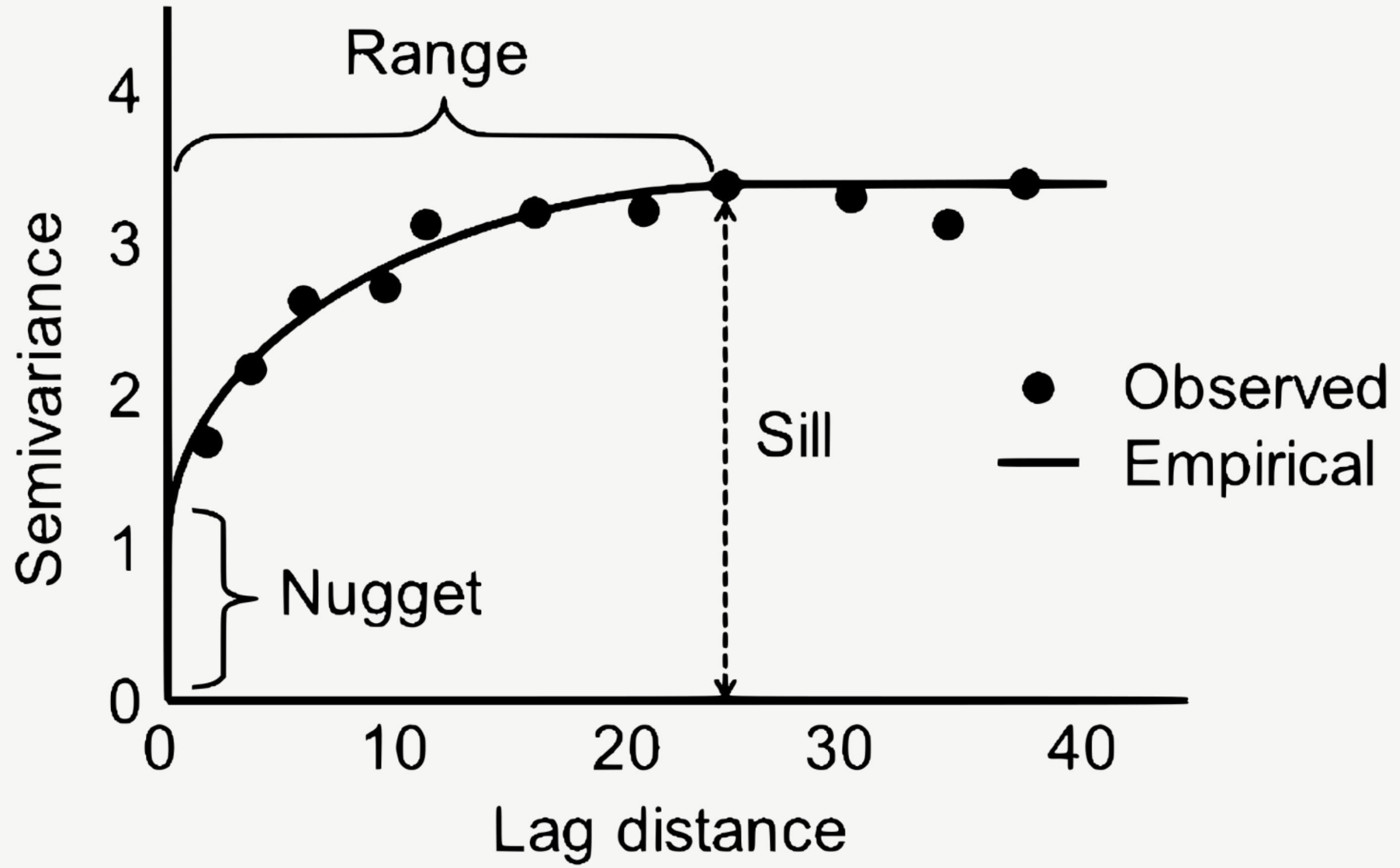
Structure functions quantify spatial dependence and partition it among distance classes.

- **Correlogram:** spatial correlation vs. distance
- **Variogram (or semi-variogram):** semi-variance vs. distance



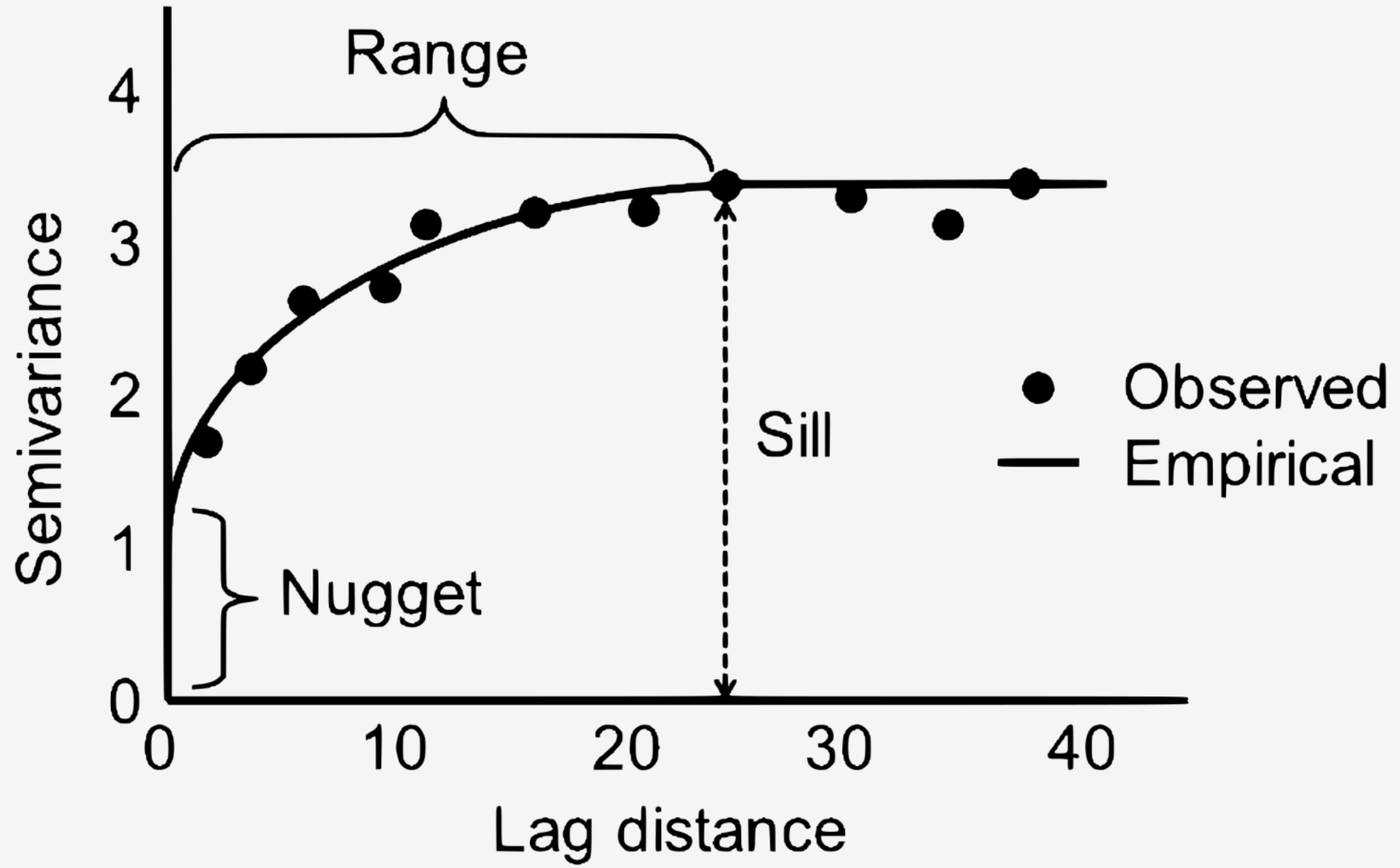
Semi-Variograms

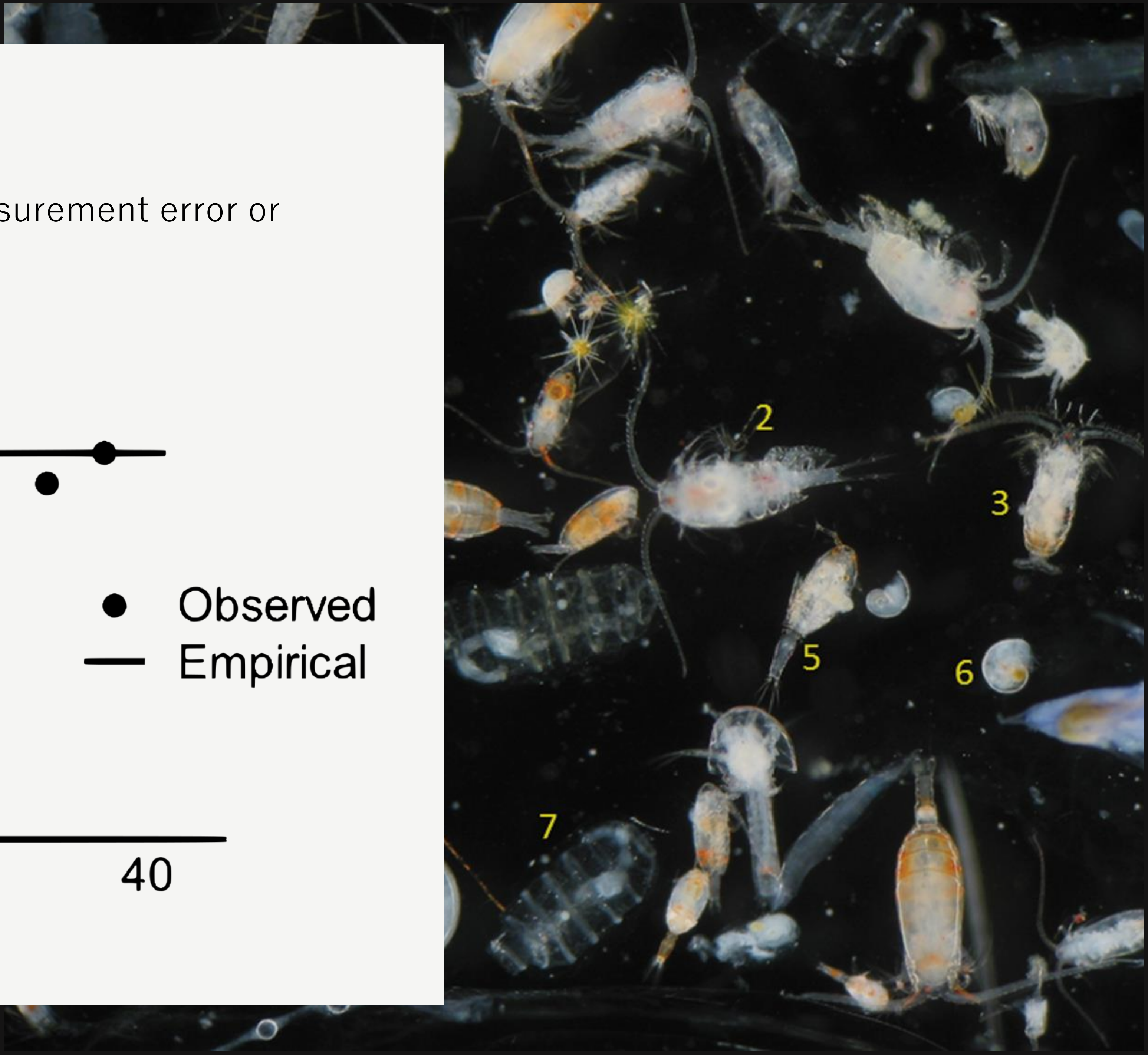
Semi-variance measures the average squared difference between values of a variable at pairs of locations.



Semi-Variograms

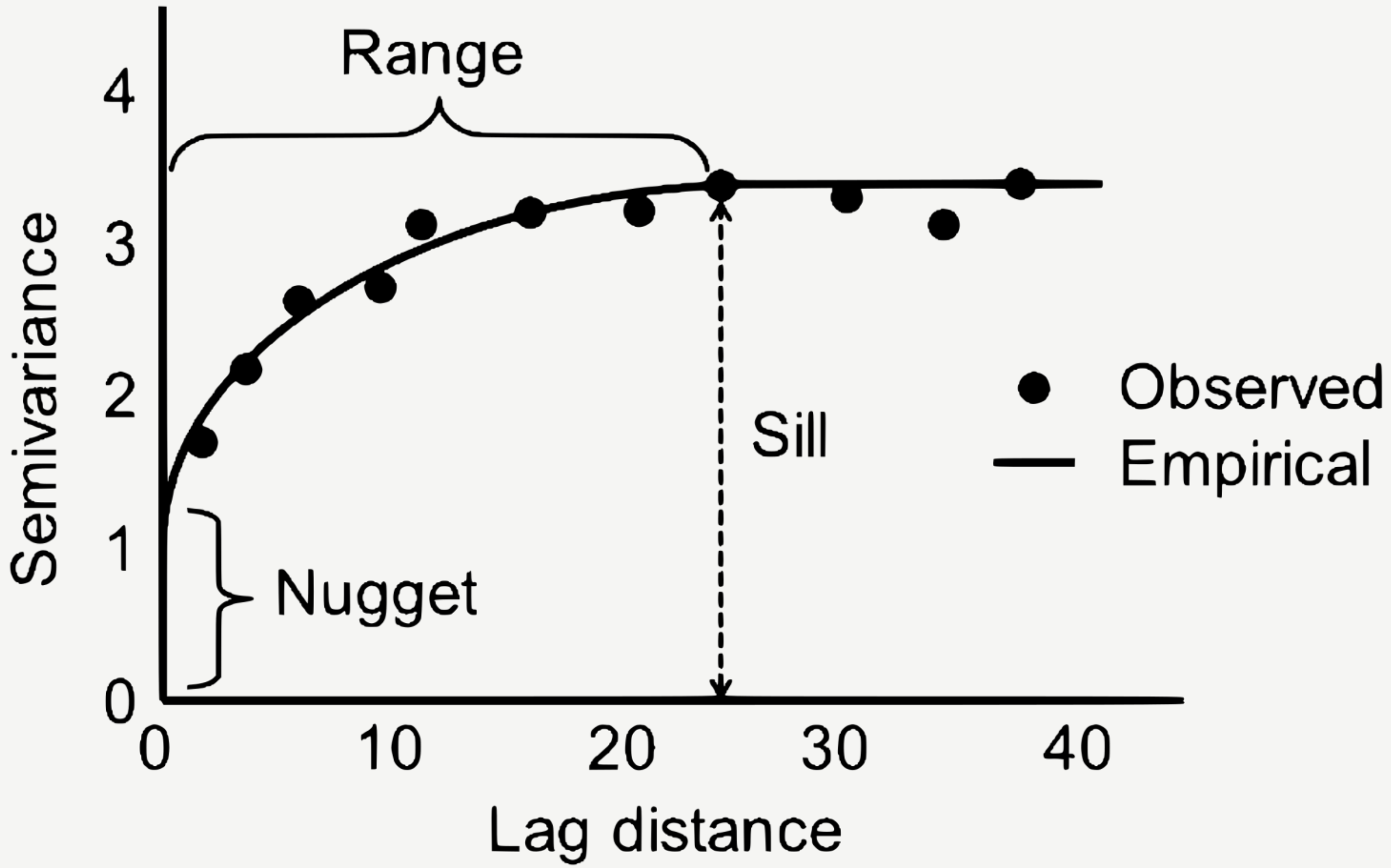
Nugget: The y-intercept represents measurement error or variability at very small scales





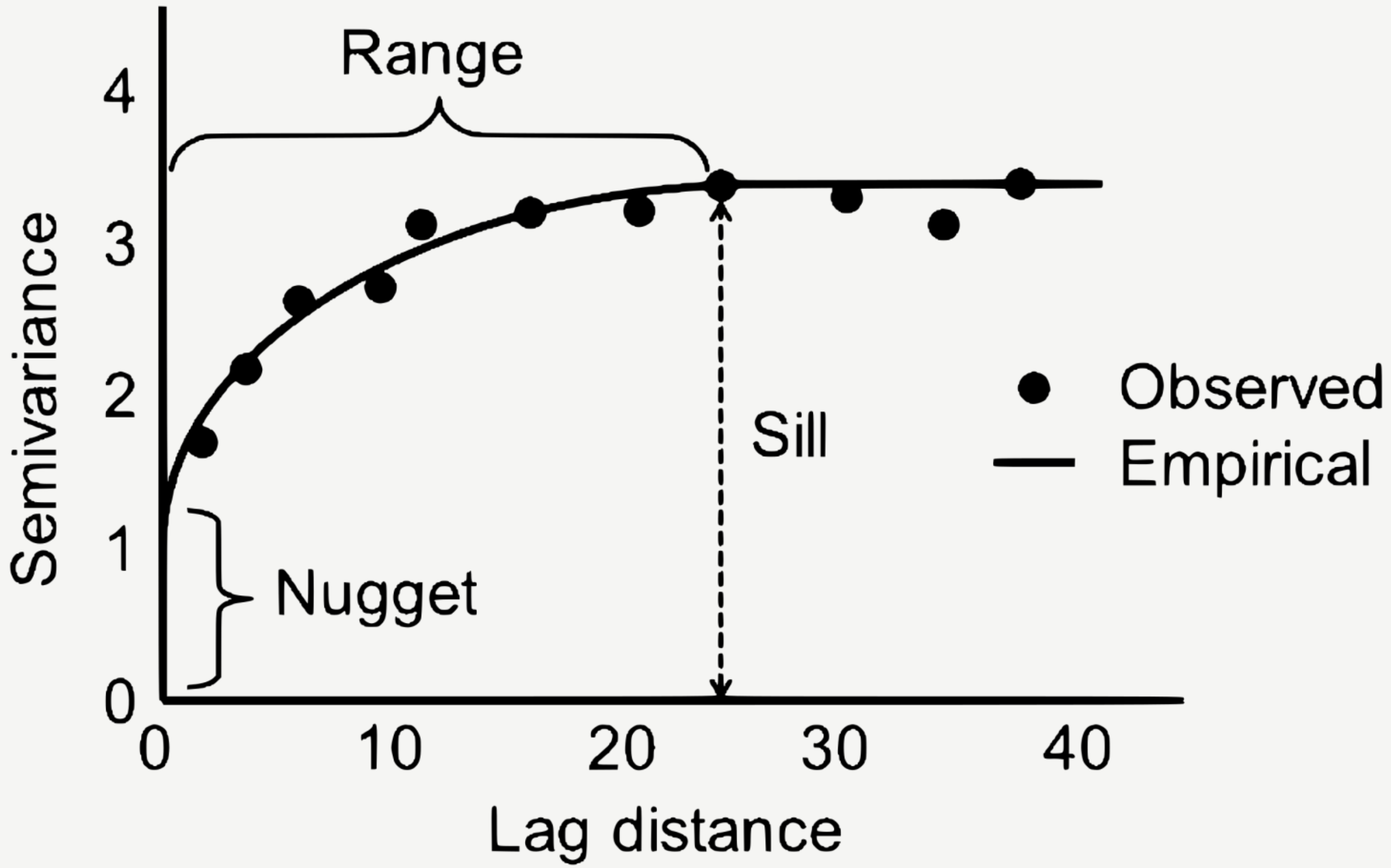
Semi-Variograms

Sill: The point where the curve flattens, indicating the maximum level of dissimilarity



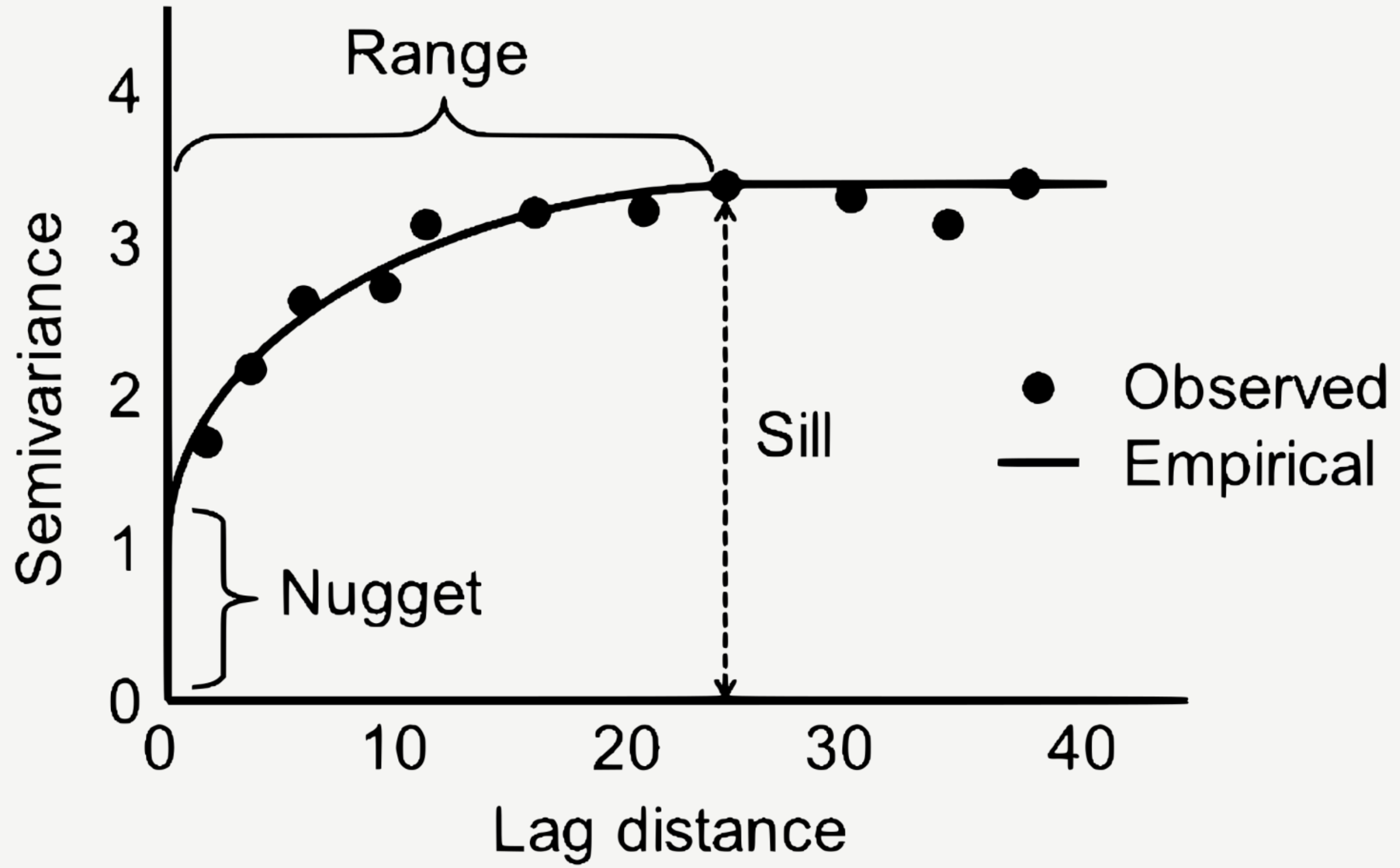
Semi-Variograms

Range: The distance beyond which points no longer influence each other.

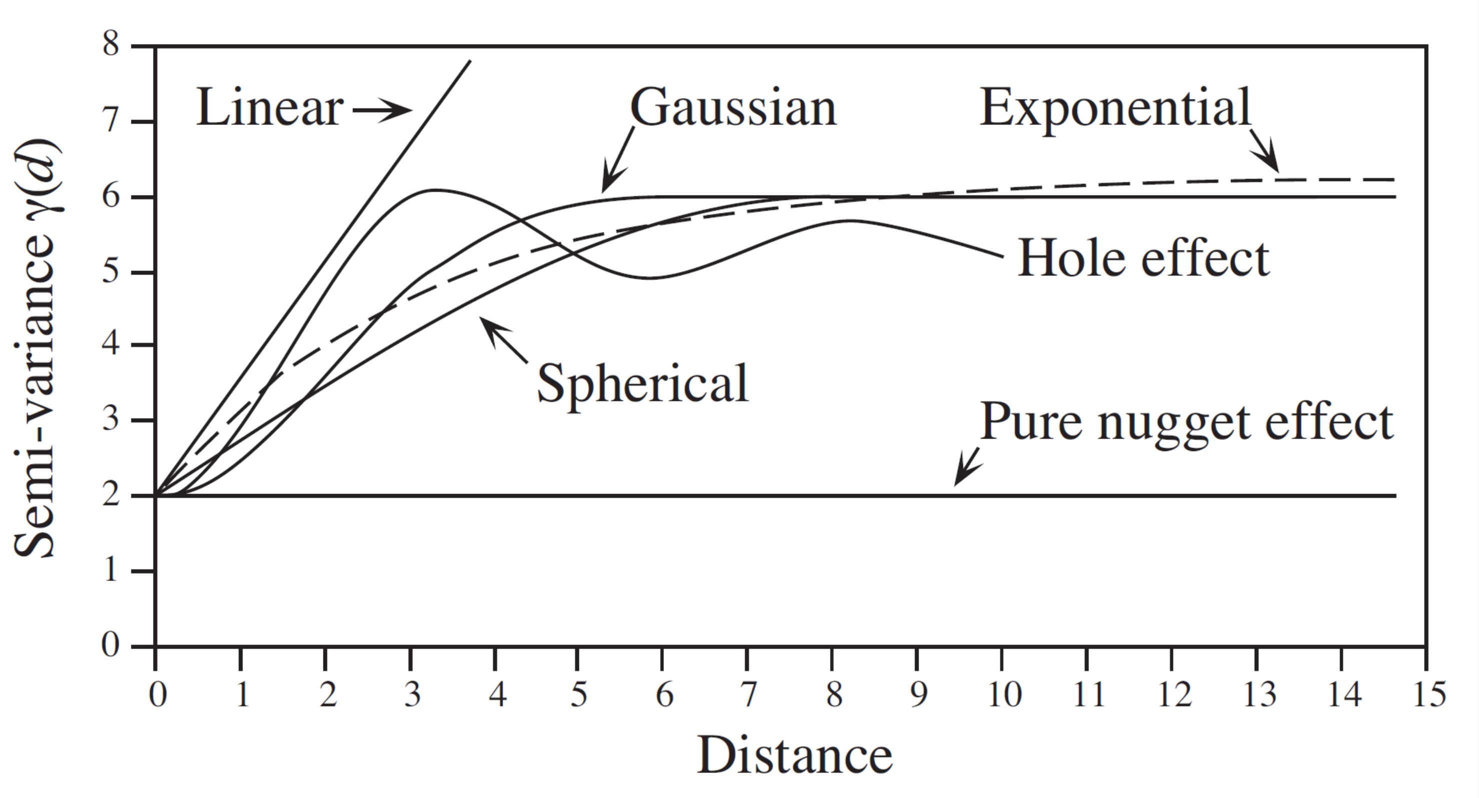


Semi-Variograms

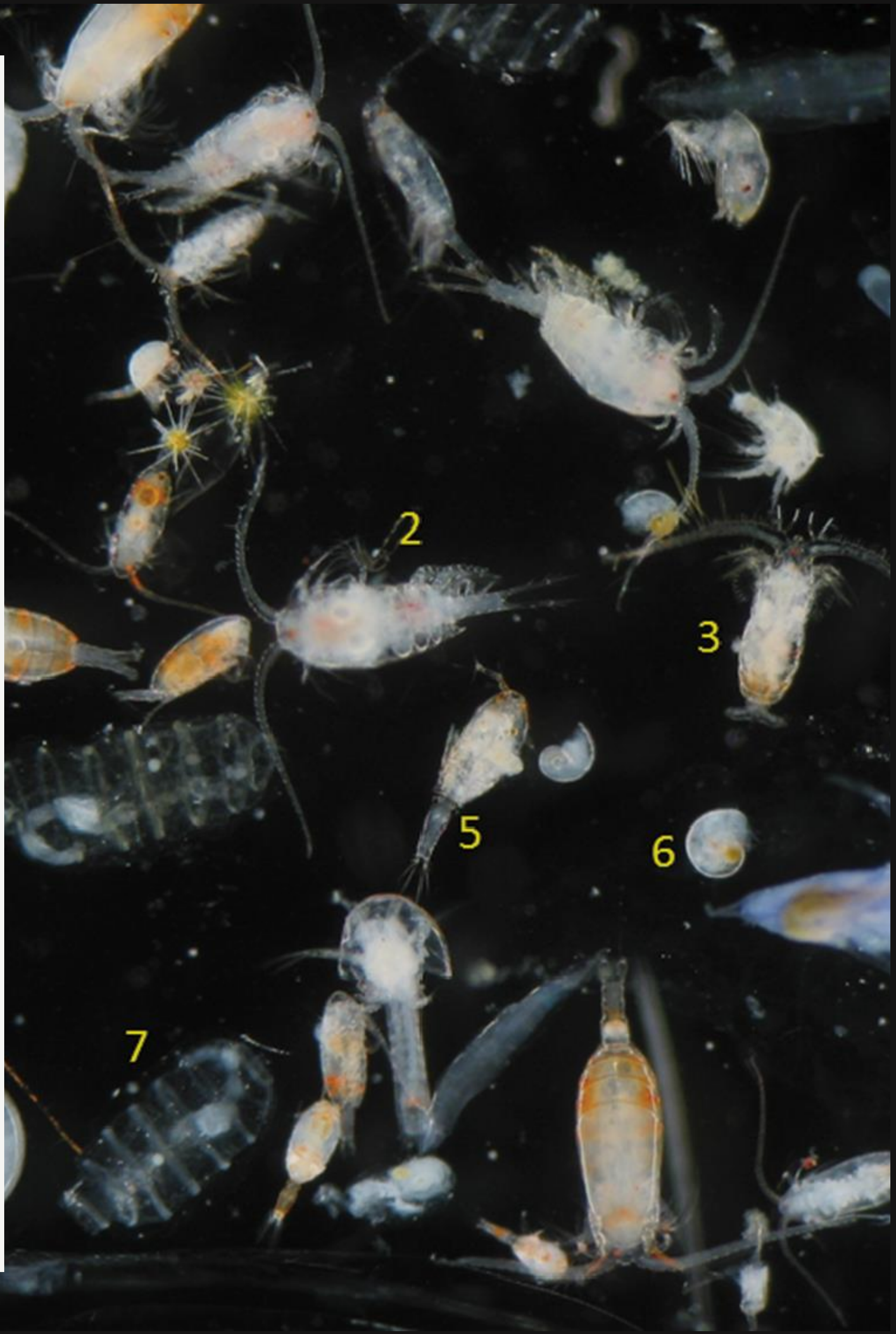
A semi-variogram is typically fit to the **model residuals**, not the original data, when used in the context of model diagnostics.



Semi-Variograms

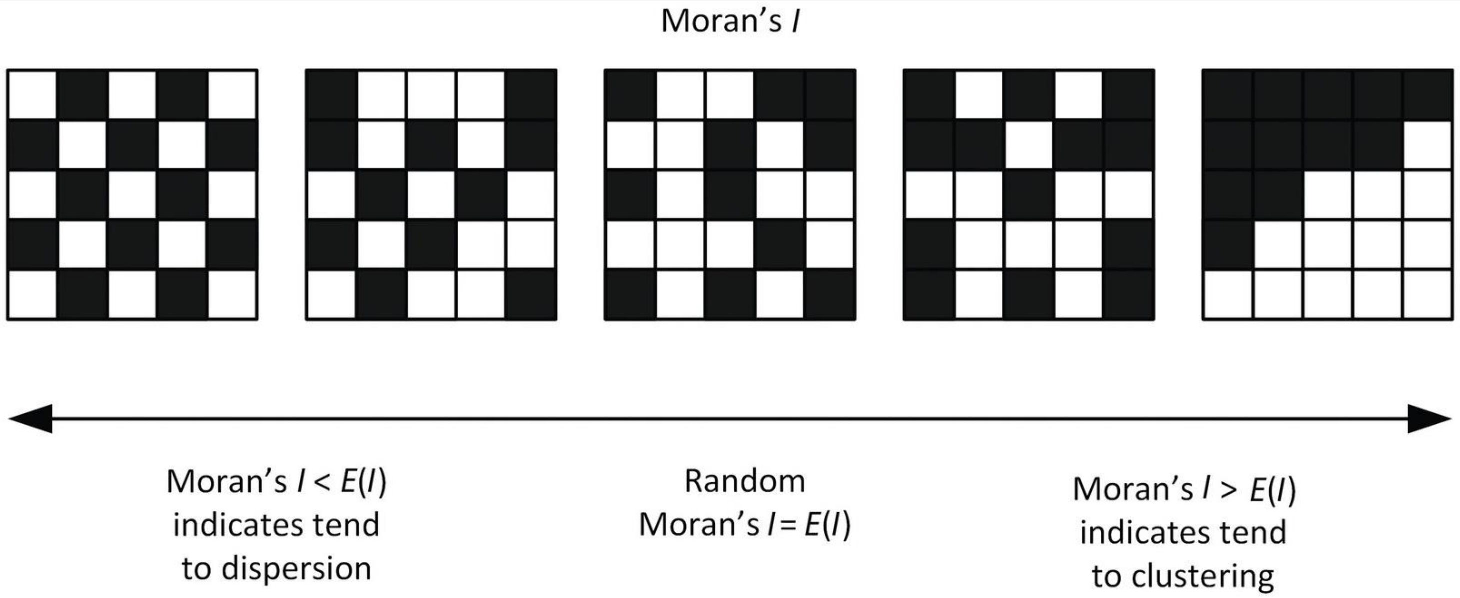


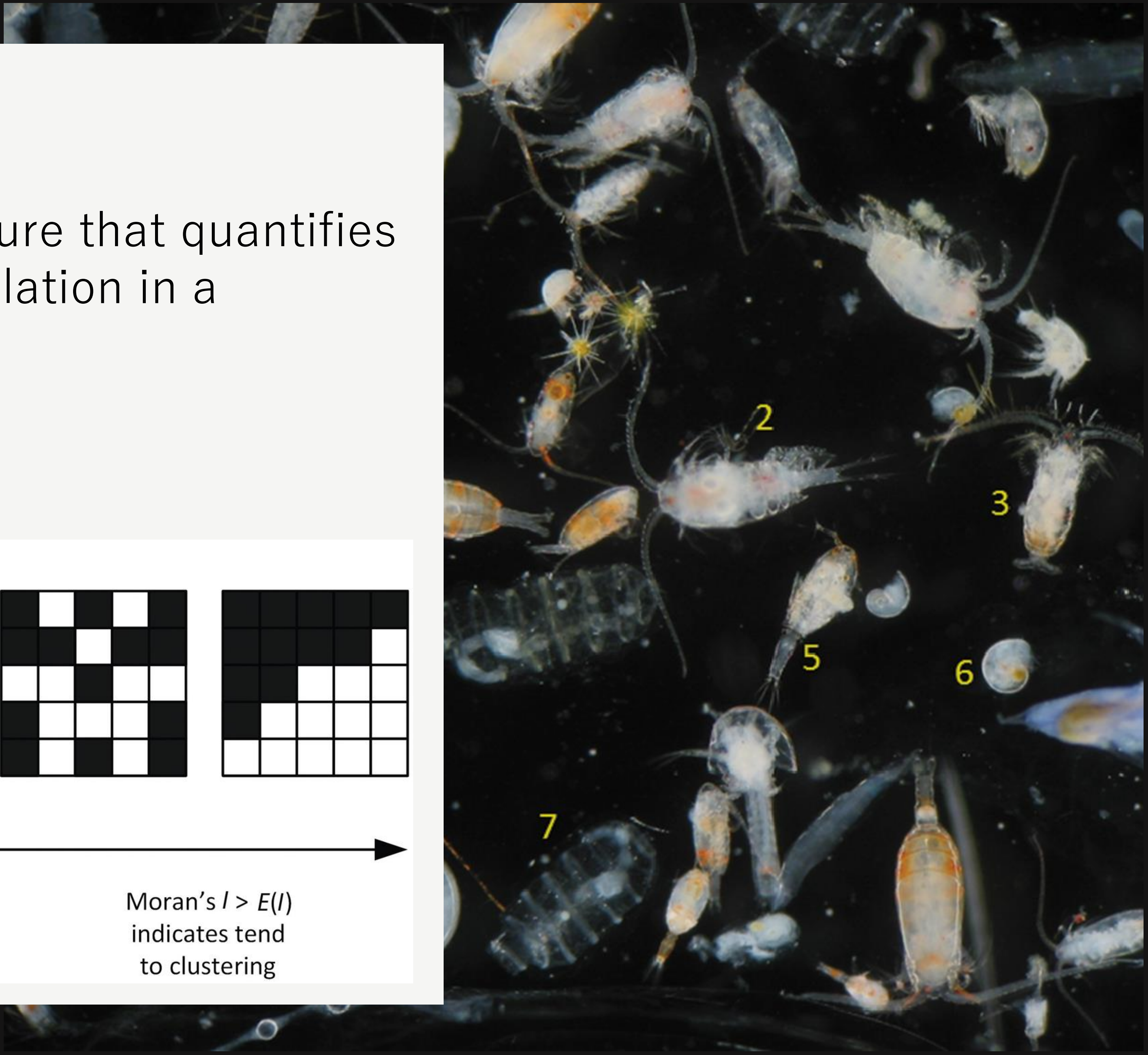
Legendre & Legendre Fig. 13.8



Moran’s *I*

Moran’s *I* is a statistical measure that quantifies the degree of spatial autocorrelation in a dataset.

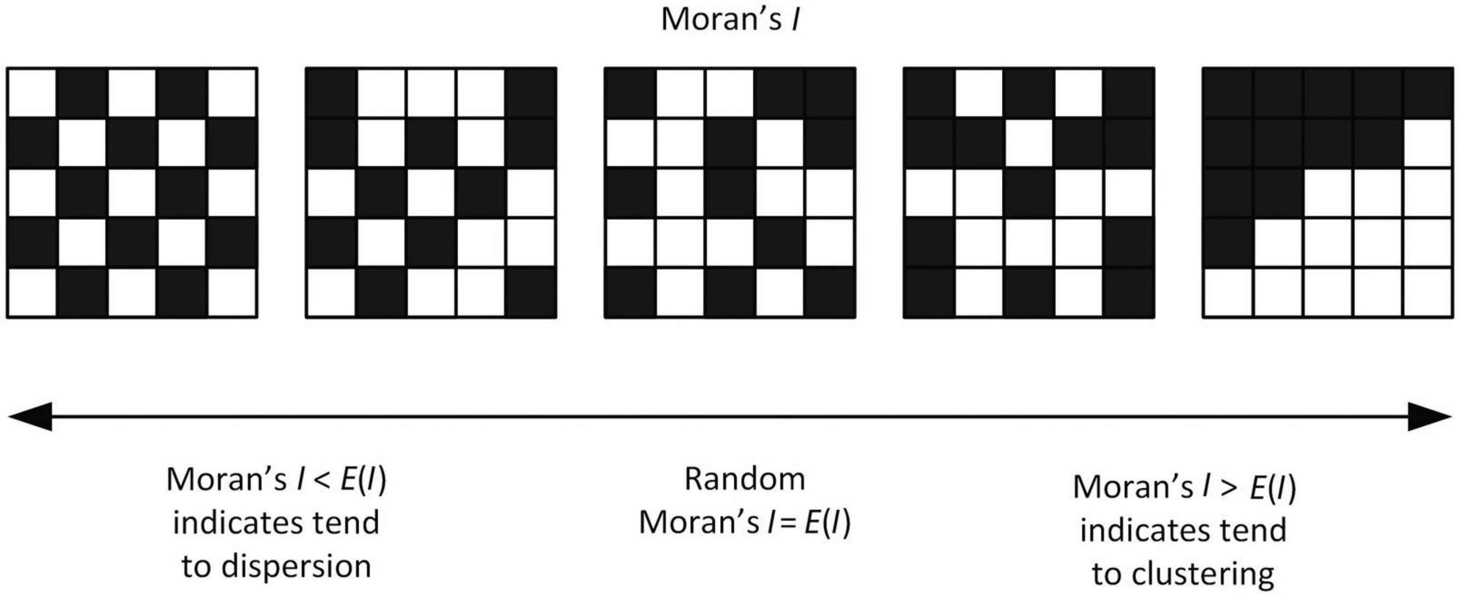




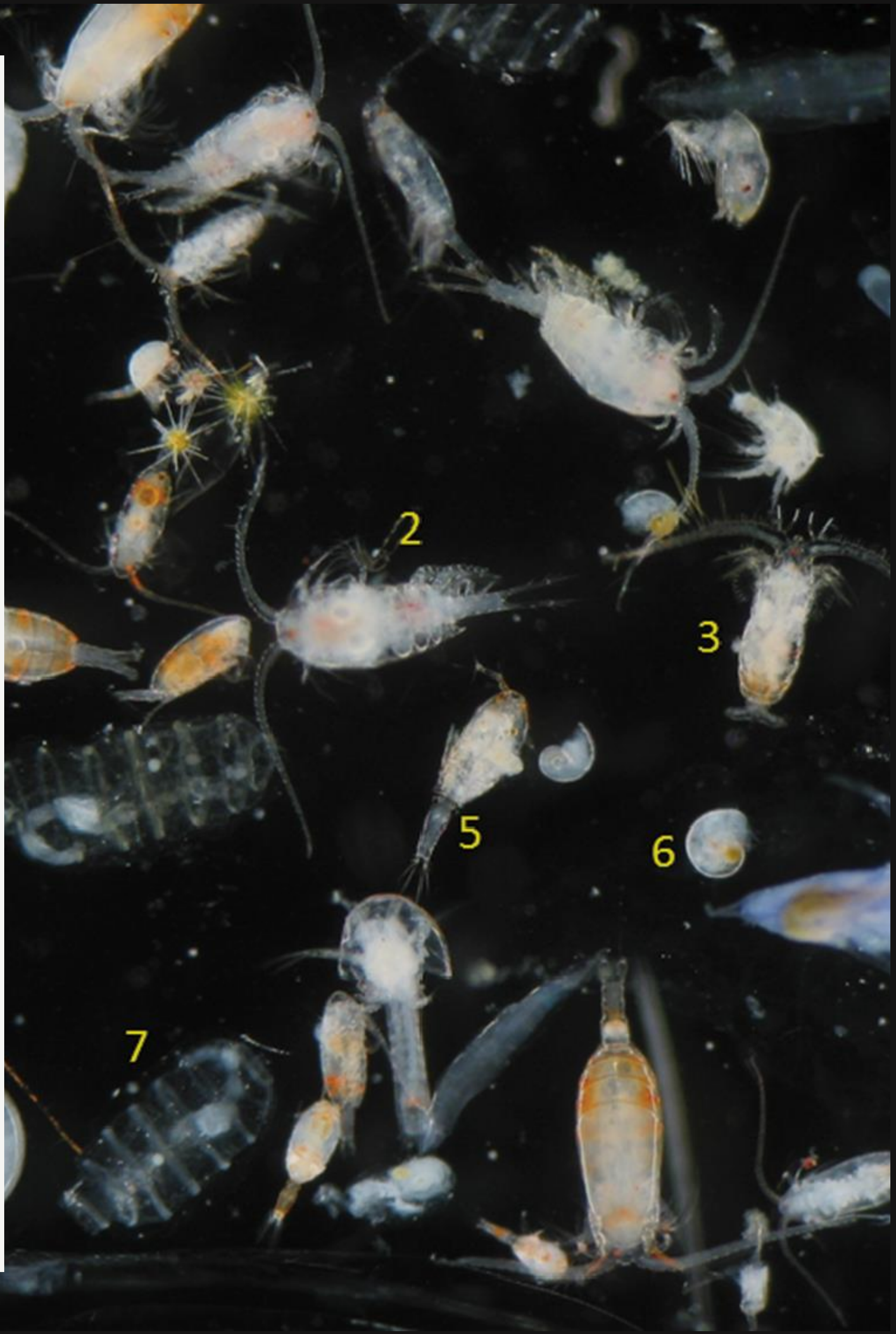
Moran’s *I*

Moran’s *I* is a statistical measure that quantifies the degree of spatial autocorrelation in a dataset.

Analogous to temporal autocorrelation

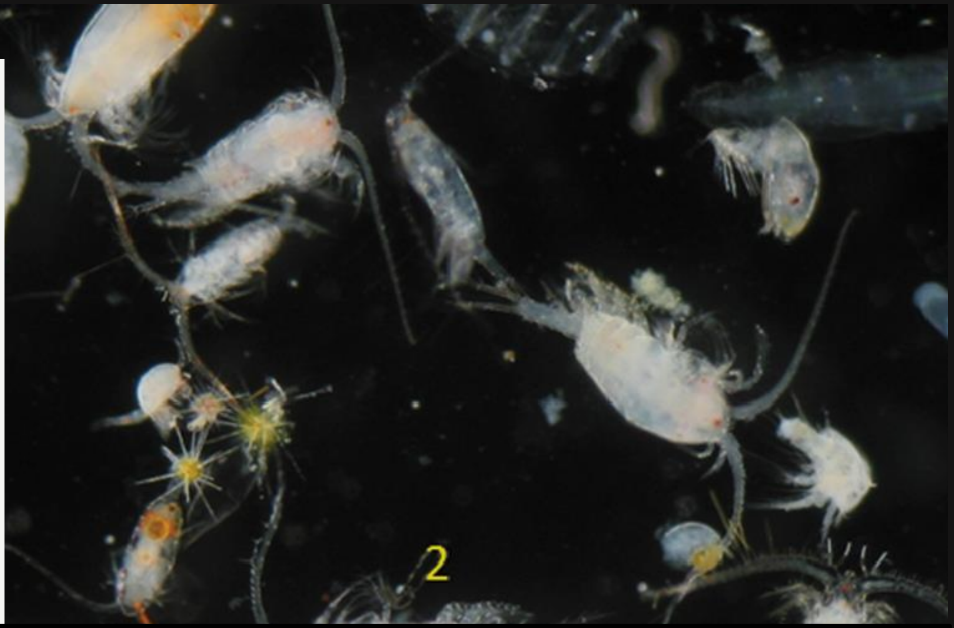


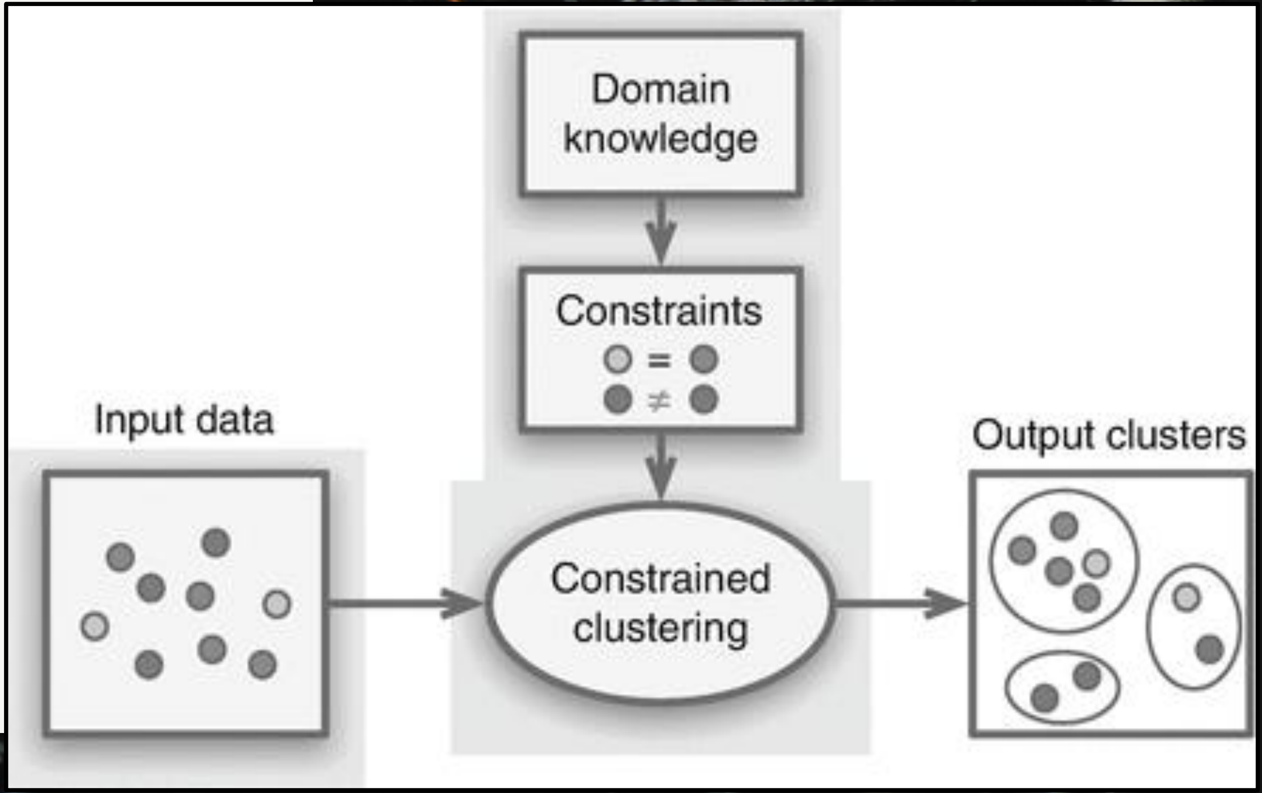
Space-Constrained Clustering



Space-Constrained Clustering

Space-constrained clustering is a type of clustering that groups data points based on both their attribute similarity *and* spatial proximity.

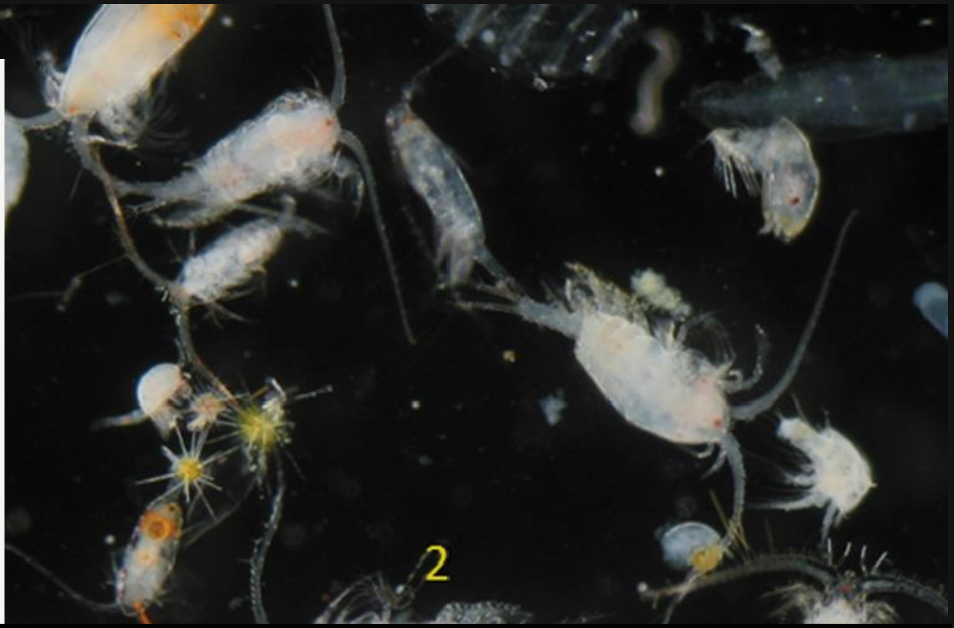


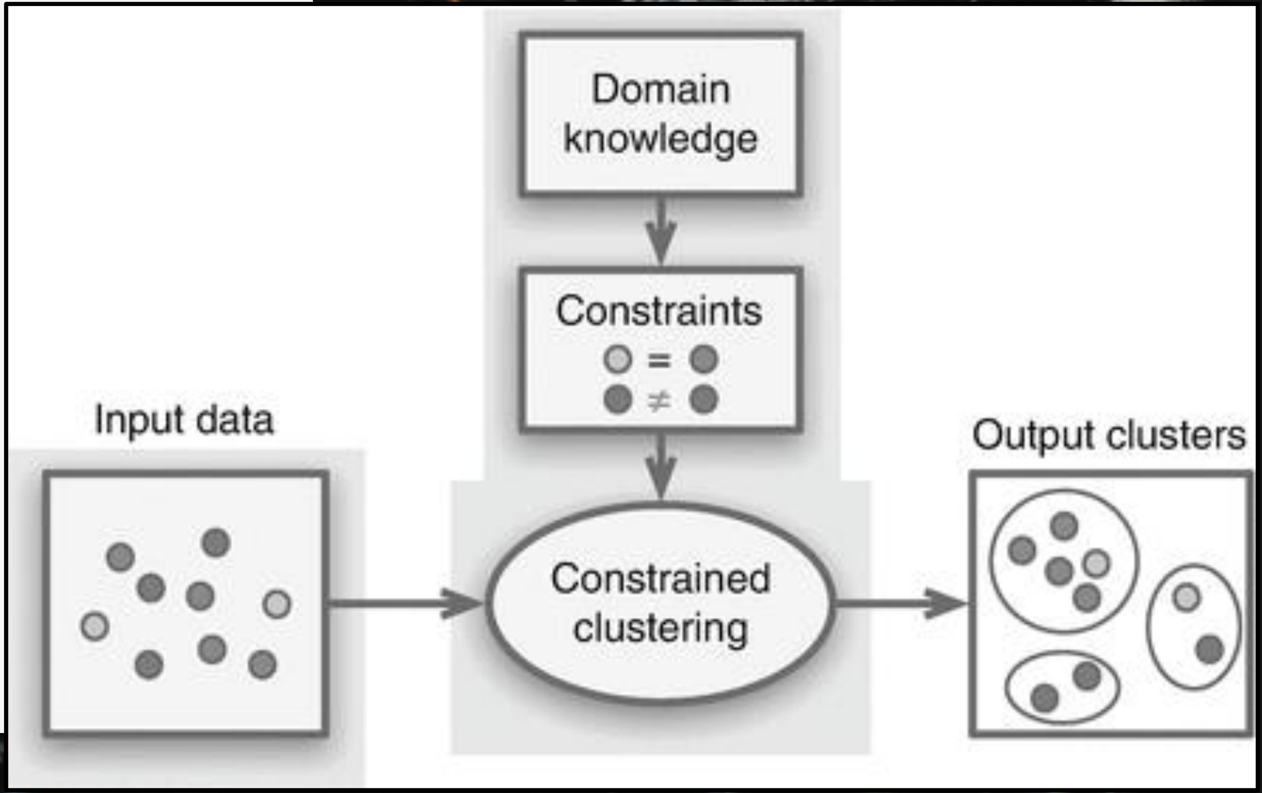


Space-Constrained Clustering

Space-constrained clustering is a type of clustering that groups data points based on both their attribute similarity *and* spatial proximity.

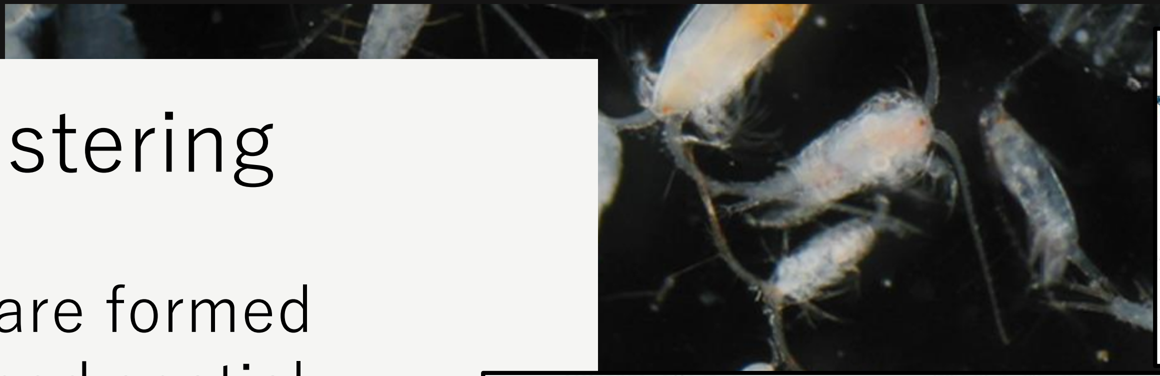
Identifies meaningful groups in spatial data while preserving the spatial relationships among data points.

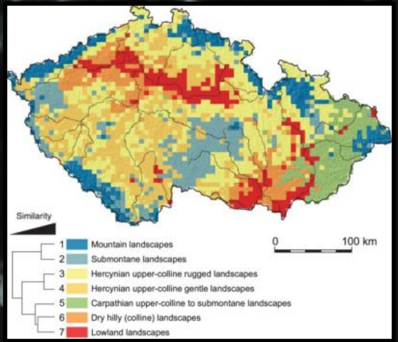


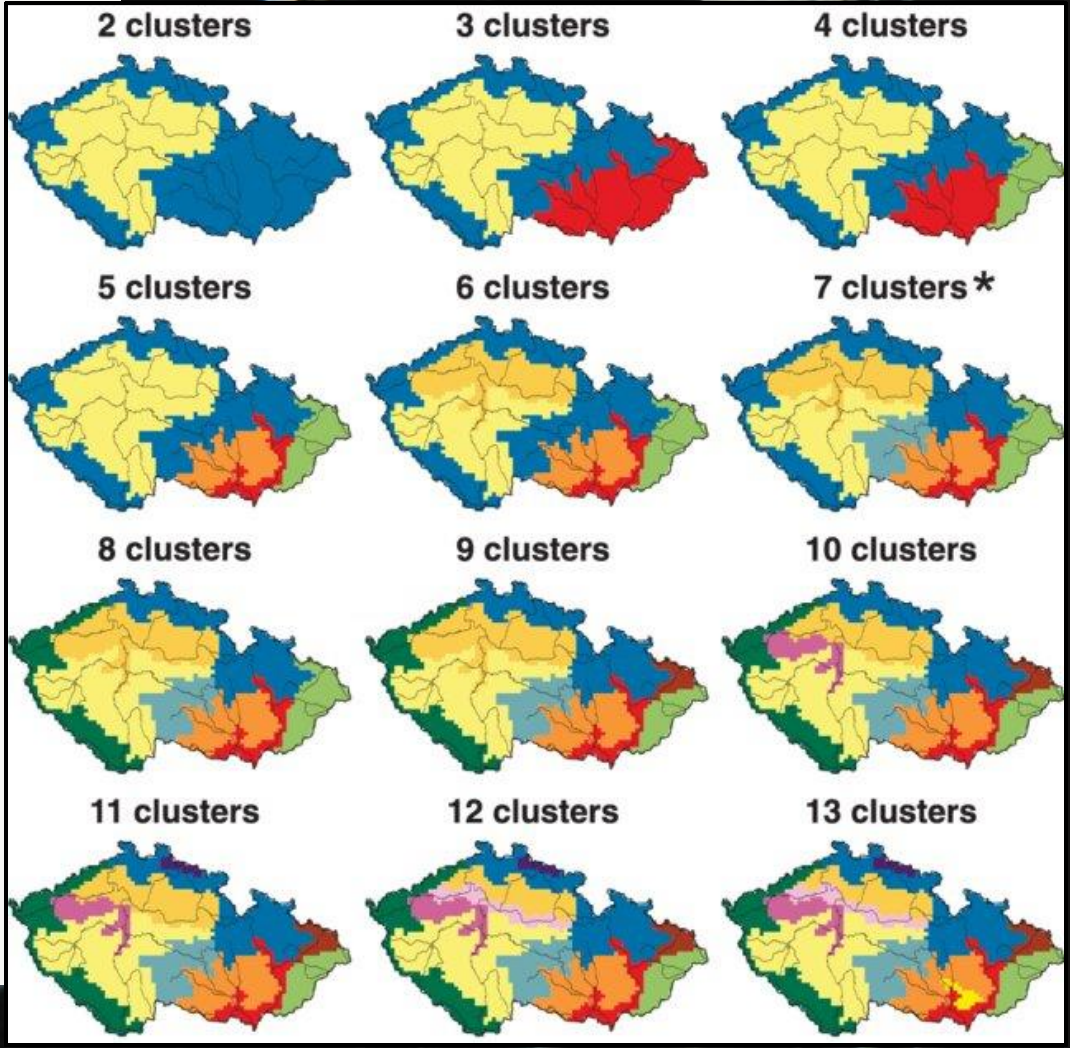


Space-Constrained Clustering

Spatial Constraints: Clusters are formed based on adjacency or predefined spatial boundaries, like geographic zones or distance thresholds.







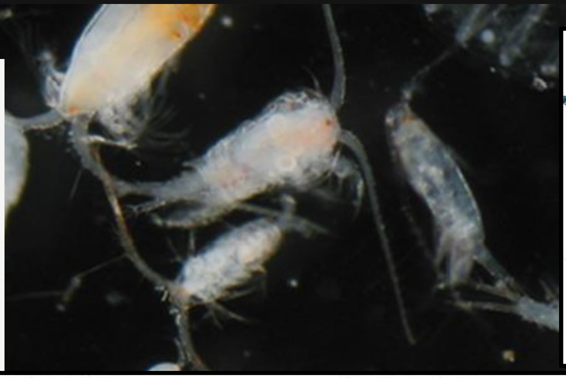
Divisek et al. 2014

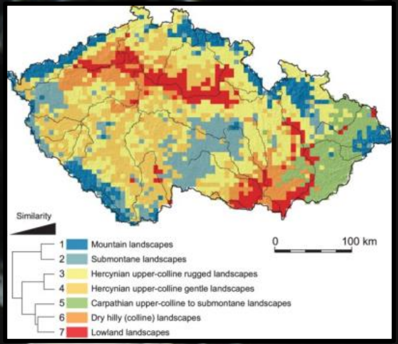
Space-Constrained Clustering

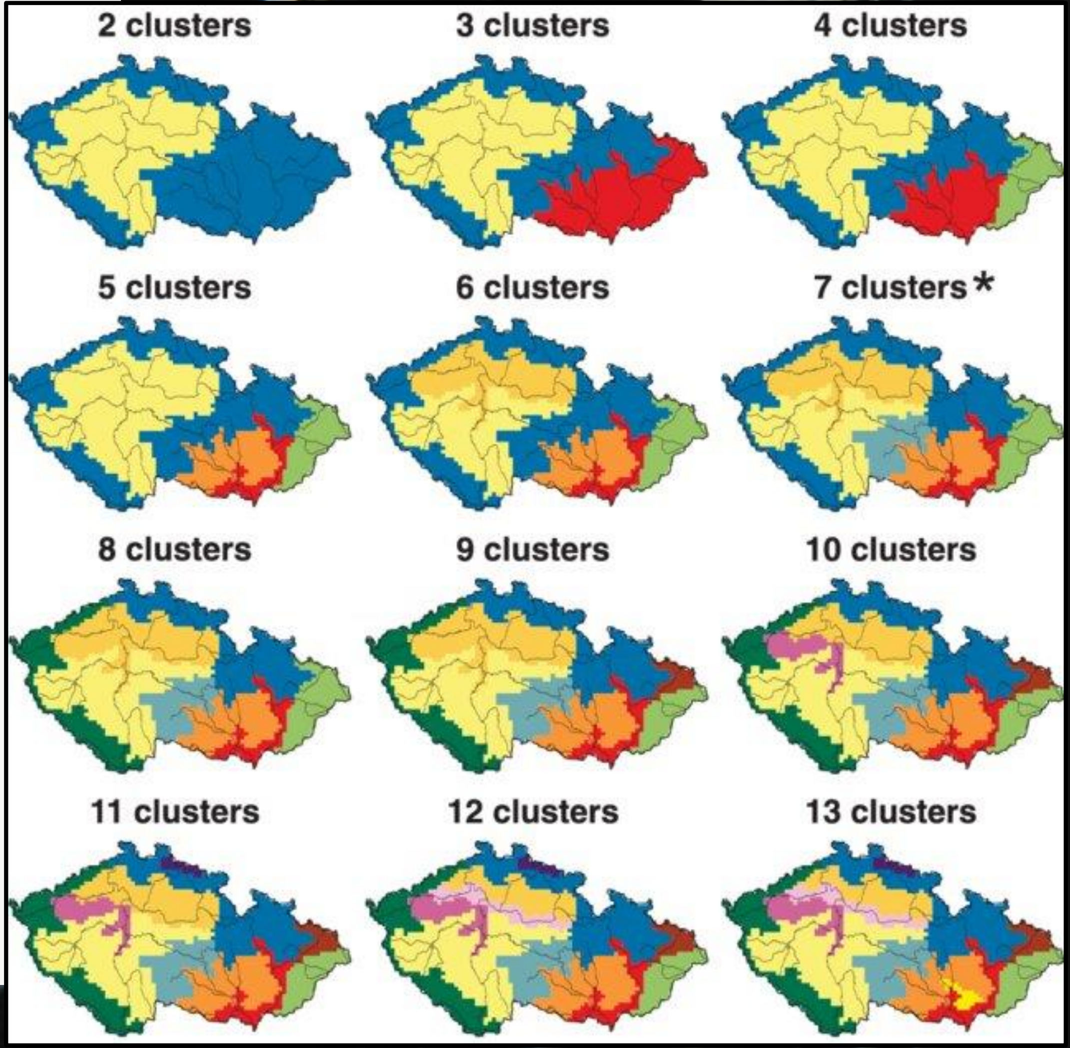
Spatial Constraints: Clusters are formed based on adjacency or predefined spatial boundaries, like geographic zones or distance thresholds.

Clustering Algorithm:

- Includes: constrained k-means, spatially constrained hierarchical clustering, or region-growing algorithms.
- Typically combine attribute distance (such as Euclidean distance in feature space) with spatial criteria (e.g., maximum allowable distance between cluster members).



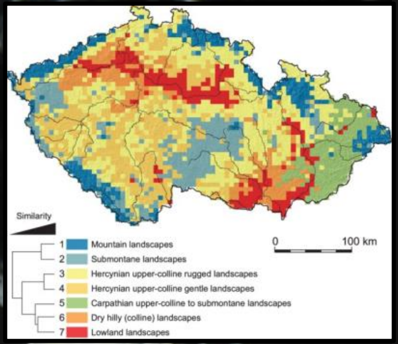


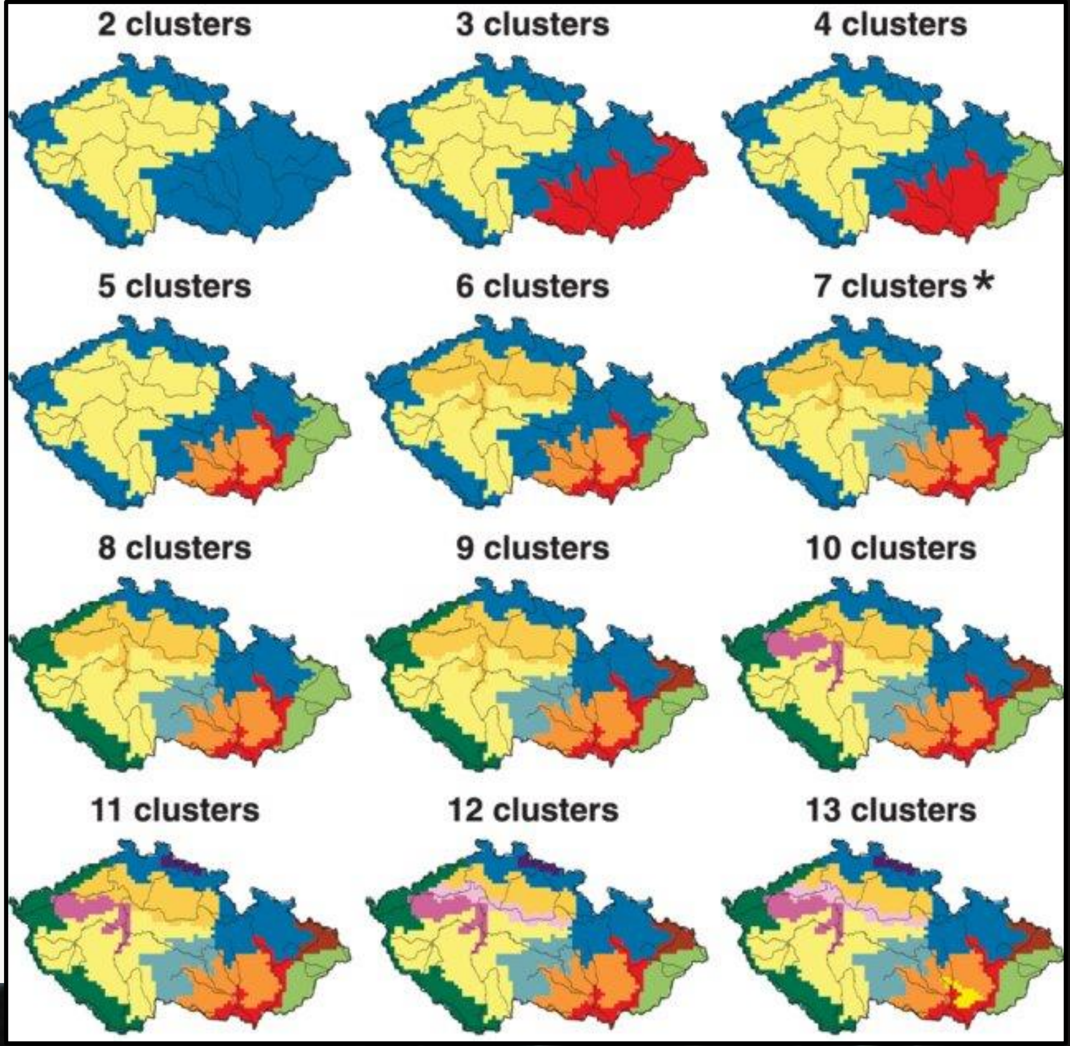


Divisek et al. 2014

Space-Constrained Clustering

Constrained solutions may be less optimal than unconstrained solutions in terms of the clustering criterion.



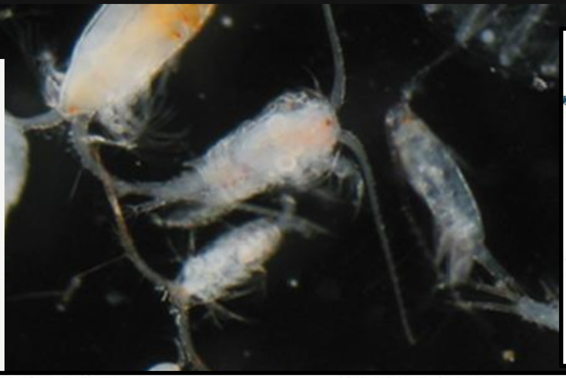


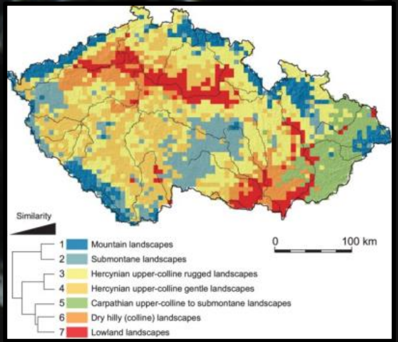
Divisek et al. 2014

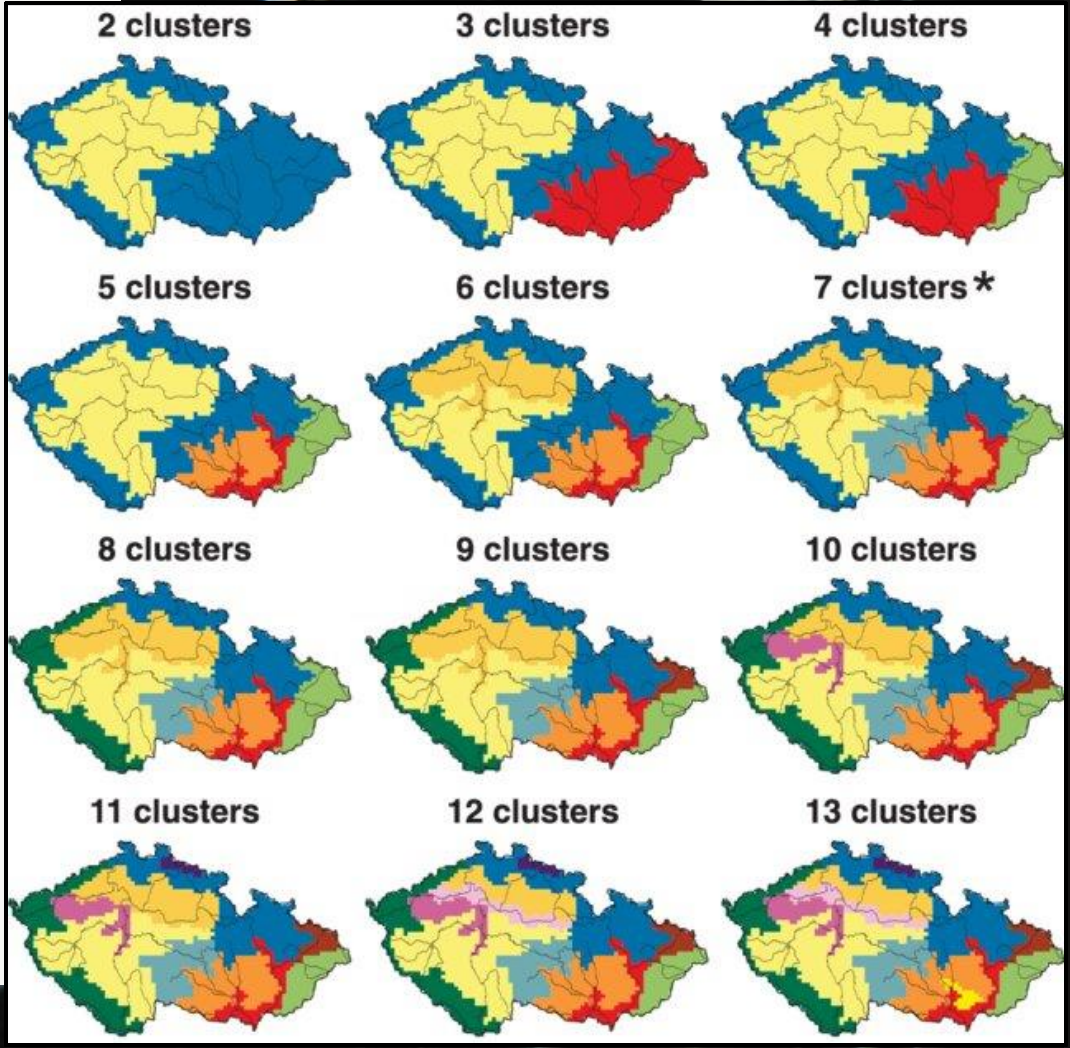
Space-Constrained Clustering

Uses:

- Delineating ecological regions, administrative units, or resource distribution networks
- Relate clusters to geographically-related causal factors (e.g., geological data)
- Designing stratified sampling programs
- Hypothesis tests of ecological similarity

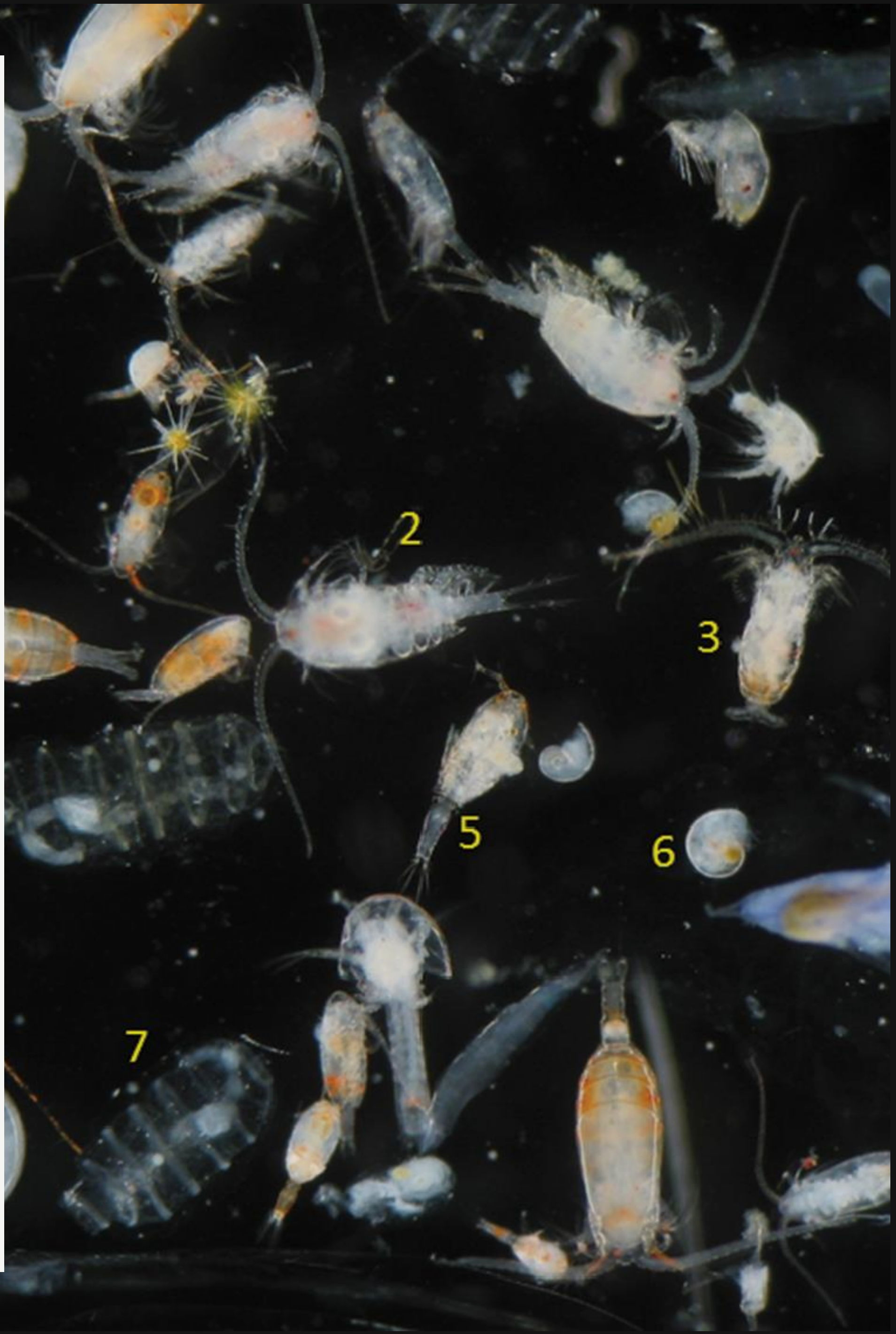






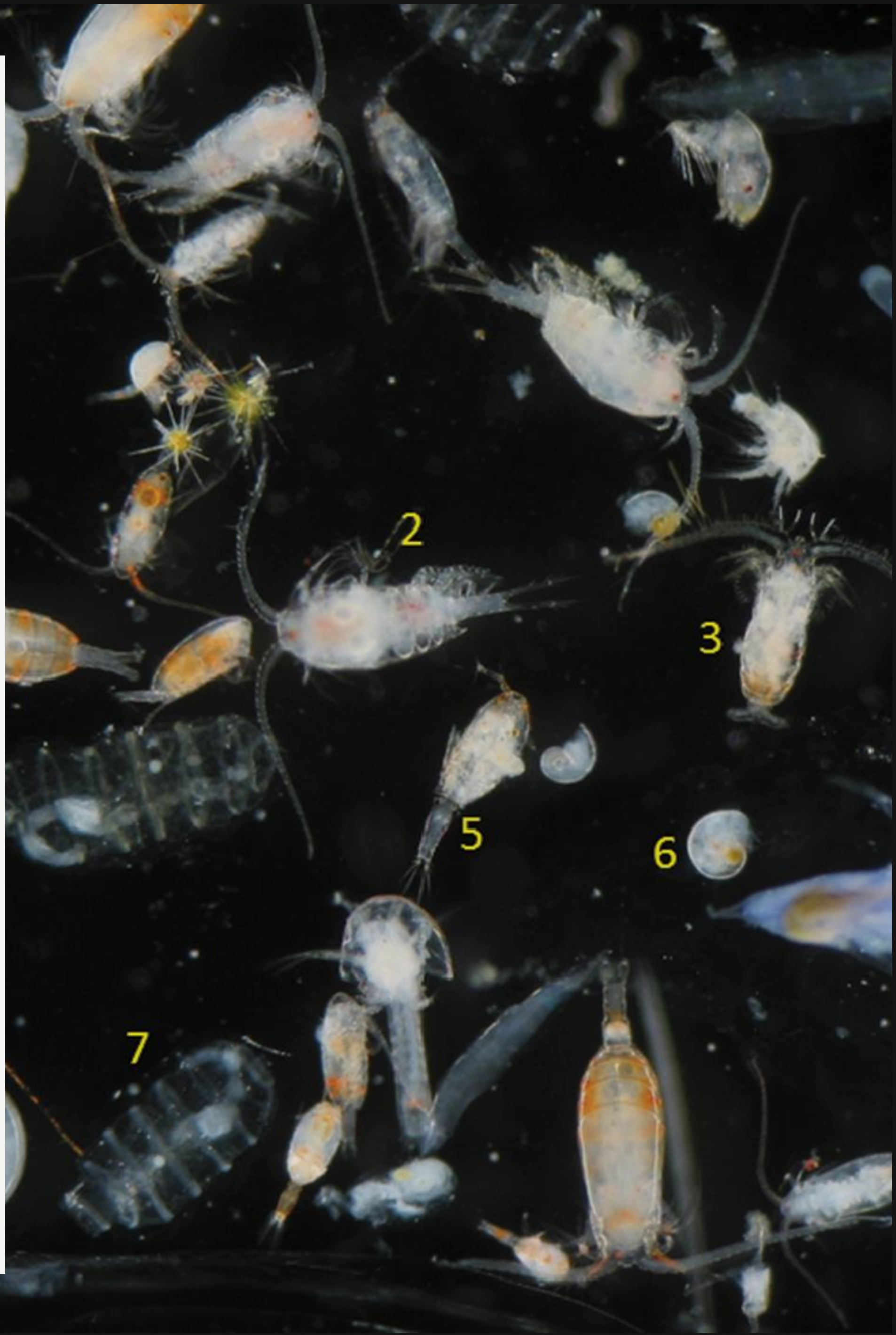
Divisek et al. 2014

Interpolated Ordination



Interpolated Ordination

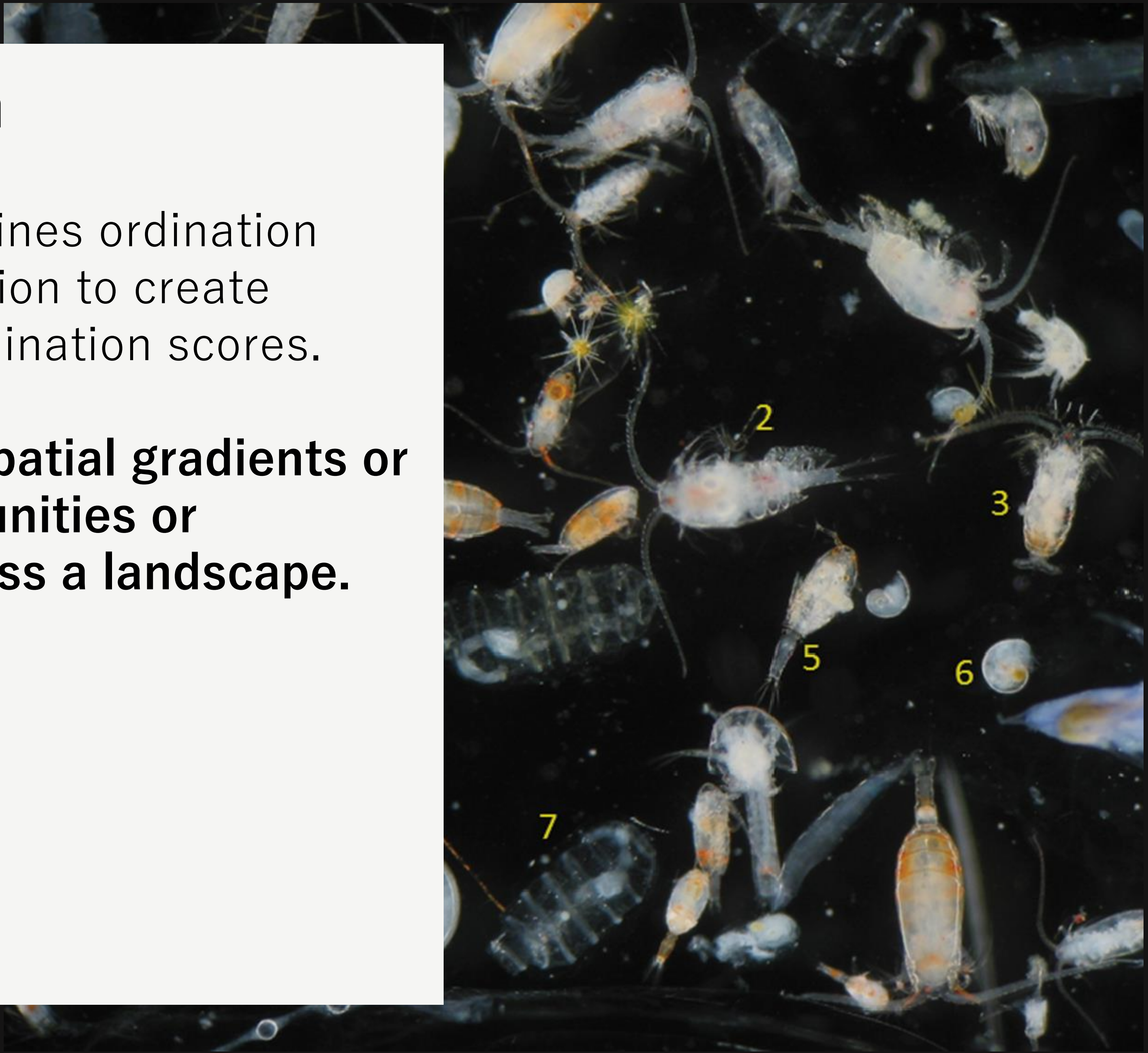
Interpolated ordination combines ordination analysis with spatial interpolation to create continuous spatial maps of ordination scores.



Interpolated Ordination

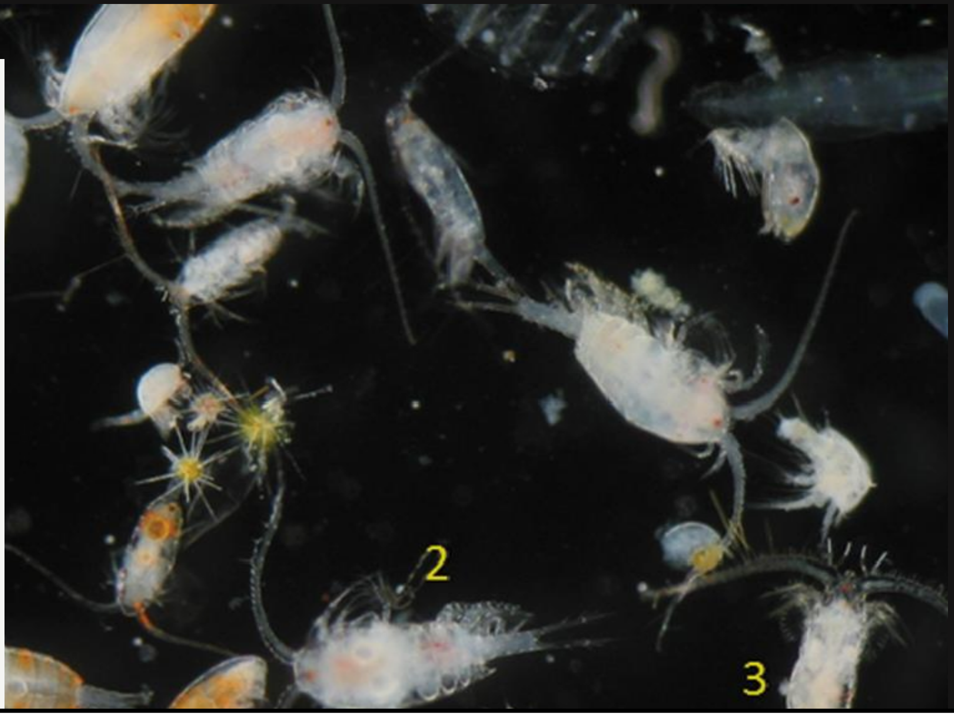
Interpolated ordination combines ordination analysis with spatial interpolation to create continuous spatial maps of ordination scores.

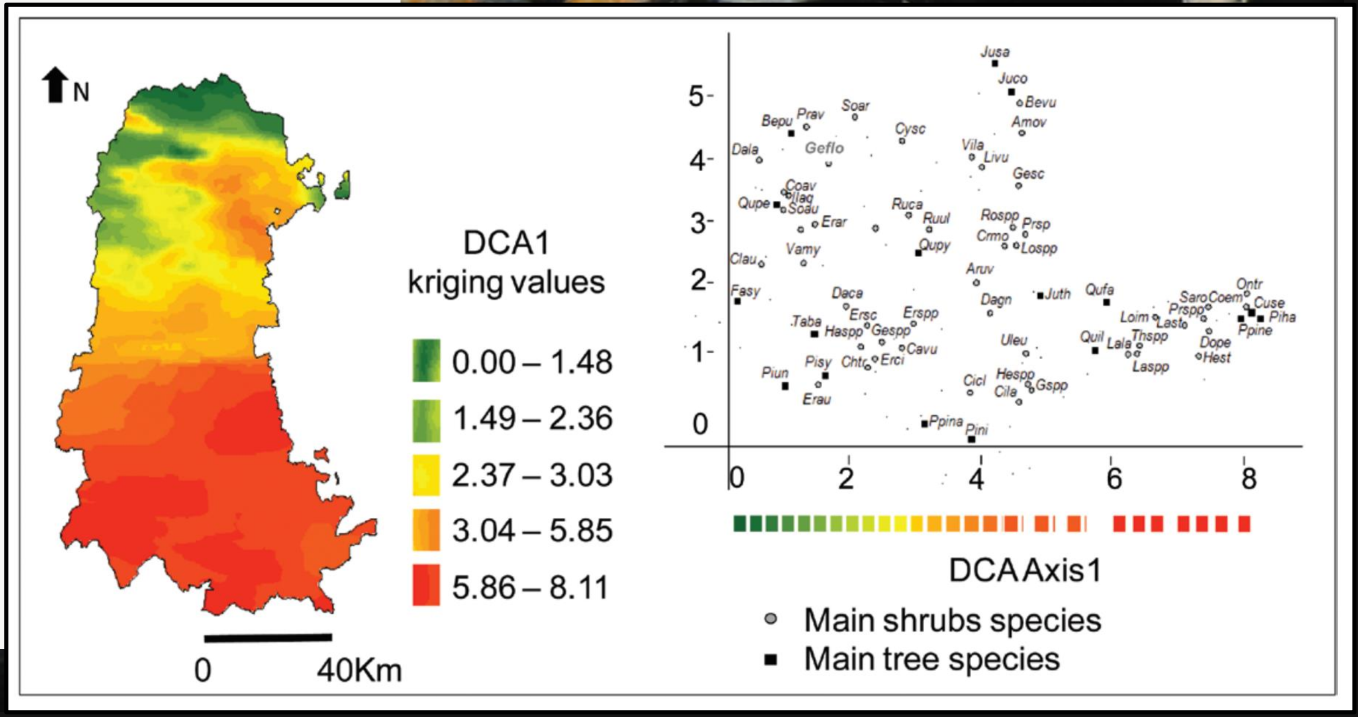
Provides a way to visualize spatial gradients or patterns in ecological communities or environmental variables across a landscape.



Interpolated Ordination

Step 1) Perform unconstrained ordination on multivariate data using **PCA, PCoA, NMDS**, etc.



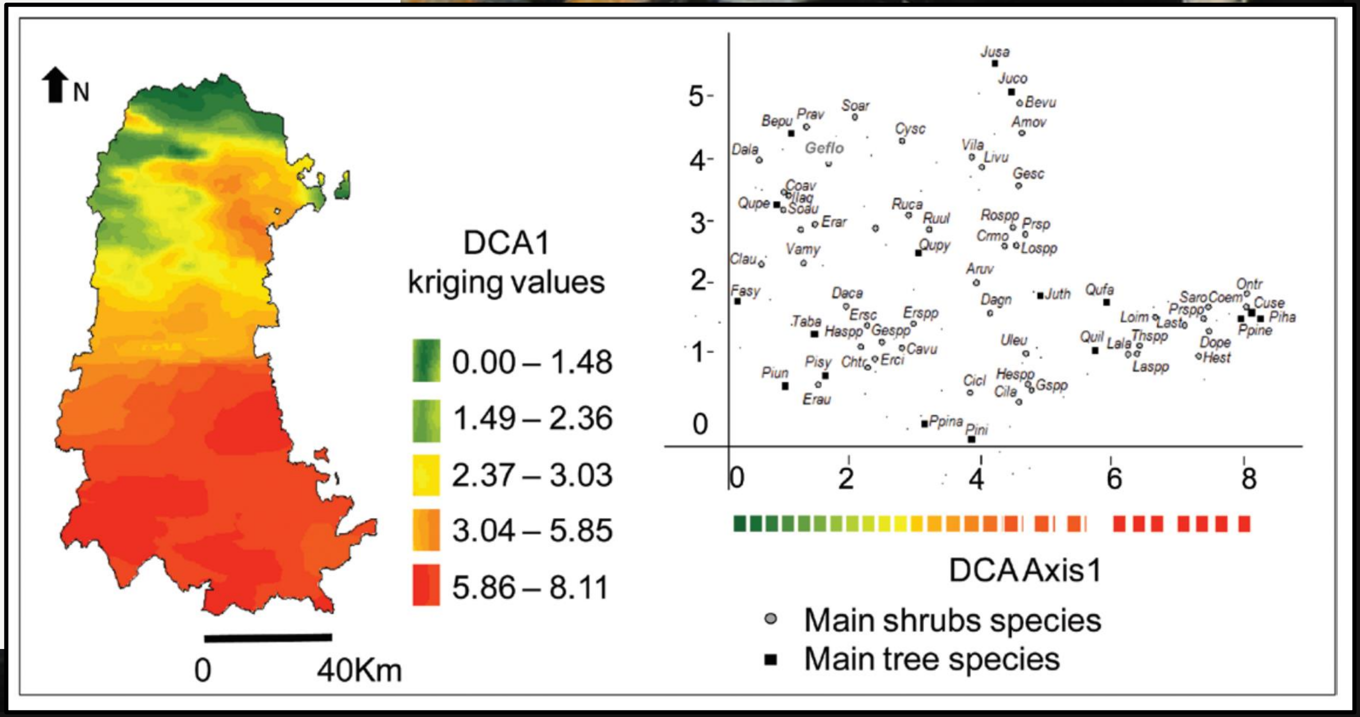


Olthoff et al. 2016

Interpolated Ordination

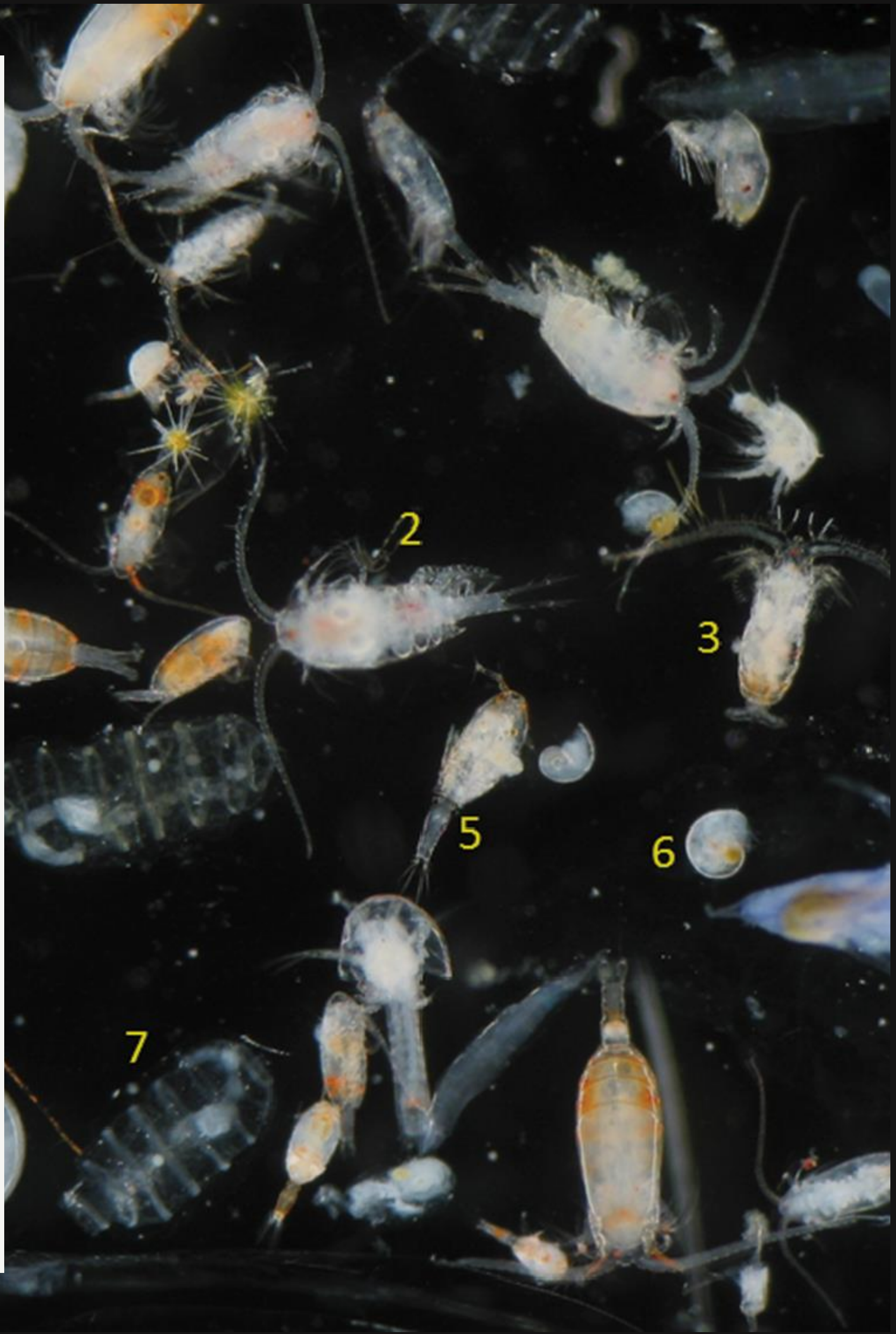
Step 1) Perform unconstrained ordination on multivariate data using **PCA, PCoA, NMDS**, etc.

Step 2) Use spatial interpolation (e.g., kriging, inverse distance weighting) on ordination scores to create continuous maps.



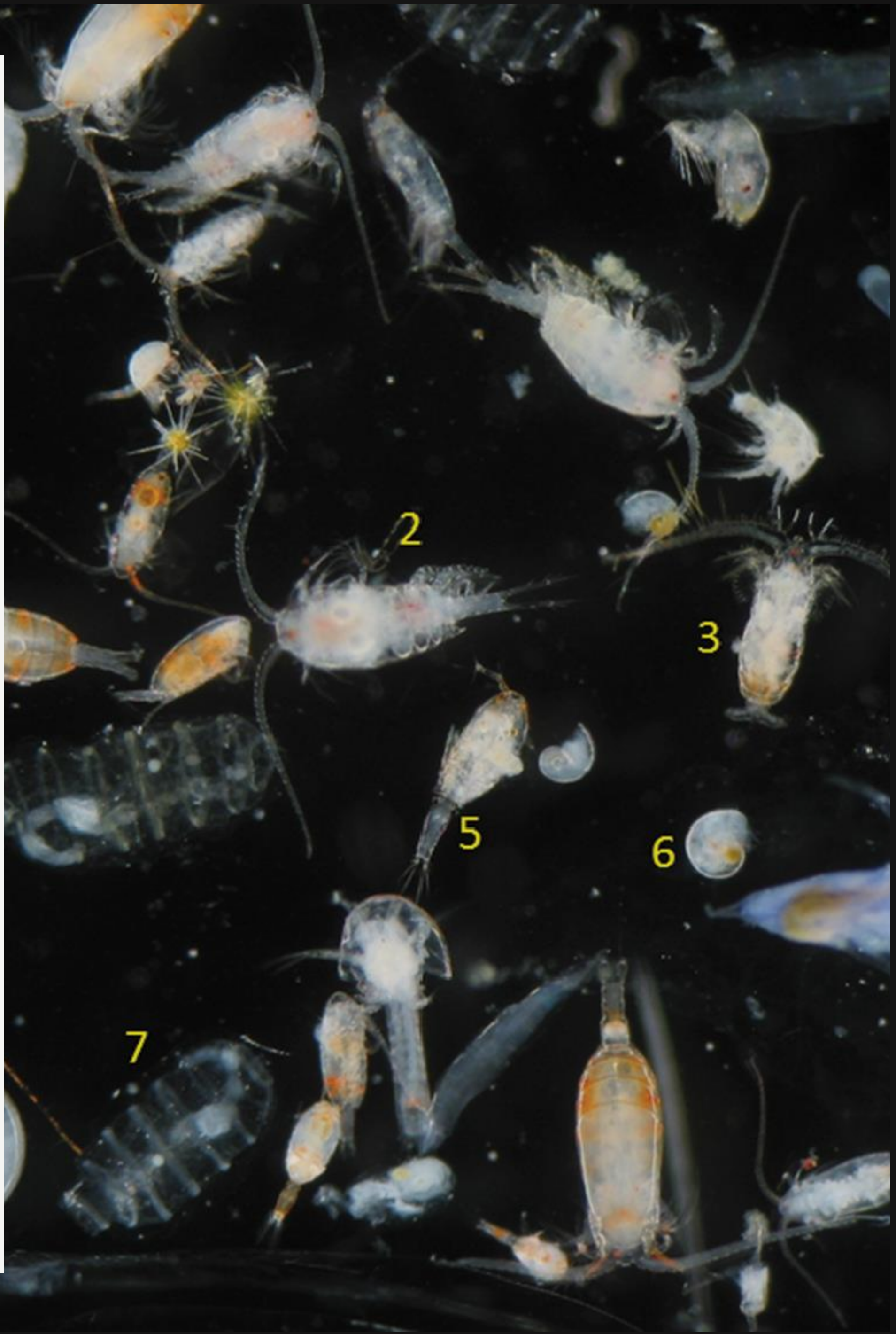
Olthoff et al. 2016

Canonical Spatial Modeling



Canonical Spatial Modeling

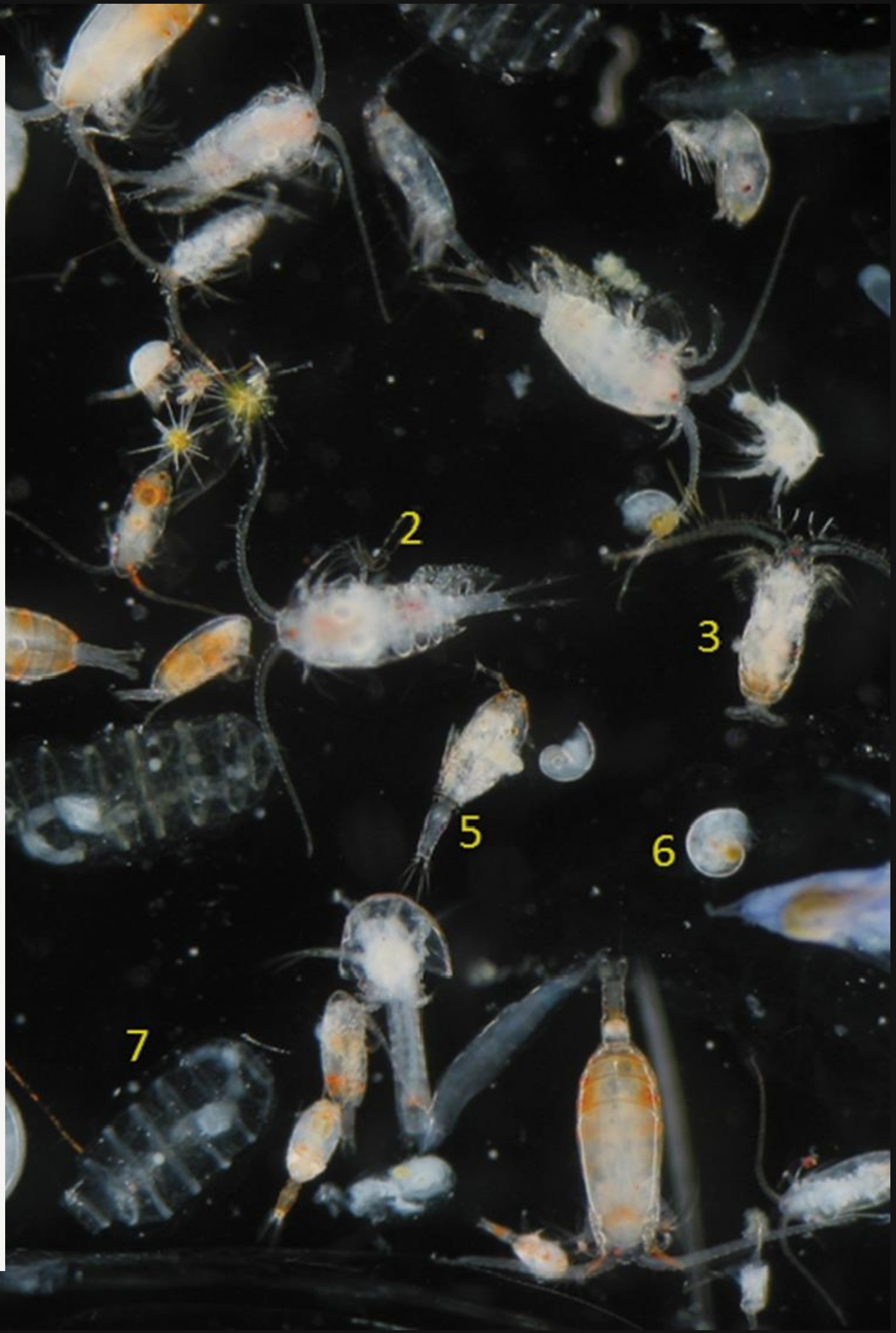
In constrained ordination (RDA/CCA), spatial variables can be included as covariates (e.g., in a **partial CCA**) to control for spatial autocorrelation, and to isolate the effect of environmental variables.



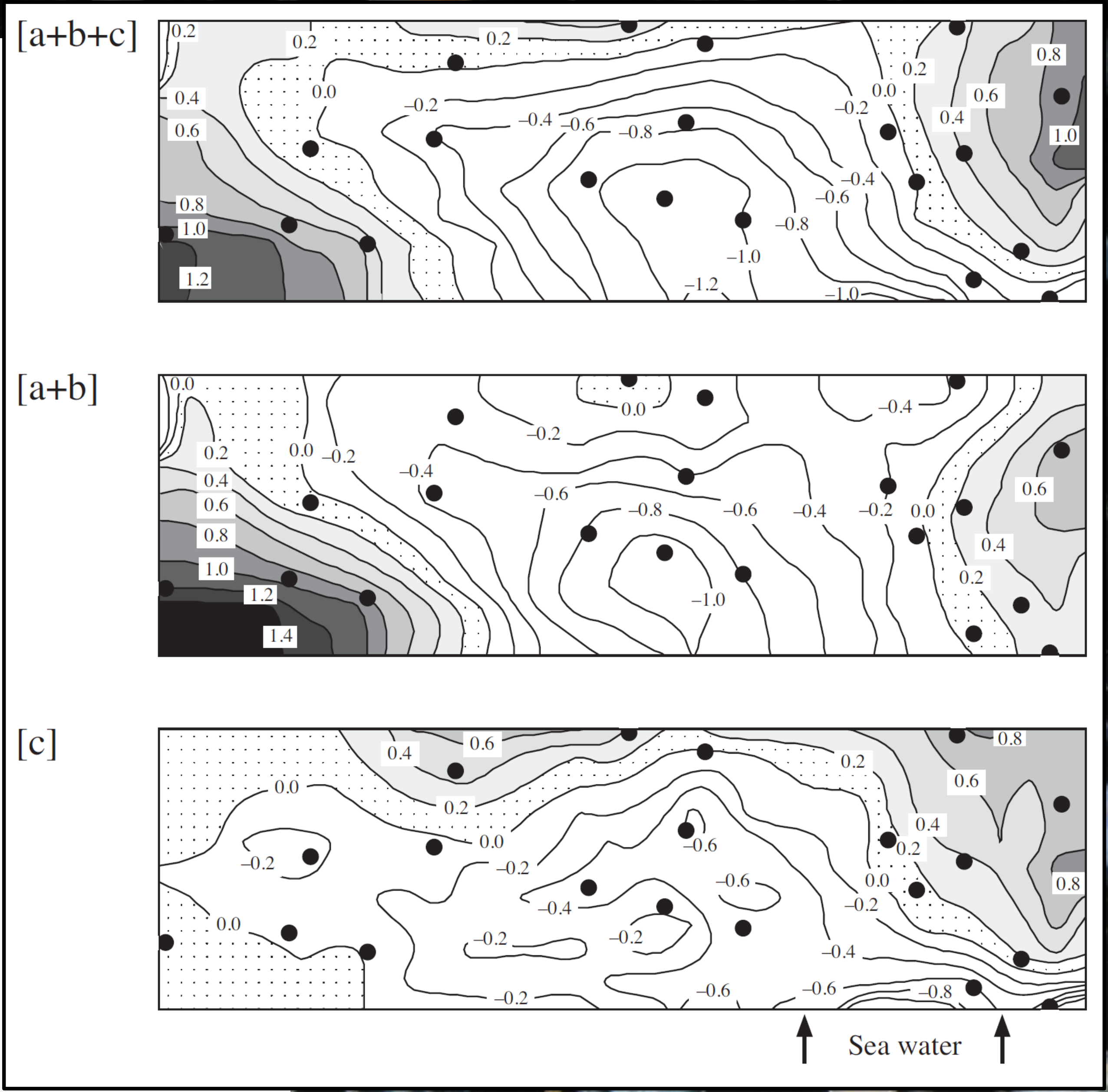
Canonical Spatial Modeling

In constrained ordination (RDA/CCA), spatial variables can be included as covariates (e.g., in a **partial CCA**) to control for spatial autocorrelation, and to isolate the effect of environmental variables.

By integrating spatial information, canonical analysis can separate the effects of spatial structure from other environmental influences.

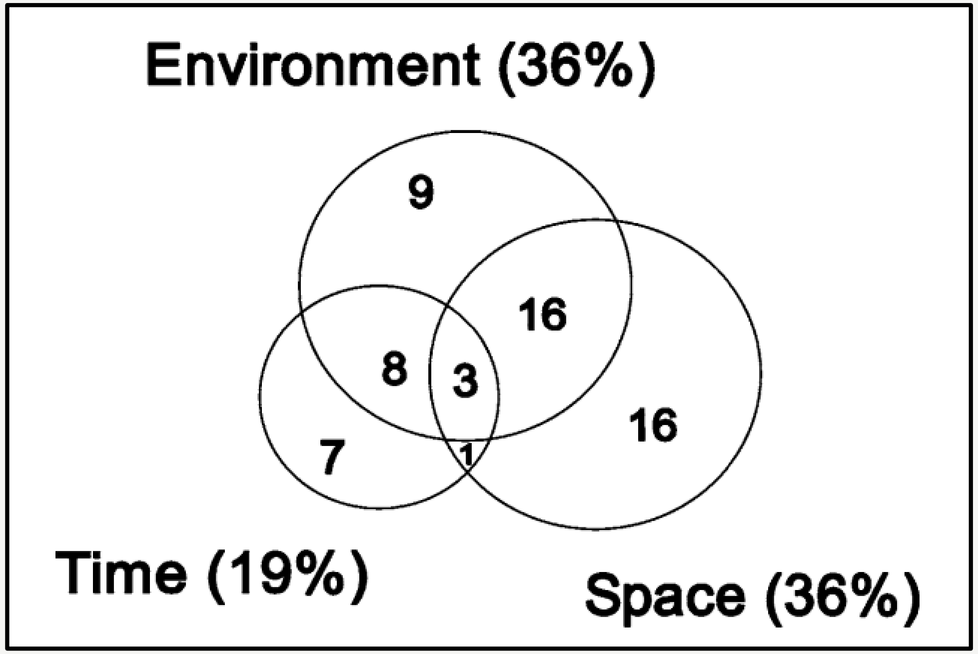


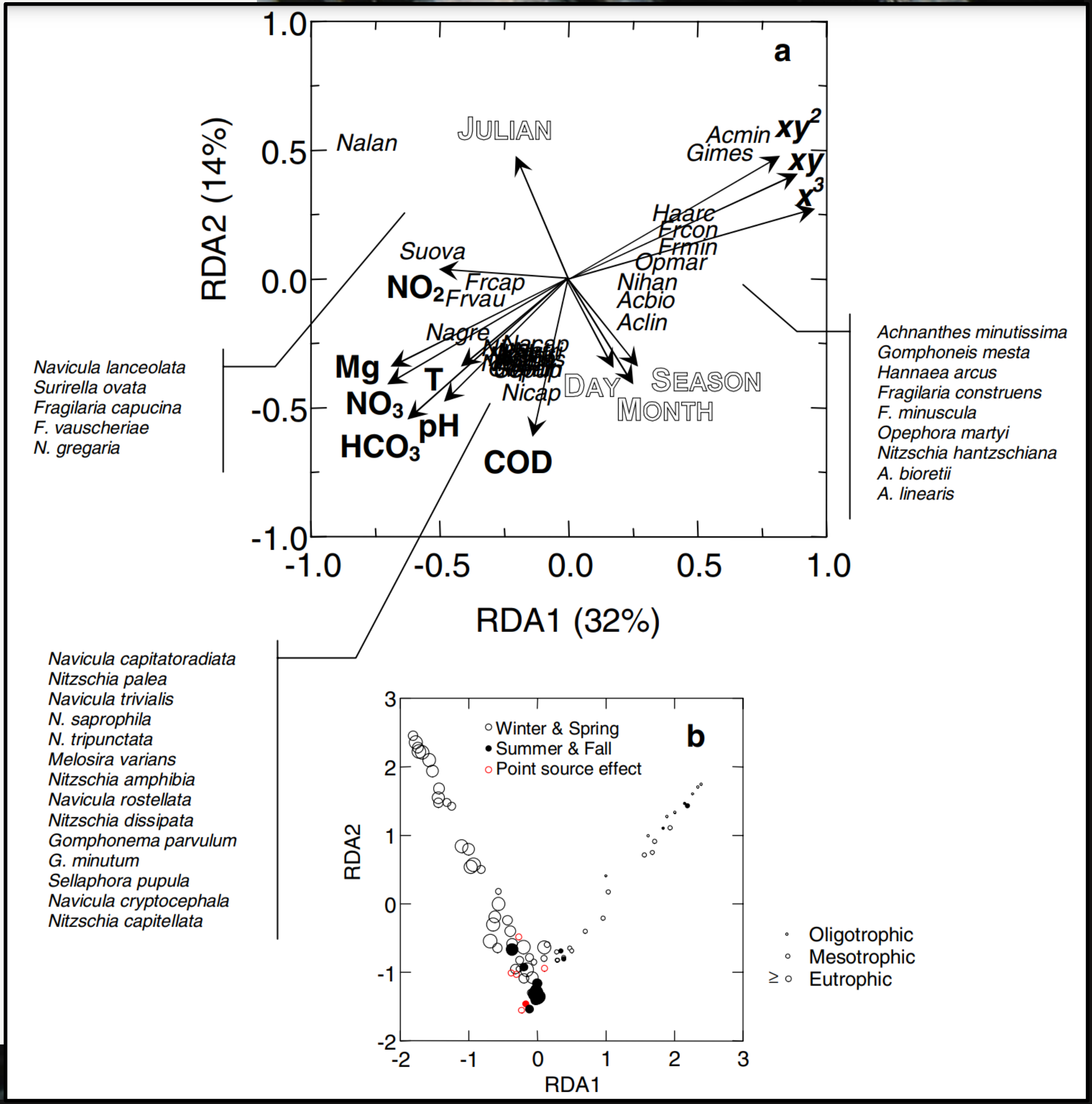
Canonical Spatial Modeling



Legendre & Legendre 13.30

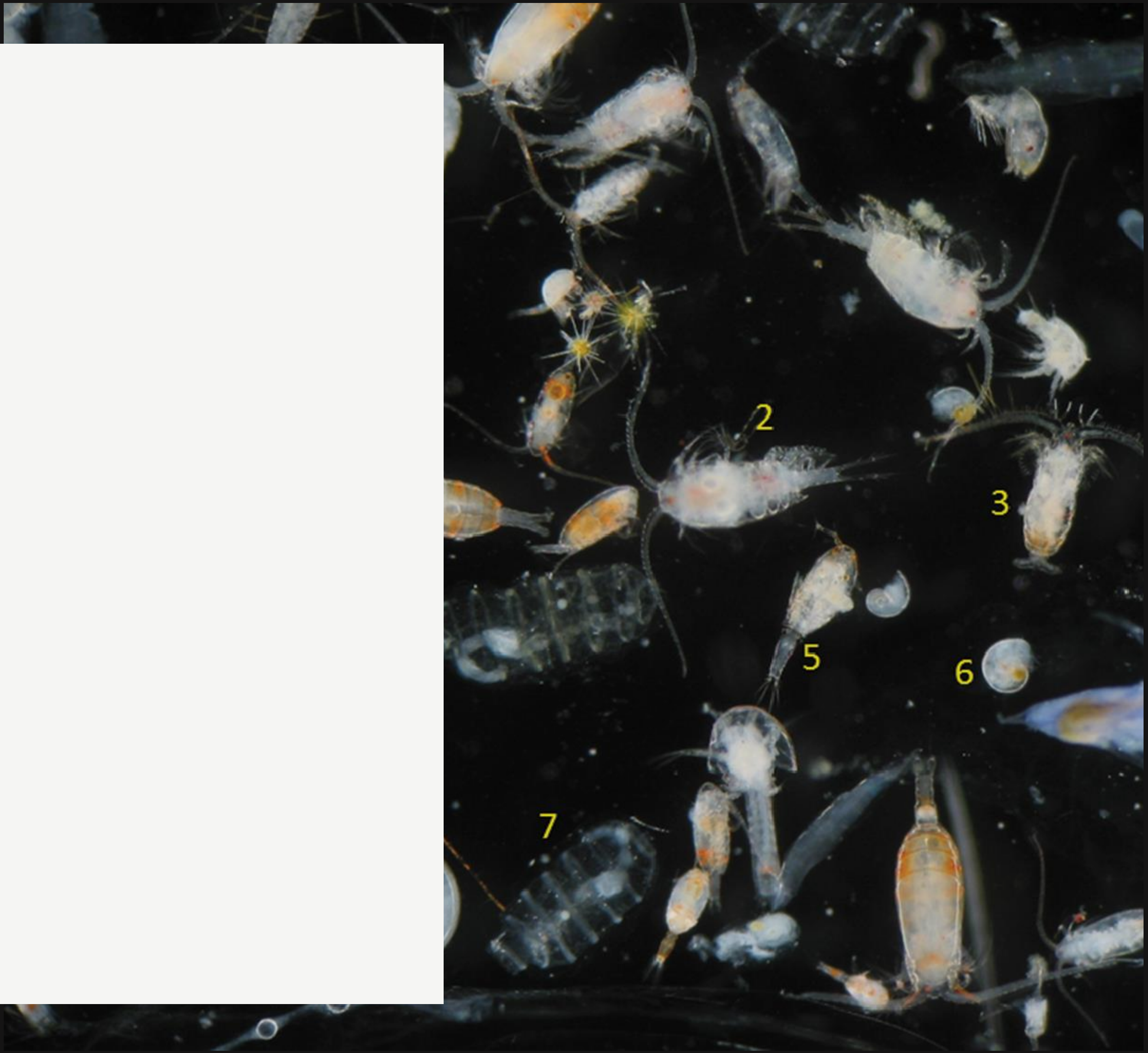
Canonical Spatial Modeling





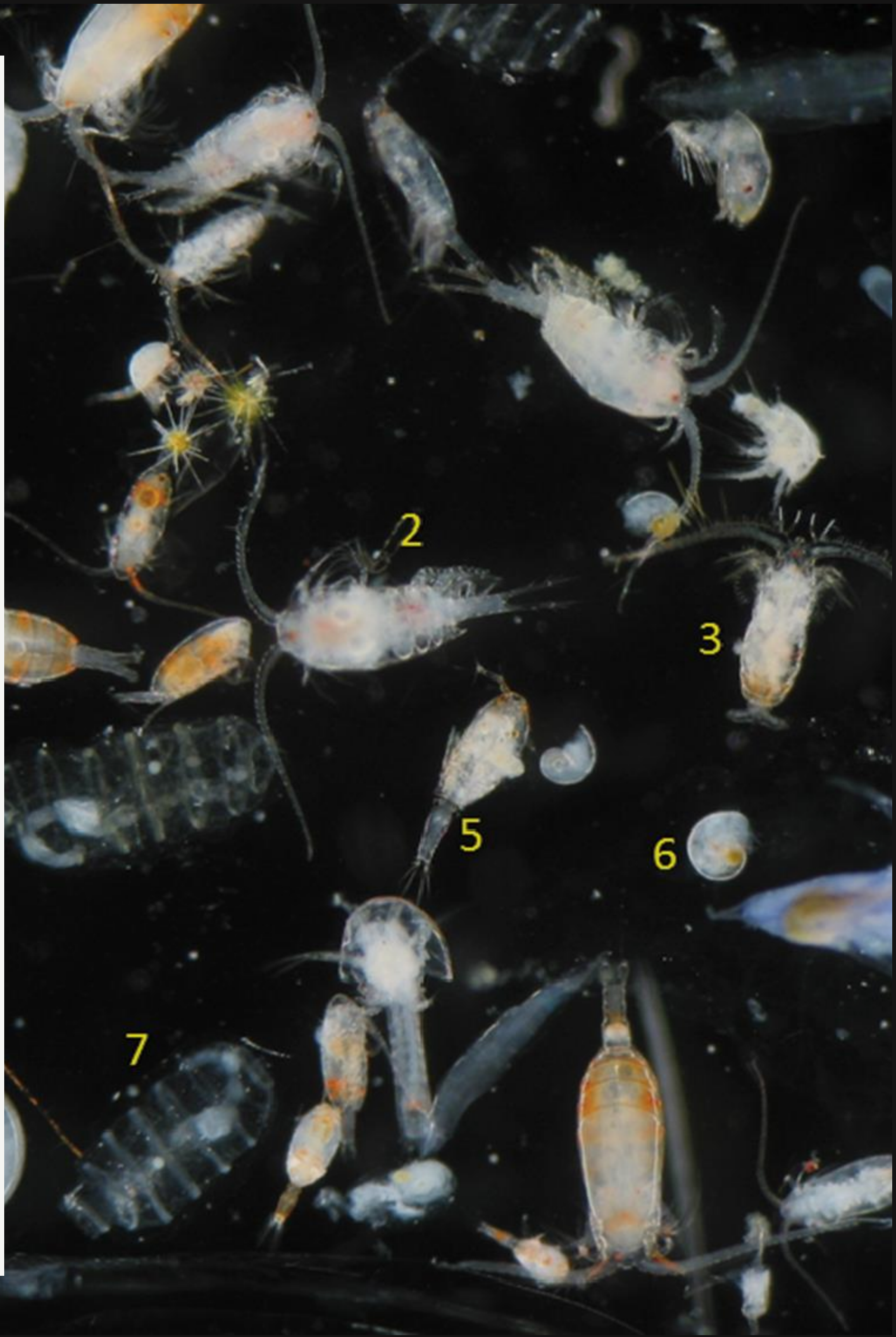
Passy 2007

Spatial Eigenfunctions



Spatial Eigenfunctions

Spatial eigenfunctions are mathematical constructs that capture spatial structure across different scales. They are used in spatial modeling to represent spatial autocorrelation patterns in data.



Spatial Eigenfunctions

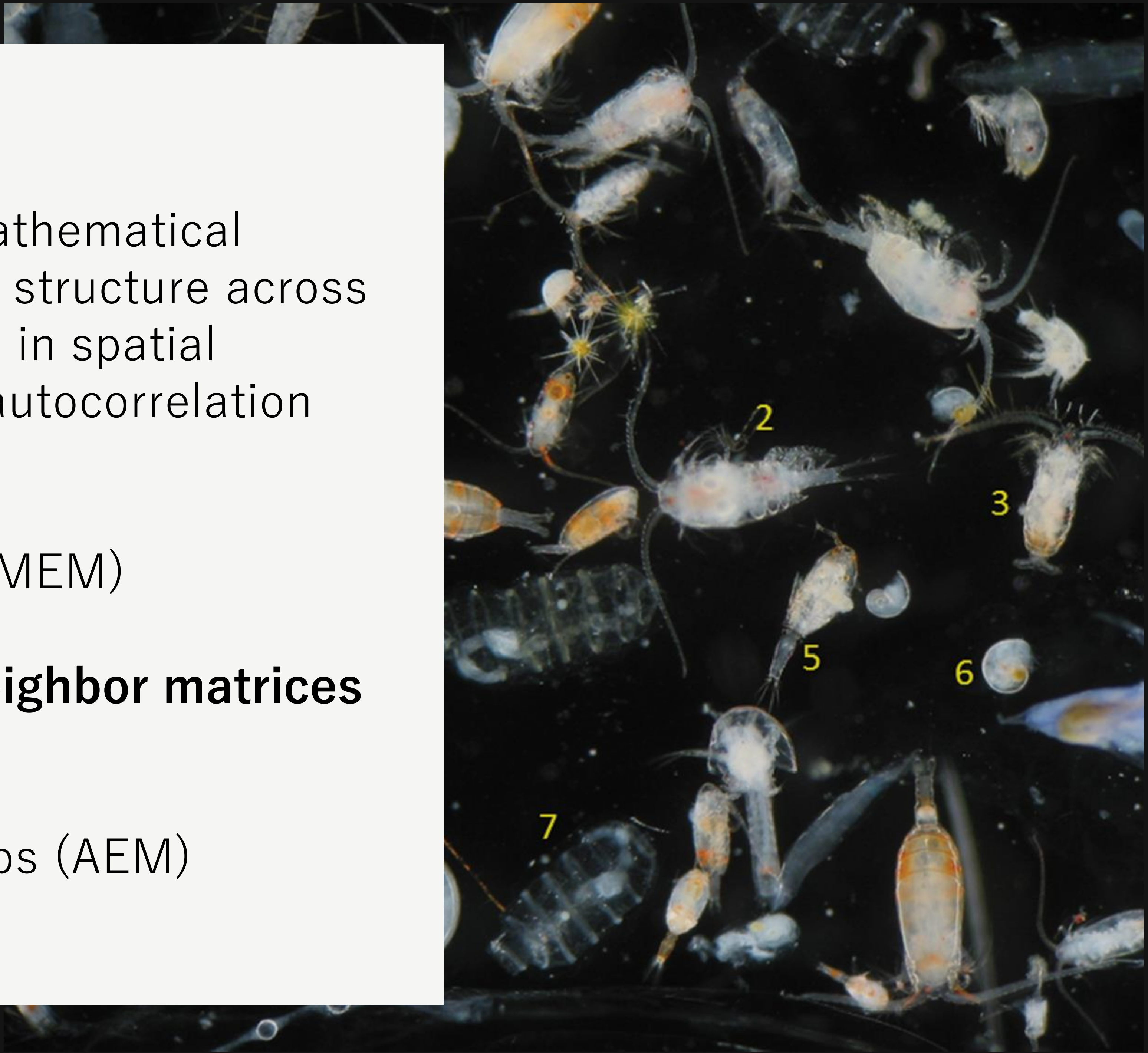
Spatial eigenfunctions are mathematical constructs that capture spatial structure across different scales. They are used in spatial modeling to represent spatial autocorrelation patterns in data.

- **Moran's eigenvector maps (MEM)**
- Principal coordinates of neighbor matrices (PCNM)
- Asymmetric eigenvector maps (AEM)

Spatial Eigenfunctions

Spatial eigenfunctions are mathematical constructs that capture spatial structure across different scales. They are used in spatial modeling to represent spatial autocorrelation patterns in data.

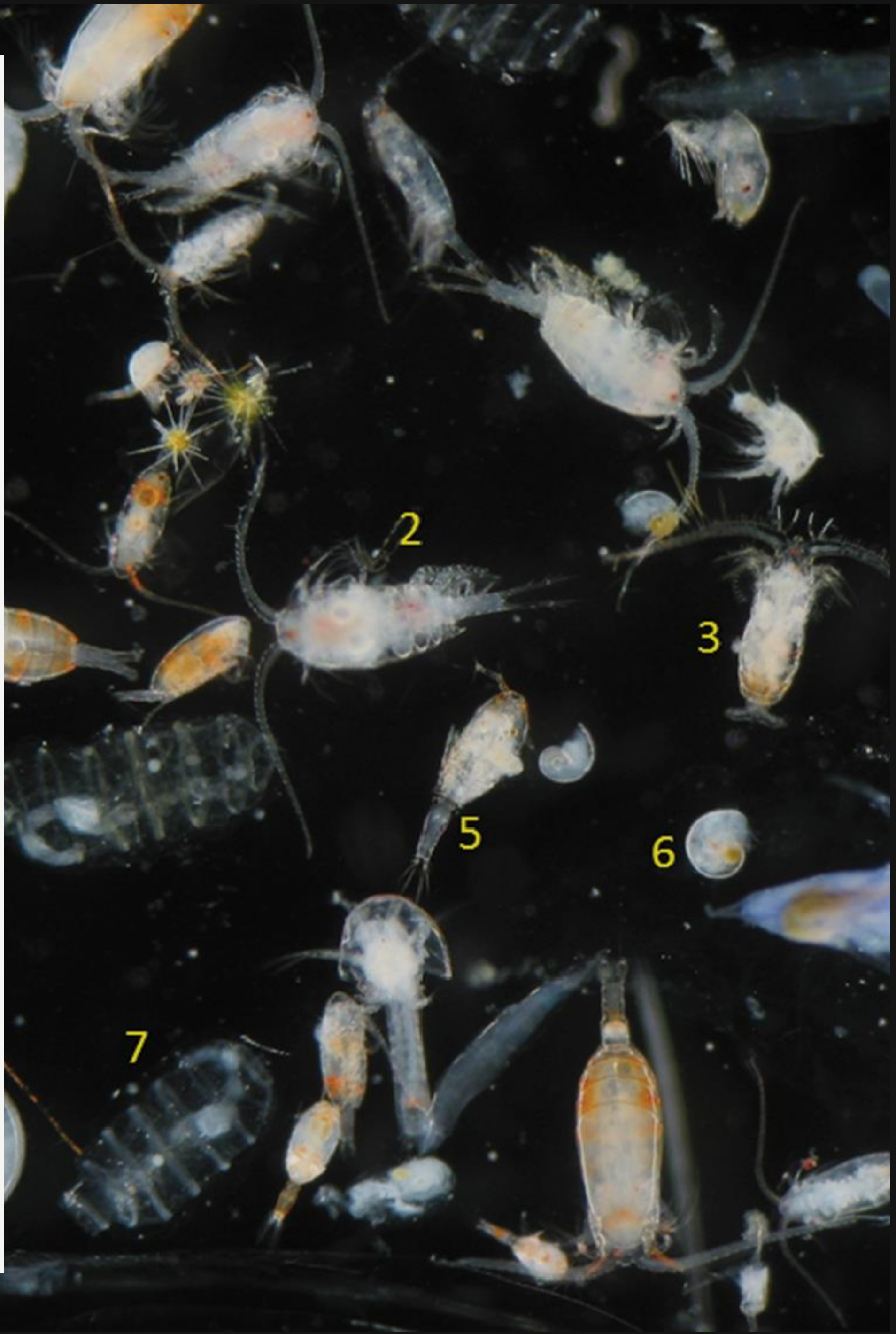
- Moran's eigenvector maps (MEM)
- **Principal coordinates of neighbor matrices (PCNM)**
- Asymmetric eigenvector maps (AEM)



Spatial Eigenfunctions

Spatial eigenfunctions are mathematical constructs that capture spatial structure across different scales. They are used in spatial modeling to represent spatial autocorrelation patterns in data.

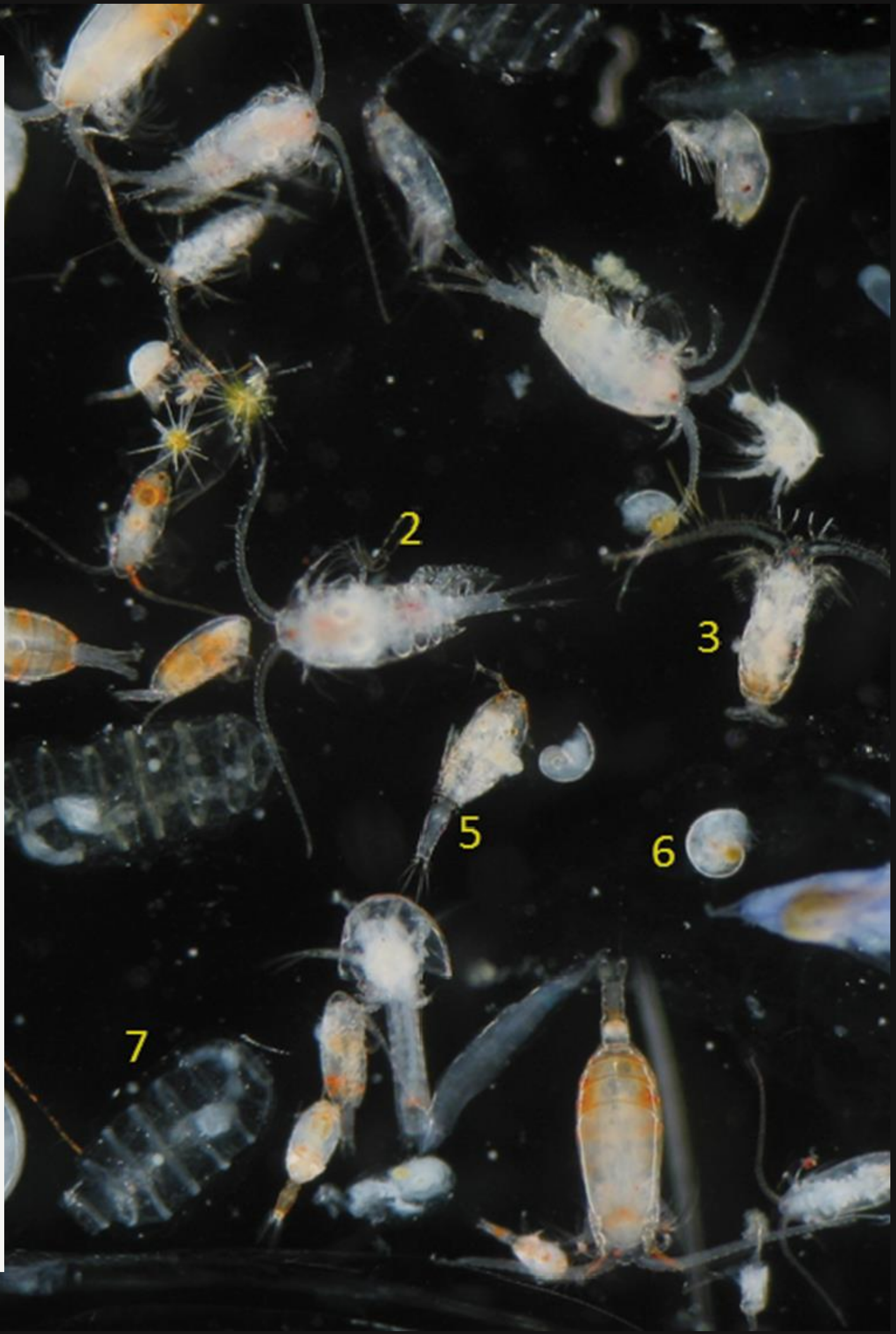
- Moran's eigenvector maps (MEM)
- Principal coordinates of neighbor matrices (PCNM)
- **Asymmetric eigenvector maps (AEM)**



Spatial Eigenfunctions

How are they used?

Spatial modeling: Eigenfunctions are included in regression or ordination models as explanatory variables to capture spatial patterns in response variables.

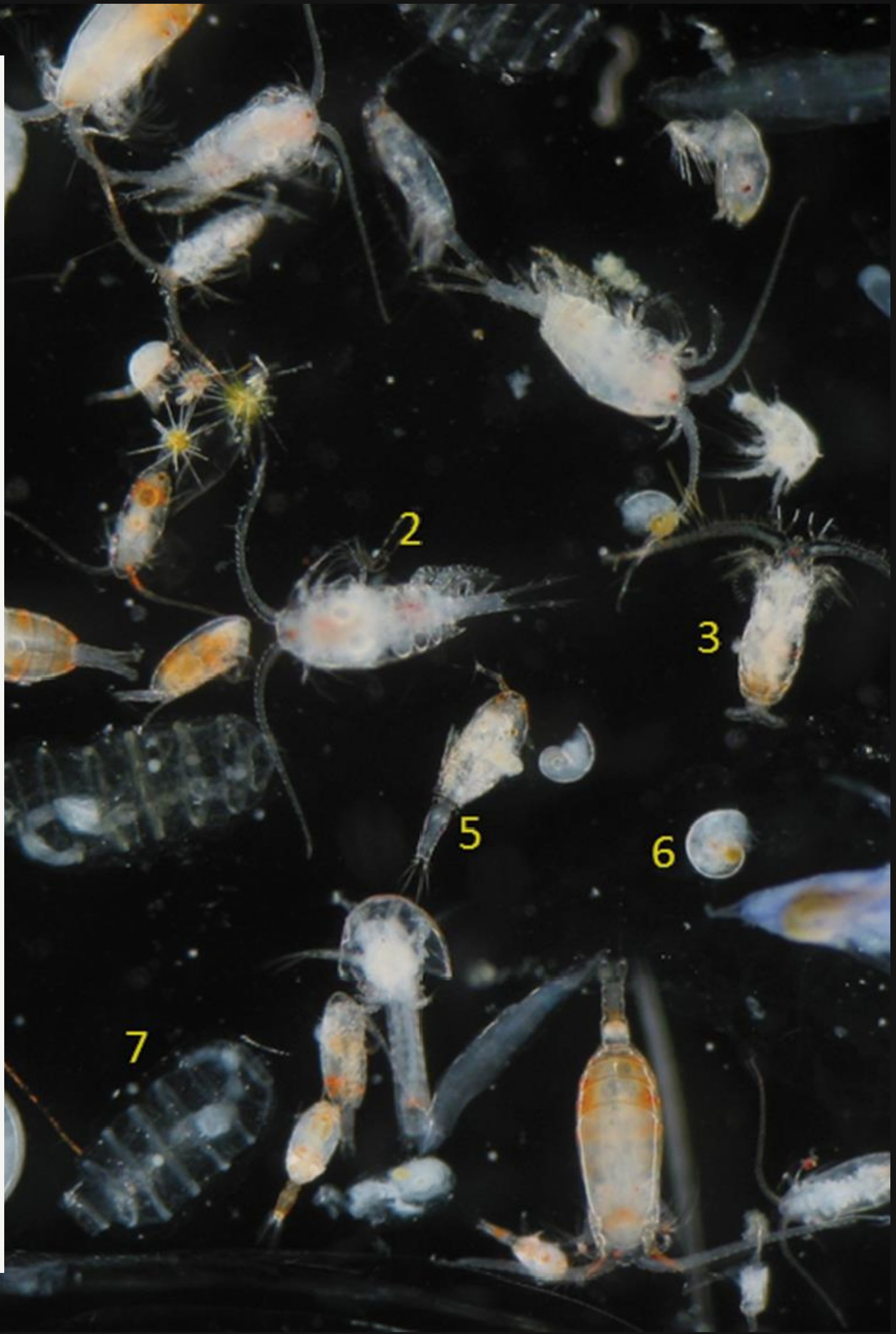


Spatial Eigenfunctions

How are they used?

Spatial modeling: Eigenfunctions are included in regression or ordination models as explanatory variables to capture spatial patterns in response variables.

Scale-Specific Analysis: By selecting specific eigenfunctions, we can focus on broad-scale or fine-scale spatial structures.



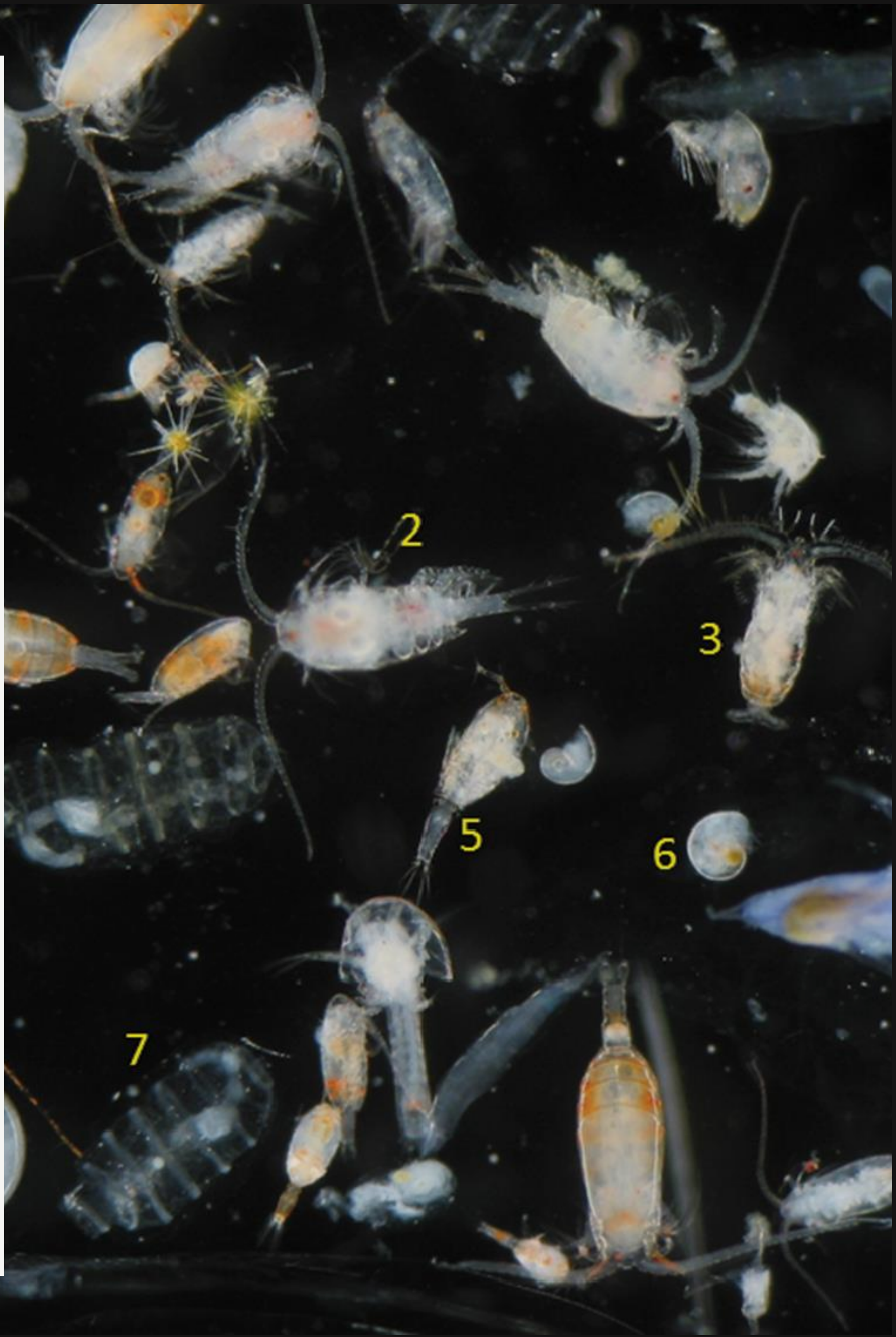
Spatial Eigenfunctions

How are they used?

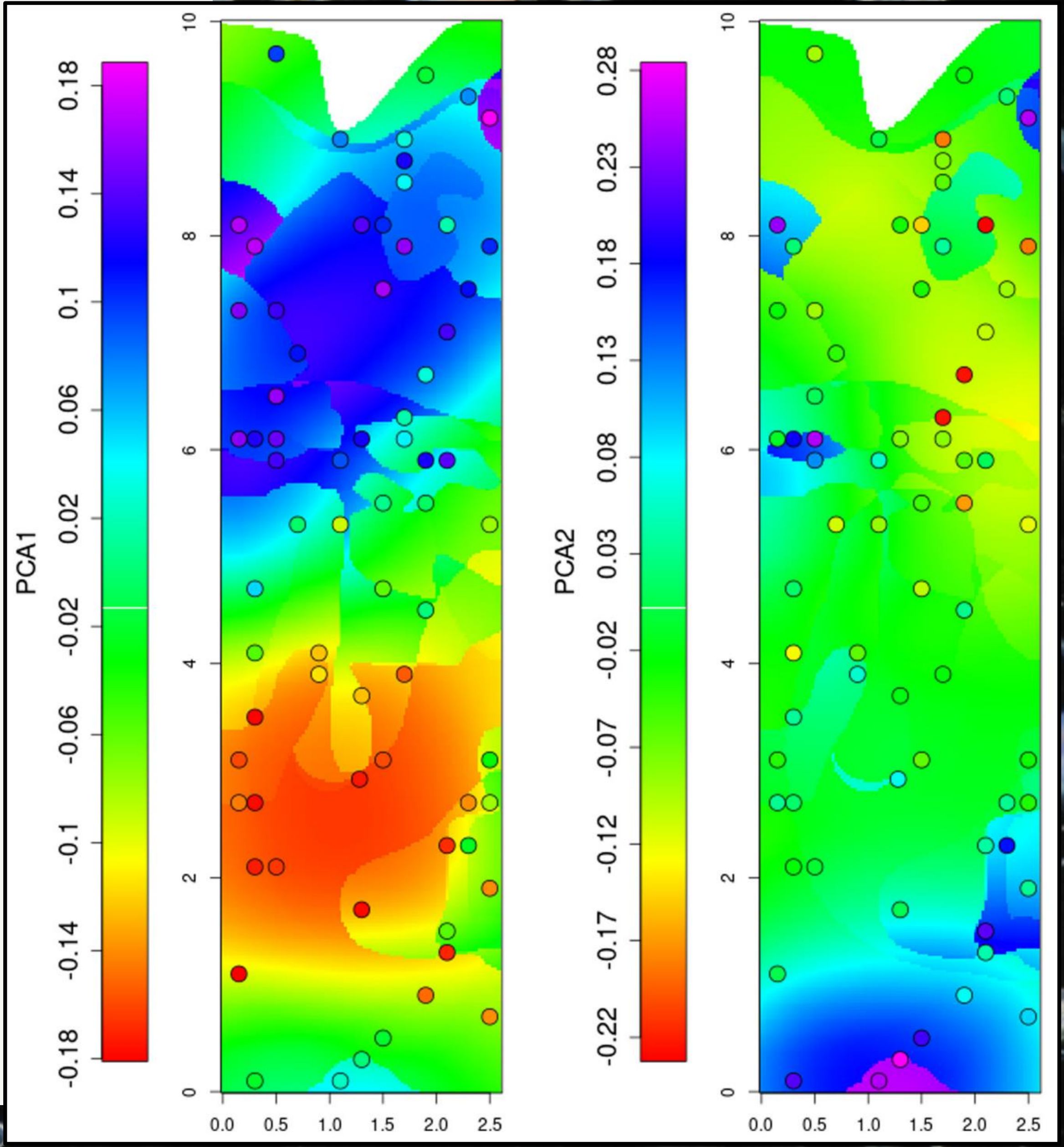
Spatial modeling: Eigenfunctions are included in regression or ordination models as explanatory variables to capture spatial patterns in response variables.

Scale-Specific Analysis: By selecting specific eigenfunctions, we can focus on broad-scale or fine-scale spatial structures.

Control for Spatial Autocorrelation: Spatial eigenfunctions account for spatial autocorrelation in the data to improve model accuracy



Spatial Eigenfunctions



Guenard & Legendre 2024

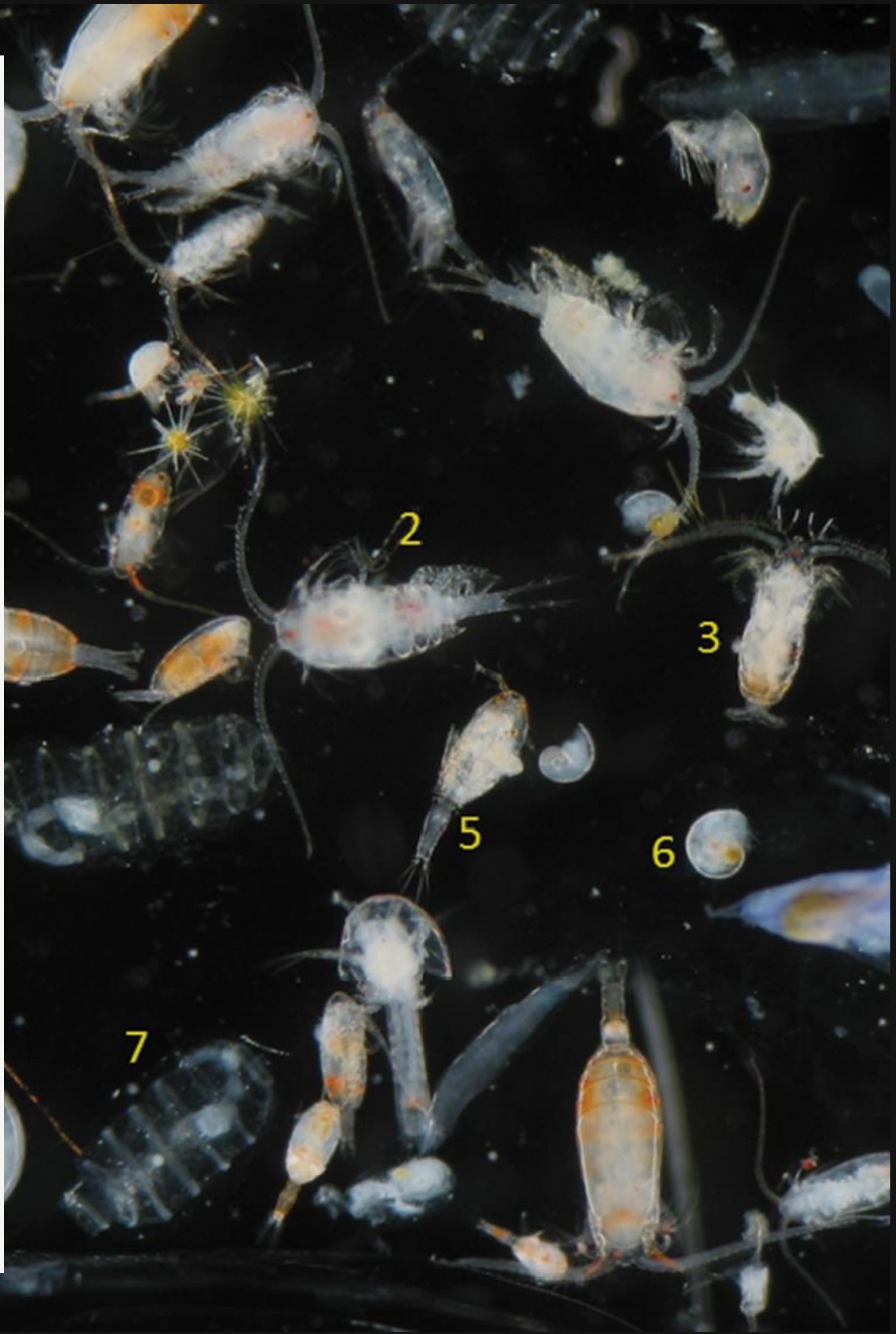
Conclusion: Summary of Key Points

- **Ecological data series** are continuous or discrete variables sampled over time or along transects in space.

- **Temporal autocorrelation** occurs when observations close in time are more similar to each other than observations farther apart.

- **Positive spatial autocorrelation** occurs when observations that are closer together tend to display values that are more similar than observations that are further apart

- **Autocorrelated values are not stochastically independent of one another!** For both parametric and non-parametric tests, the computed test statistics will exhibit Type 1 error (false significance) when autocorrelation is present.



Questions?