Monterey Bay Aquarium
Research Institute

Dynamical time-series analysis of midwater communities

Marisol Figueroa, Moss Landing Marine Labs: Physical Oceanography Lab

Mentors: Dr. Steven Haddock

Summer 2019

Keywords: Empirical Dynamic Modeling, Convergent Cross Mapping, Takens' Theorem, Time-Series, Lorenz attractors, 3D Modeling, Midwater Communities, Embedding, State-Space Reconstruction, Forecast

ABSTRACT

Most models used in research rely heavily on correlation, past observations of organisms/systems and try to predict the outcomes based on simple modeling techniques that are not optimal for complex scenarios. These models do not account for chaos or chaotic error in time series specifically and can miss spatial data that can be useful when research parameters that live within spatiotemporal variability. The ocean itself is a prime example of a chaotic system, therefore it is imperative when modeling how organisms and oceanographic conditions interact to utilize Empirical Dynamic Modeling (EDM). EDM is practical for studying systems that change over time, and the model itself infers patters and associations from data whereas other approaches are only making use of correlation-based inferences that can make invisible the real interactions of a system trying to be studied. These time series are projections of the manifold dynamics of the coordinate axes, and we can thus reconstruct this manifold utilizing time series data

collected over many years. Need more organisms and or more conditions. Keep it going over the years and see how the mathematical figure changes as more data is given so a full space reconstruction/ shadow can be created. Maybe include the El Nino year data however that might cause some bidirectional causality

INTRODUCTION

Most models used in research rely heavily on correlation, past observations of organisms/systems and try to predict the outcomes based on simple modeling techniques that are not optimal for complex scenarios. These models do not account for chaos or chaotic error in time series specifically and can miss spatial data that can be useful when research parameters that live within spatiotemporal variability. Even simple nonlinear processes are known to give rise to the phenomena "mirage correlations", where variables can appear to be correlated but this correlation may vanish or even change sign over different time periods. Eventually these models can suggest coupled variables are not casually related at all, and for variables such as oceanographic physical processes or biological processes of different organisms these are directly related and are driving forces for one another, as will be discussed later in this paper. The ocean itself is a prime example of a chaotic system, therefore it is imperative when modeling how organisms and oceanographic conditions interact to utilize Empirical Dynamic Modeling (EDM). EDM is practical for studying systems that change over time, and the model itself infers patters and associations from data whereas other approaches are only making use of correlation based inferences that can make invisible the real interactions of a system trying to be studied.

The basic underlying goal of EDM is to reconstruct behavior of dynamic systems from time series data, and the state of this dynamic system can be thought of as a location in a state space whose coordinates are on an axes x, y, and z [state what they are once decided]. The system state changing, evolving and the axes interacting create trajectories and trace out a physical 3-dimnesional geometric object called a Lorenz attractor. This manifold (*M)* consists of all trajectories on the Lorenz, where *t* is time and where *m(t)* is a point on the manifold, we may view a time series as a projection of that manifold onto a coordinate axis. As the point *m(t)* is being plotted in the 2d cartesian plane, the displacement of the time series on x is traced as well as y and z resulting in three simultaneous time series. These time series are projections of the manifold dynamics of the coordinate axes, and we can thus reconstruct this manifold utilizing time series data collected over many years. When conducting EDM, one must also talk about lags *X(t), X (t-T)* since these are essential. This paper will outline the materials and methods needed to conduct a full EDM analysis, results will vary when applied to visualize different parameters in 3D space on the manifold.

MATERIALS AND METHODS

Below is the code and packages that can be used to set up the EDM modeling. The data was pulled from Dr. Steven Haddock's data from the vars repo that inters have access to during the internship. All Nan's are filtered out of the data, set up of target depth, dive (period of year) and count is based off the organisms chosen for this model.

Generating a data set that has one entry per species per dive is important to integrate across depth and dive. Grouping and summarizing by commands is helpful. Filter by species when making that integrated (reduced) data set. If you do it on the whole thing, it can use that as the basis for choosing species or environmental pairs for comparison. For this example, it was grouped by Diversity and Concept and then count concepts (many entries per dive). And one to get the physical parameters around the 200-meter mark for each dive as explained earlier (**one entry per dive**). Merged them with a left.join into the table called together. Once these filters are set, the CCM data should knit in R when using R markdown. But if it does not knit, download the data onto your local and try again. Sometimes the vars data repo acts up.

```
#Copy paste into R and change what is necessary to fit your model

#Packages
library(tidyverse)
require(lubridate)
Sys.getenv("R_USER")
Sys.setenv(TZ='GMT')

# Start of code
dat = read_tsv("~/repos/vars-atlas/data/HaddockAfter2005.concepts",
na=c("NaN","None","NA","NAN"),col_types = cols())
TargetDepth=200
# (Sum concepts by dive)
onedim = dat %>%
  group_by(DiveStr,ConceptName) %>%
  summarise(counts = length(ConceptName))

# Get physical one value per dive
# Convert to pacific time now.
physical = dat %>% filter(TargetDepth-10 < Depth, Depth < TargetDepth+10) %>%
  group_by(DiveStr) %>%
  summarise(dates=first(with_tz(RecordedDate,"US/Pacific")),temp =
mean(Temperature,na.rm = TRUE),
                                        oxygen = mean(Oxygen,na.rm = TRUE), sal =
mean(Salinity,na.rm = TRUE)) %>%
  # Subtracting extra 0.3 days for ~8 hour offset
```

mutate(dayinterval=difftime(dates,first(dates),units="days"))

# merge two datasets above
together=left_join(onedim, physical, by = "DiveStr")

head(together)
tail(together)
view(together)

```
TargetDepth=200
# (Sum concepts by dive)
onedim = dat %>%
  group_by(DiveStr,ConceptName) %>%
  summarise(counts = length(ConceptName))

# Get physical one value per dive
physical = dat %>% filter(TargetDepth-10 < Depth, Depth < TargetDepth+10) %>%
  group_by(DiveStr) %>%
  summarise(dates=first(RecordedDate),temp = mean(Temperature,na.rm = TRUE),
oxygen = mean(Oxygen,na.rm = TRUE), sal = mean(Salinity,na.rm = TRUE))

# merge two datasets above
together=left_join(onedim, physical, by = "DiveStr")
```

The mutate function adds new variables and keeps existing ones. dayinterval was made and it was rounded with difftime calculating differences between the dates. min(dates) gives the minimum in the data set of the dates in the table you create and removes missing values. Units are in days, and the zero means the data was made into integers, it does not start at zero. Variable physical as the data being filtered by the Target depth (200-25) which is less depth (raw value from table), then raw value depth again less than Target depth plus 50. This created a range of values. Sidenote, for this data ser there were anomalies in Mexico within the data and I filtered it out by setting up the lat/long for the Monterey Bay Submarine Canyon (MBSC).

When using this code, summarize the function for mean to makes one column per group, defined variable dates as first date in time data (made into US/Pacific) with temperature being the mean of all the temperatures and remove missing values. Oxygen is the mean of

oxygen in data, remove missing values. Salinity in the code is defined as mean of salinity and the missing values were removed. Nextone is defined as a time series lag and or it shifts the time base of the dayinterval that was created above and the dataframe is sorted by dates. Defined interval as an integer with the time lag subtracted from the dayinterval that was created in the code. The select function is selecting the lag and all rows that have any missing values. When you are done filtering, you should have graphs and time series that show lags and correlations as a function of time and space like below.

(a) Moran effect in adult-recruitment model

$$R_i(t+1) = r_i N_i(t)[1 - N_i(t)]e^{-\Psi_i V(t)}$$
$$N_i(t+1) = s_i N_i(t) + \max[R_i(t-D_i), 0]$$

(b) Linear correlation $r_r = 0.557$

(c) Mirage correlation in two species competition model

$$M_1(t+1) = 3.8 M_1(t)[1 - M_1(t)] - 0.02 M_1(t) M_2(t)$$
$$M_2(t+1) = 3.5 M_2(t)[1 - M_2(t)] - 0.08 M_2(t) M_1(t)$$

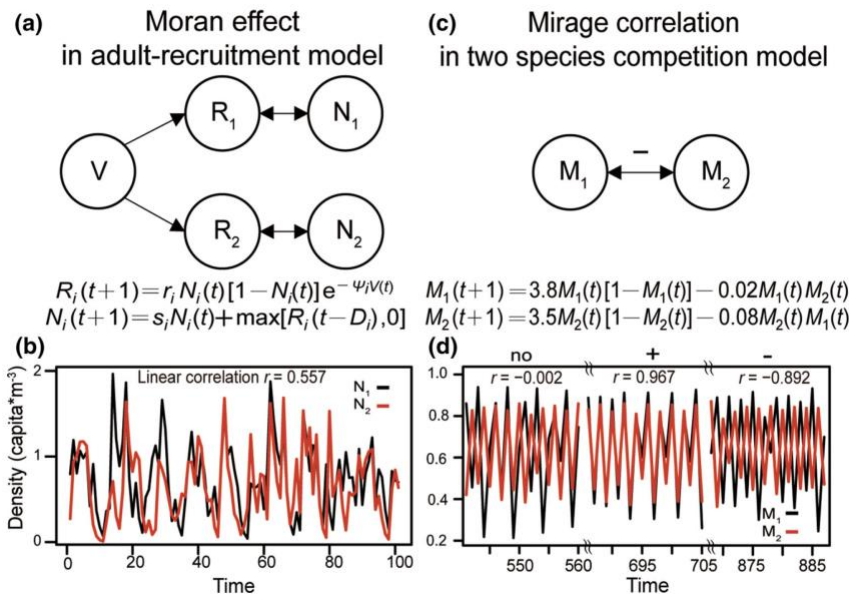**(d)** no $r = -0.002$ + $r = 0.967$ − $r = -0.892$

Figure 1: Physical representation of recruitment model and mathematical time lags and an example of the mirage correlations seen in time series with competition models. Paper cited below

```r
library(tidyverse)

require(lubridate)

Sys.getenv("R_USER")

Sys.setenv(TZ='GMT')

dat=read_tsv("~/repos/vars-atlas/data/HaddockAfter2005.concepts",
na=c("NaN","None","NA","NAN"),col_types = cols())

TargetDepth=200 #Can change this from 200 to 100 vice versa because NA's showed up.
No CTD readings at 100.

# (Sum concepts by dive)

onedim = dat %>%

  group_by(DiveStr,ConceptName) %>%

  summarise(counts = length(ConceptName))

# Get physical one value per dive

# Convert to pacific time now.

#interv =

physical = dat %>% filter(TargetDepth-25 < Depth, Depth < TargetDepth+50) %>%

  group_by(DiveStr) %>% #This range of values was then grouped by the dive

  summarise(dates=first(with_tz(RecordedDate,"US/Pacific")),temp                        =
mean(Temperature,na.rm = TRUE),

  oxygen = mean(Oxygen,na.rm = TRUE), sal = mean(Salinity,na.rm = TRUE)) %>%

   # Subtracting extra 0.3 days for ~8 hour offset

mutate(dayinterval=round(difftime(dates,min(dates,na.rm=TRUE),units="days"),0),

 nextone=lag(dayinterval, order_by = dates)

interv=as.integer(dayinterval-nextone)) %>%
```

select(-c("nextone")) %>% drop_na()

#latfilterX=dat%>%filter(Latitude-something<Latitude, Latitude<Latitude+something)

#longfilterX=dat%>%filter(Longitude-something<Longitude,
Longitude<LOngitude+something)

merge two datasets above

together=left_join(onedim, physical, by = "DiveStr") # variable together defined as a left
join (a,b, by="x1"), joining rows from b to a.

head(physical)

test = filter(together,ConceptName %in% c("Praya","Praya dubia")) %>% drop_na()
#Just seeing if the data works 😊

head(test)

tail(test)

view(dat)

Plot interval between dives

ggplot(physical,aes(x=dayinterval,y=interv)) + geom_col() + geom_point() #Must take
out Mexico

ggplot(physical,aes(x=dayinterval,y=temp)) + geom_line() + geom_point() #Must take
out Mexico

RESULTS

Results will vary based on where *t* is time and where *m(t)* is on what point on the
manifold. It will also vary on the time series used as it will be a projection of that
manifold onto a coordinate axis of your choice. As the point *m(t)* is being plotted in the
2d cartesian plane, the displacement of the time series on x is traced as well as y and z

resulting in three simultaneous time series for your chosen x,y, and z as well as where the lags *X(t), X (t-T) exist on the manifold* since these are essential.
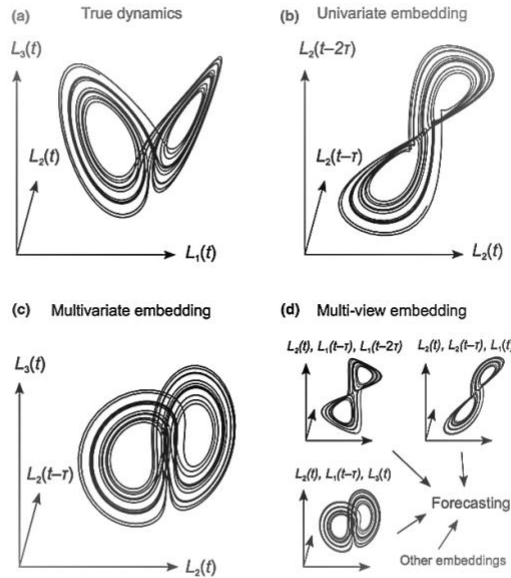

Fig. 4 Conceptual explanation of univariate, multivariate, and multi-view embedding, using the Lorenz attractor ($L_1(t)$, $L_2(t)$, and $L_3(t)$). The original attractor (a) is approximated using state space reconstruction by univariate (b), multivariate (c), and multi-view embedding (d). Under a univariate embedding (b), only one ($L_2(t)$, $L_2(t - \tau)$, and $L_2(t - 2\tau)$). In the multivariate embedding (c), multiple variables (L_2 and L_3 in this example) are used based on a combination of more than one lagged time series ($L_2(t)$, $L_2(t - \tau)$, and $L_3(t)$). Moreover, the attractor can be reconstructed by multi-view embedding (d), that is, combinations of various multiple

Figure 2: Full Lorenz examples. Paper cited in references.

DISCUSSION

The EDM approach worked very well this data from vars due to its ability to take time series data that was provided, and turn it into a 3D representation in the spatiotemporal space within R. In chaotic systems, linear modeling has a lot of limitations whereas EDM account form variability in chaotic, not only chaotic, but everchanging systems like the ocean. With this, we can determine the complexity of our system, distinguish linear versus nonlinear, determine casual variables, forecasting. Can provide effective policy and management decisions, recommendations about the survivorship of ecosystems, climate research etc.

CONCLUSIONS/RECOMMENDATIONS

Need more organisms and or more conditions. Suggestion: Continuous monitoring of three variables on temporal timescales to visualize. The mathematical figure will change as more data is given; with enough time series data a full space reconstruction and or Lorenz shadow can be created. Good Luck!!

ACKNOWLEDGEMENTS (Normal, Times New Roman, 12 pt, bold)

References:

Chang, CW., Ushio, M. & Hsieh, Ch. Empirical dynamic modeling for beginners. *Ecol Res* **32**, 785–796 (2017). https://doi.org/10.1007/s11284-017-1469-9

Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. Science 338:496–500. doi:10.1126/science.1227079

Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young LS (eds) Dynamic systems and turbulence. Springer, New York, pp 366–381 van Nes EH, Scheffer M, Brovkin V, Lenton TM, Ye H, Deyle E. doi: 10.1007/BFb0091924

Sugihara G (2015) Causal feedbacks in climate change. Nat Clim Change 5:445–448. doi:10.1038/nclimate2568

Whitney H (1936) Differentiable manifolds. Ann Math 37:645–680. doi:10.2307/1968482

Ye H, Sugihara G (2016) Information leverage in interconnected ecosystems: overcoming the curse of dimensionality. Science 353:922 doi: 10.1126/3