Université de Strasbourg

UFAZ FRENCH-AZERBAIJANI UNIVERSITY

Applications of Evolutionary Algorithms in climate projections

Master’s thesis defence

Kanan Jafarli

Supervisors : Prof. Rodrigo Abarca Del Rio, Prof. Pierre Collet and Dr. Ulviyya Abdulkarimova

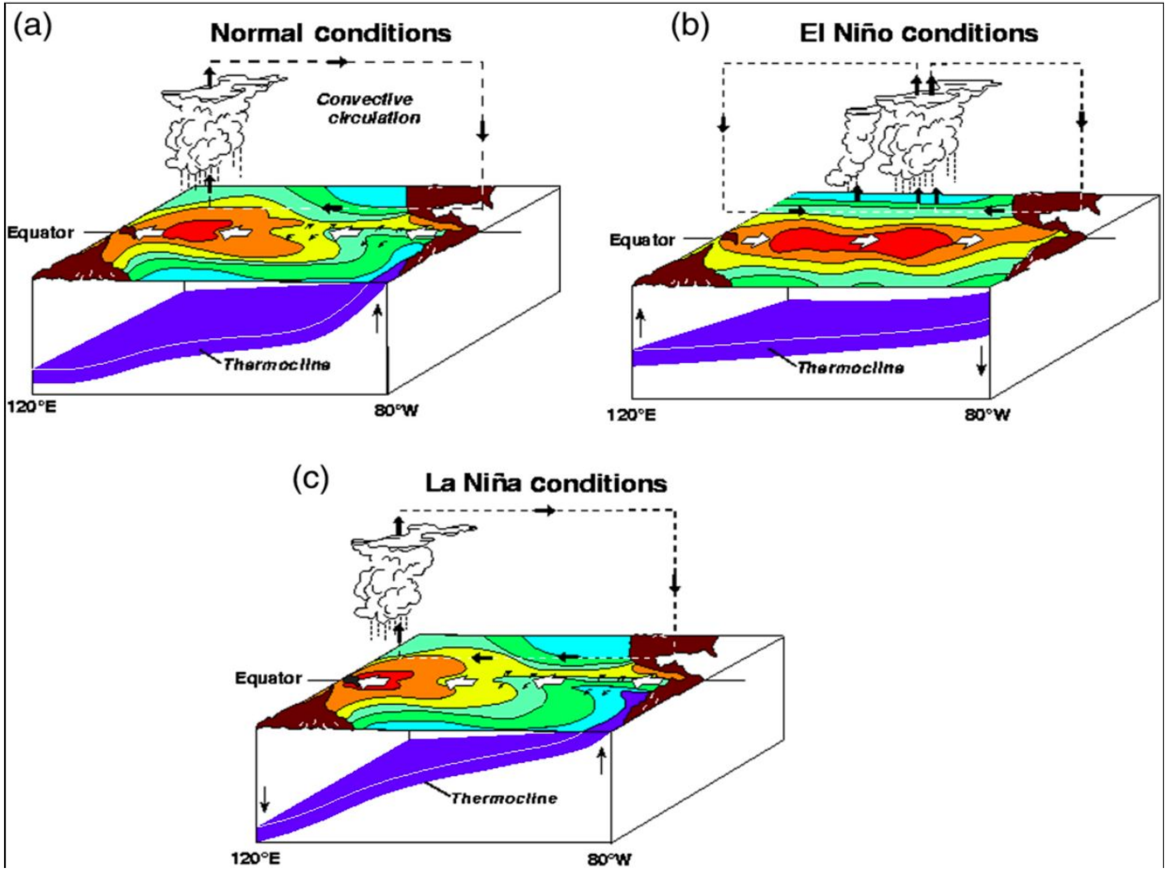
EL Nino Southern Oscillation (ENSO)

El Nino and La Nina

- El Nino is warm event of ENSO, means “Little Boy” or “Christ Child”.

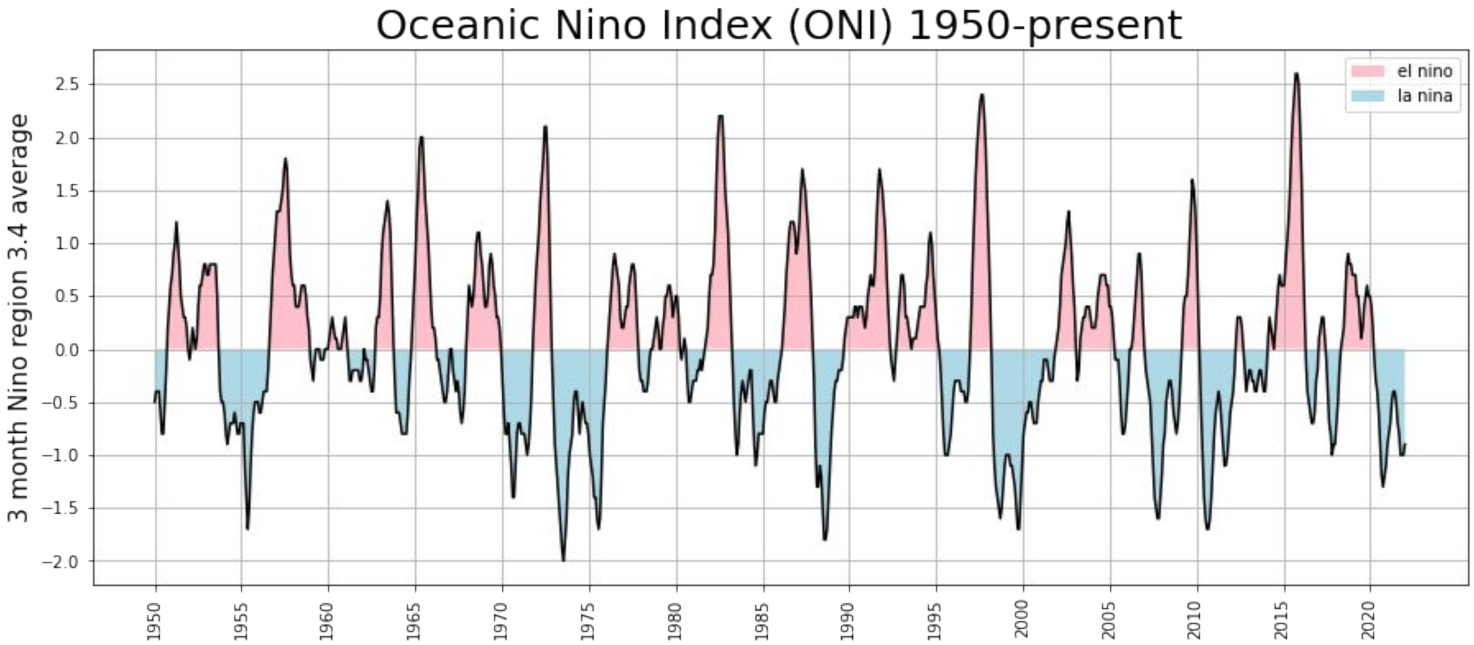
- La Nina is cold event of ENSO, means “Little Girl”.

Source: NASA; see www.pmel.noaa.gov/elnino/schematic-diagrams



Oceanic Nino Index (ONI) data

from Climate Prediction Center of the National Oceanic and Atmospheric Administration (NOAA).



Data source : https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

Category	Weak	Moderate	Strong	Very Strong
Threshold interval	[0.5 - 0.9]	[1.0 - 1.4]	[1.5 - 1.9]	[2.0 - ∞]

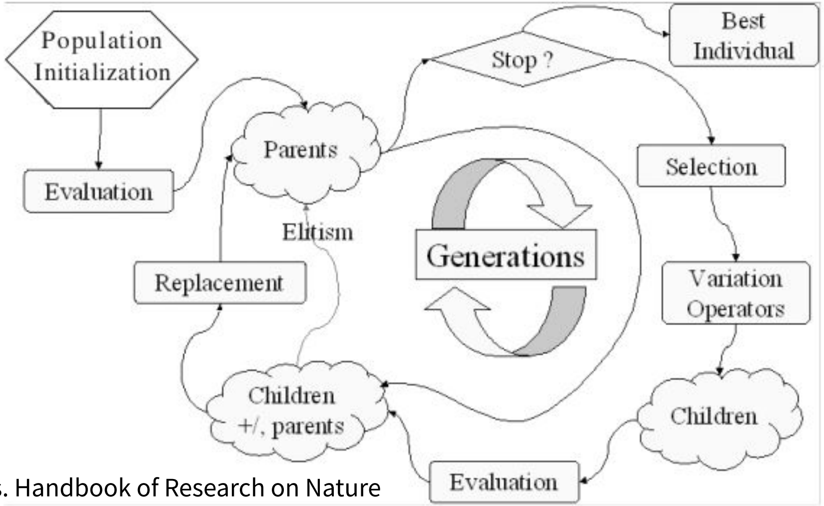
Objectives

- Classification of the strength of El Nino and La Nina events by amplitude (tested with ML classifiers)
- Classification of climate events El Nino, La Nina by shapes (by eye, by ML)
- Application of Evolutionary Algorithms to find simple equation to describe the events

Evolutionary Algorithms

- EA, inspired by Charles Darwin's theory of natural evolution and initially presented by Hans-Paul Schwefel amd Ingo Rechenberg in 1965 (Evolutionary Strategies) and then John Holland in the 1960's and early 1970's (Genetic Algorithms), are stochastic search and optimization algorithms based on the principles of natural evolution.

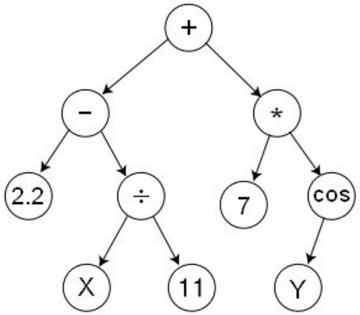
- Main parameters used in EA:
  - ❏ Population size and number of generations
  - ❏ Crossover and mutation probabilities
  - ❏ Number of children per generation
  - ❏ Selection pressure



P. Collet and J. P. Rennard. Stochastic optimization algorithms. Handbook of Research on Nature Inspired Computing for Economics and Management, 04 2007.

Genetic Programming

- GP is developed by Cramer in 1985 and developed by Koza in 1992.
- Each individual is a potential solution
- Potential solution is equation.
- Individuals are represented as LISP-like functions
- Used -, +, * operators and x, sin(x) functions

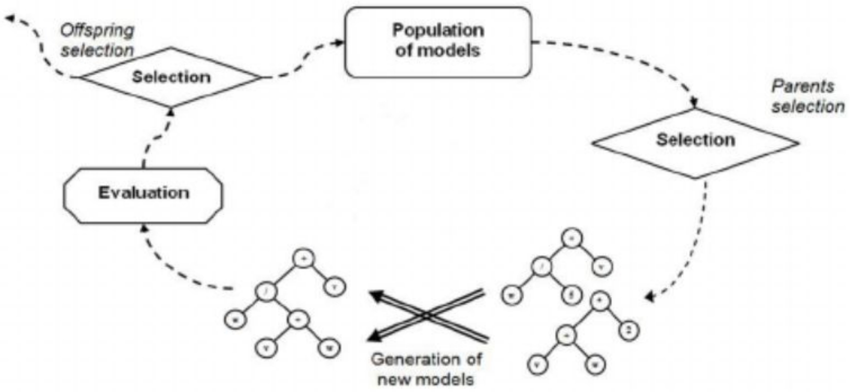


$$\left(2.2 - \left(\frac{X}{11}\right)\right) + \left(7 * \cos(Y)\right)$$

- Steps of GP :
  -initialization
  -evaluation

$$fitness = \sqrt{\sum_{i=1}^{n} \frac{|f(x_i) - s_i|^2}{n}}$$

  -selection (tournament)
  -crossover (single-point)
  -mutation



M. Kommenda, G. Kronberger, St. Winkler, M. Affenzeller, S. Wagner, L. Schickmair, and B. Lindner. Application of genetic programming on temper mill datasets. pages 58–62, 09 2009. doi: 10.1109/LINDI.2009.5258766.

Real valued genetic algorithm (RVGA)

- Stochastic algorithm that is used to find solutions to optimization problems
- Belongs to the family of EAs
- Each individual is a vector of real values corresponding to parameters of the equation
- Each parameter is represented by a IEEE754 single or double precision floating point number
- Potential solution is parameter values

- Steps of RVGA :
  -initialization
  -evaluation

$$fitness = \sqrt{\sum_{i=1}^{n} \frac{|f(x_i) - s_i|^2}{n}}$$

  -selection (tournament)
  -crossover (single point)
  -mutation

EASEA platform

- Evolutionary Algorithm Specification Language (EASEA) is an Artificial Evolution software platform.
- Created back in year 2000
- Automatically parallelizes EAs on parallel architectures
- Its compiler generates C++ source files.

- How to run GP :

  easena filename.ez -cuda_gp

  make

  ./filename

- How to run RVGA :

  easena filename.ez -cuda

  make

  ./filename

Parallelization of Evolutionary Algorithms

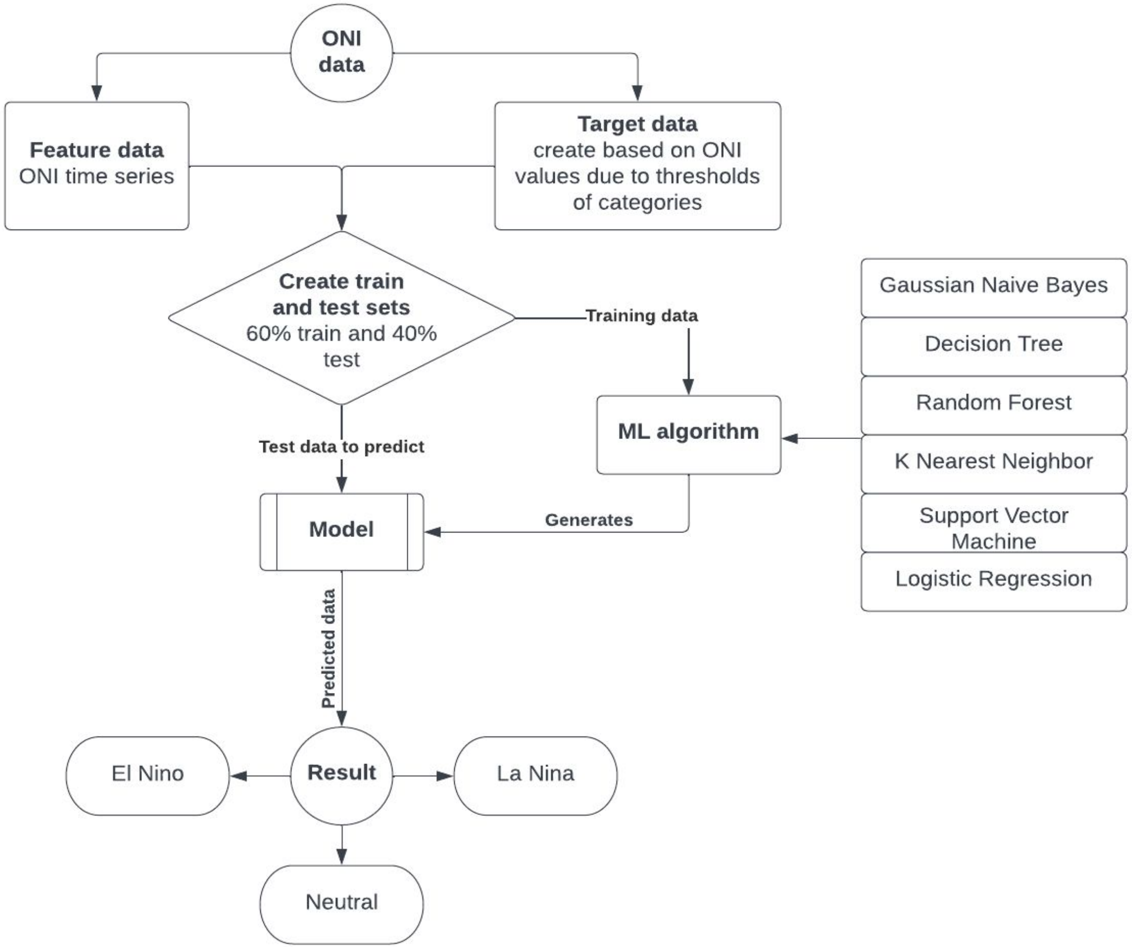
- EAs are inherently parallel.

- EAs can be parallelized using a single machine with multiple processor cores or a system of machines sharing the population over the network :
  - ➔ GPGPU parallelization
    - ➔ parallelization can be done automatically on NVIDIA GPGPU cards using EASEA platform by compiling the source file with the -cuda option
  - ➔ Island parallelization
    - ➔ Exchanging individuals/knowledge between different machines (islands)

Runs are done on using PARallel System for Evolutionary Computing (PARSEC) machines at the French-Azerbaijani University

Experimental results

ML Classification



Classification of different datasets with ML models

- 2 classes with El Nino and anything else

	accuracy	mean	std
Random Forest	1.000000	1.000000	0.000000
Logistic Regression	1.000000	1.000000	0.000000
Gaussian NB	0.962209	0.972032	0.017098
Decision Tree	1.000000	1.000000	0.000000
Support Vector Machine	1.000000	1.000000	0.000000
K Nearest Neighbor	1.000000	1.000000	0.000000

- 3 classes with El Nino, La Nina and regular year

	accuracy	mean	std
Random Forest	1.0	1.0	0.0
Logistic Regression	1.0	1.0	0.0
Gaussian NB	1.0	1.0	0.0
Decision Tree	1.0	1.0	0.0
Support Vector Machine	1.0	1.0	0.0
K Nearest Neighbor	1.0	1.0	0.0

- 5 classes with strong EN, strong LN, regular year, weak EN and weak LN

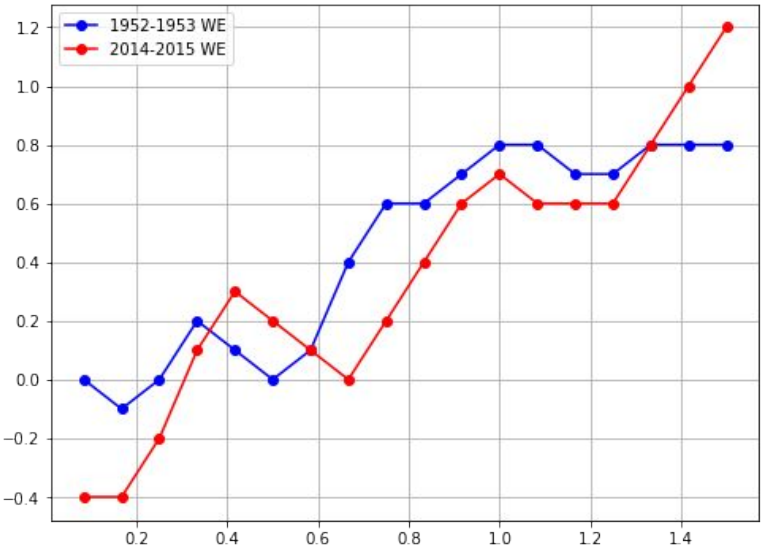
	accuracy	mean	std
Random Forest	1.000000	1.000000	0.000000
Logistic Regression	0.872093	0.884721	0.022366
Gaussian NB	1.000000	1.000000	0.000000
Decision Tree	1.000000	1.000000	0.000000
Support Vector Machine	0.979651	1.000000	0.000000
K Nearest Neighbor	1.000000	1.000000	0.000000

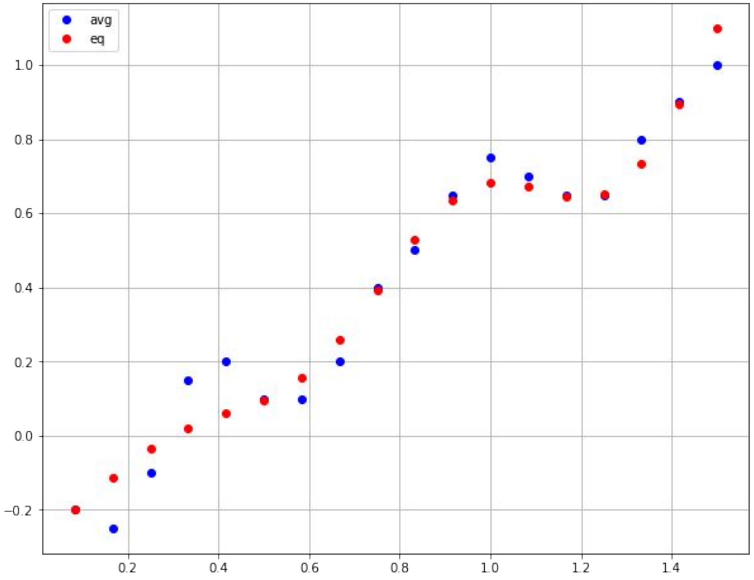
- 5 classes with weak EN, moderate EN, strong EN, very strong EN and anything else

	accuracy	mean	std
Random Forest	1.000000	1.000000	0.000000
Logistic Regression	0.857558	0.921977	0.021429
Gaussian NB	1.000000	1.000000	0.000000
Decision Tree	1.000000	1.000000	0.000000
Support Vector Machine	0.973837	1.000000	0.000000
K Nearest Neighbor	0.994186	0.997674	0.004651

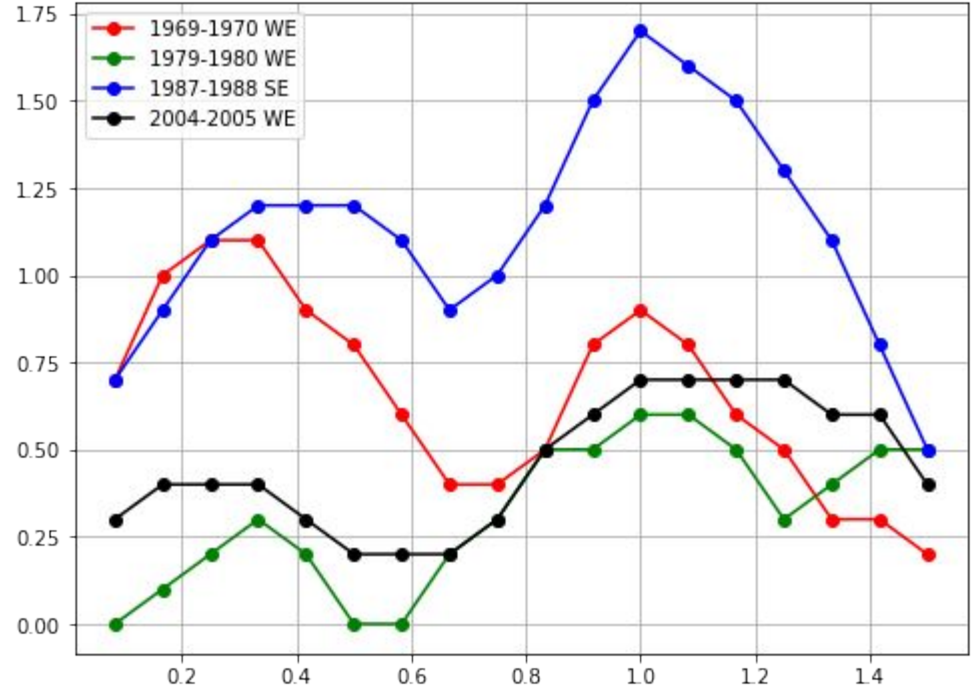
Classified El Nino events by eye and found equation of average in the each group using GP

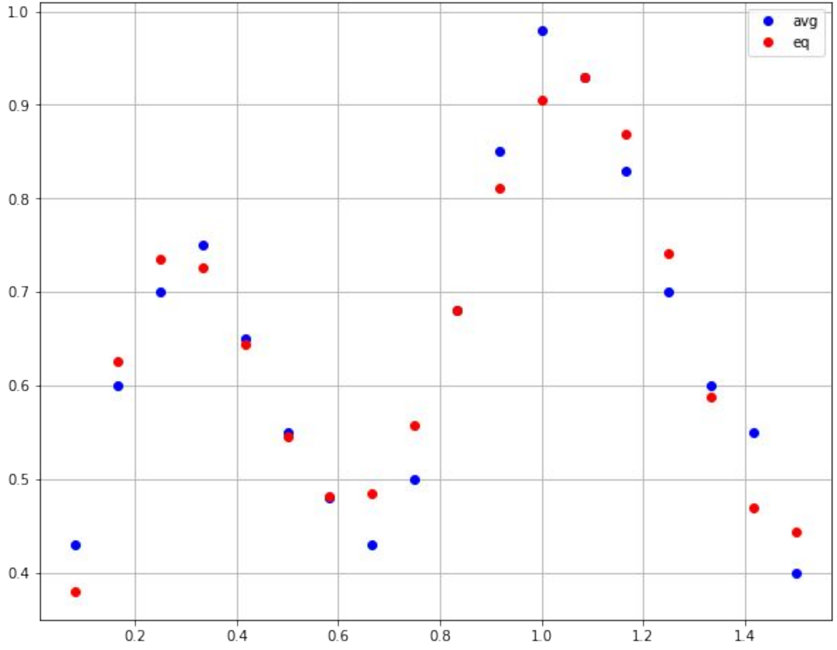
- eq1 = 0.128129 * x * sin(8.640880 * x) + 0.870872 * x − 0.2798906



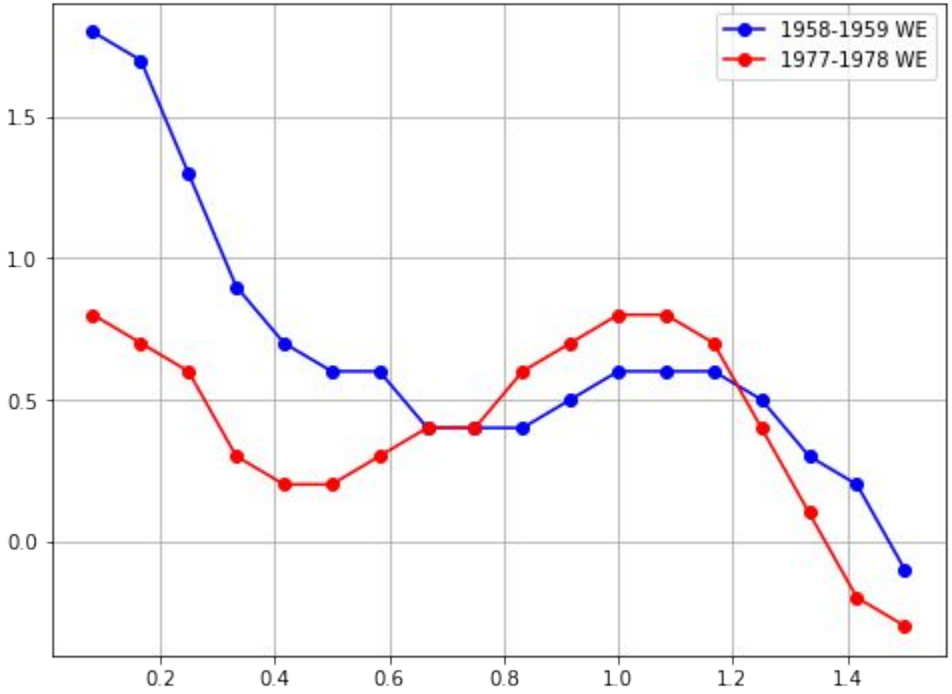


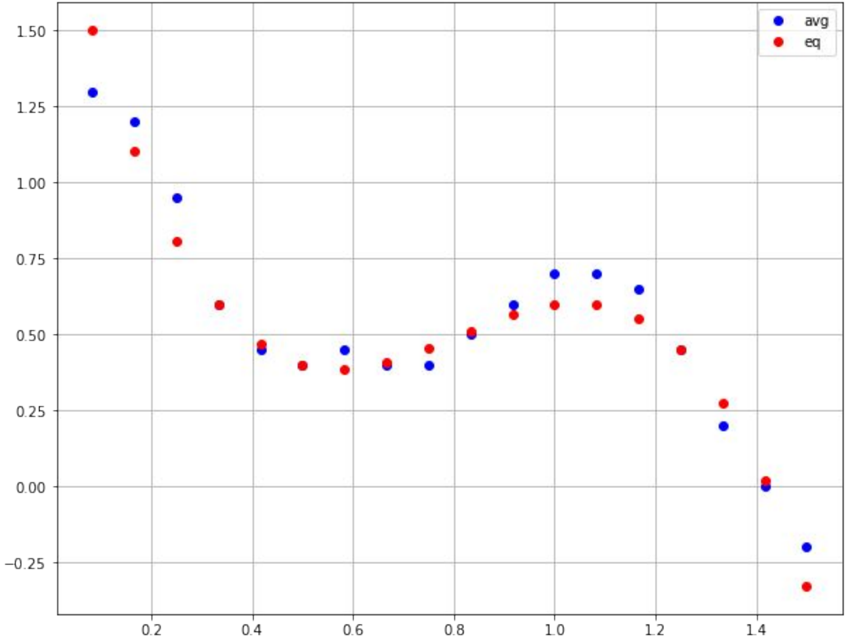
- $eq2 = 1.189035 * \sin(0.699276 * x) + (-0.920625 * x + 0.755611) * \sin(5.417646 * x) + 0.014571$



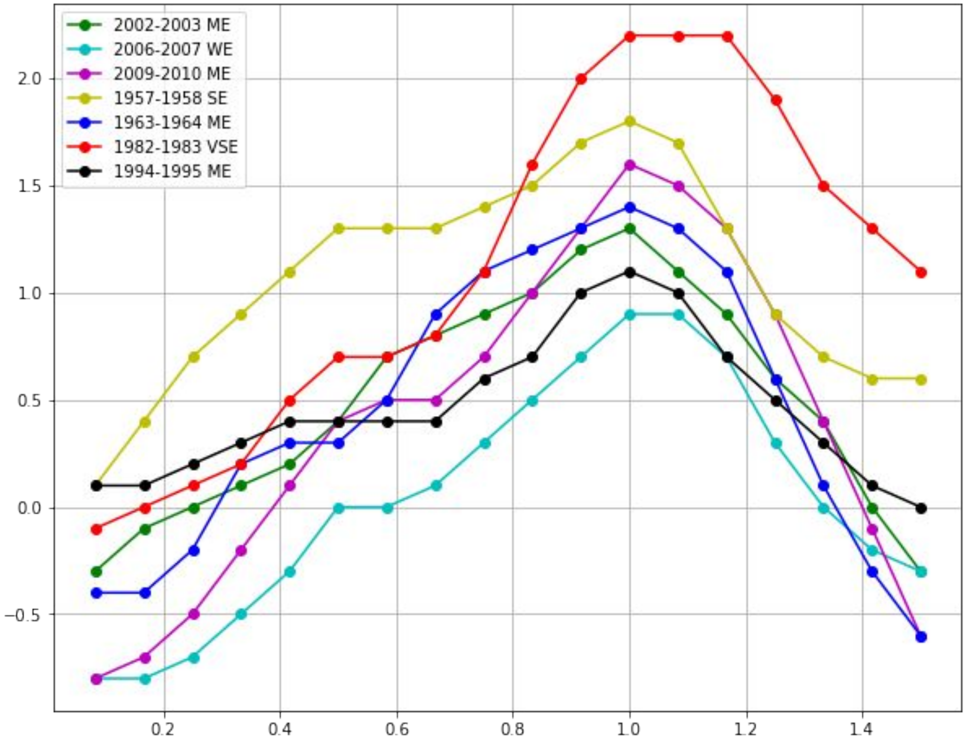


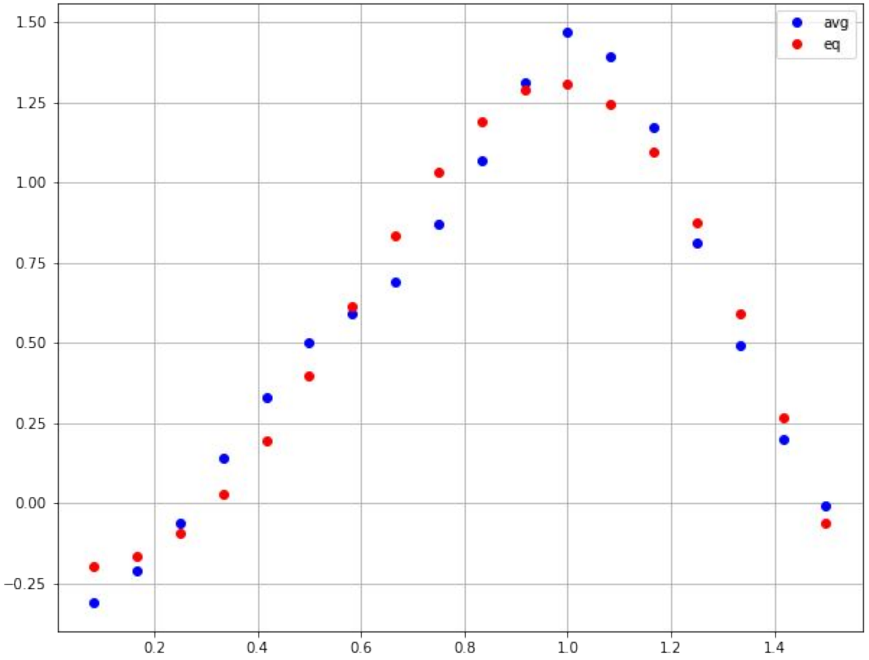
- $eq3 = (1.4221796 - x) * (2 * x * (x - 0.995554) + (2 * x - 0.64533426) * \sin(x - 0.694731) + 1.0)$



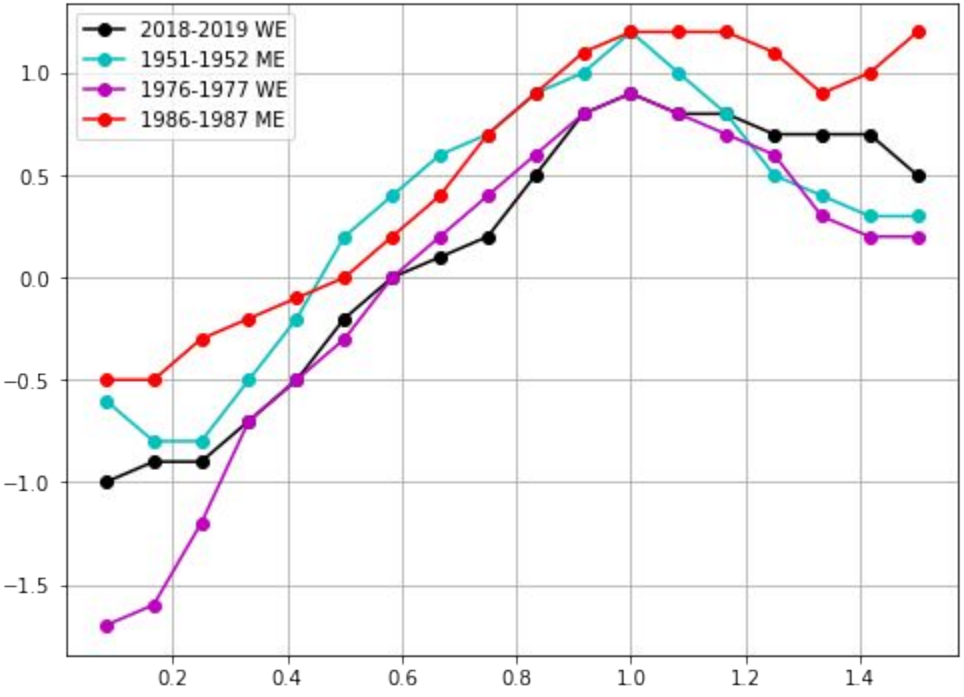


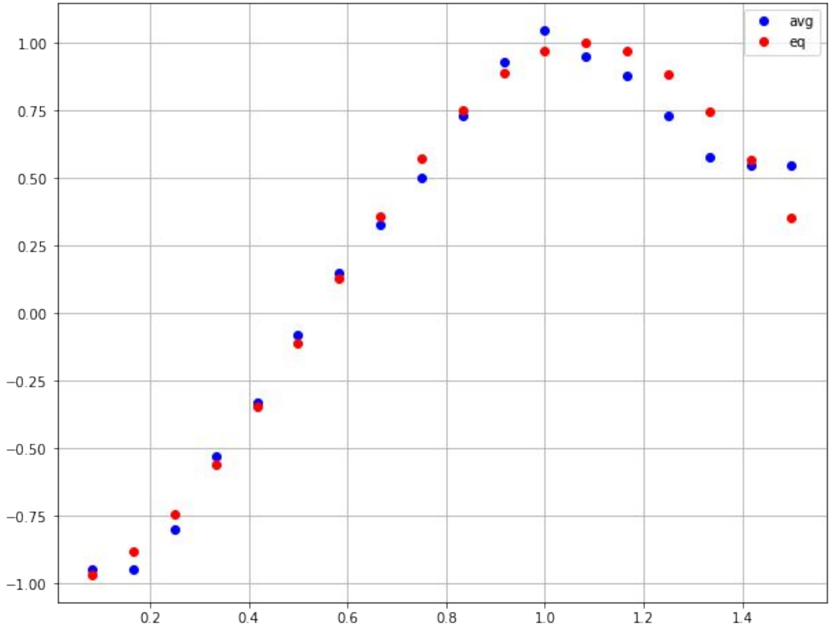
- eq4 = 0.660694 * x + x * sin(3.211015 * x − 1.065241) − 0.192678



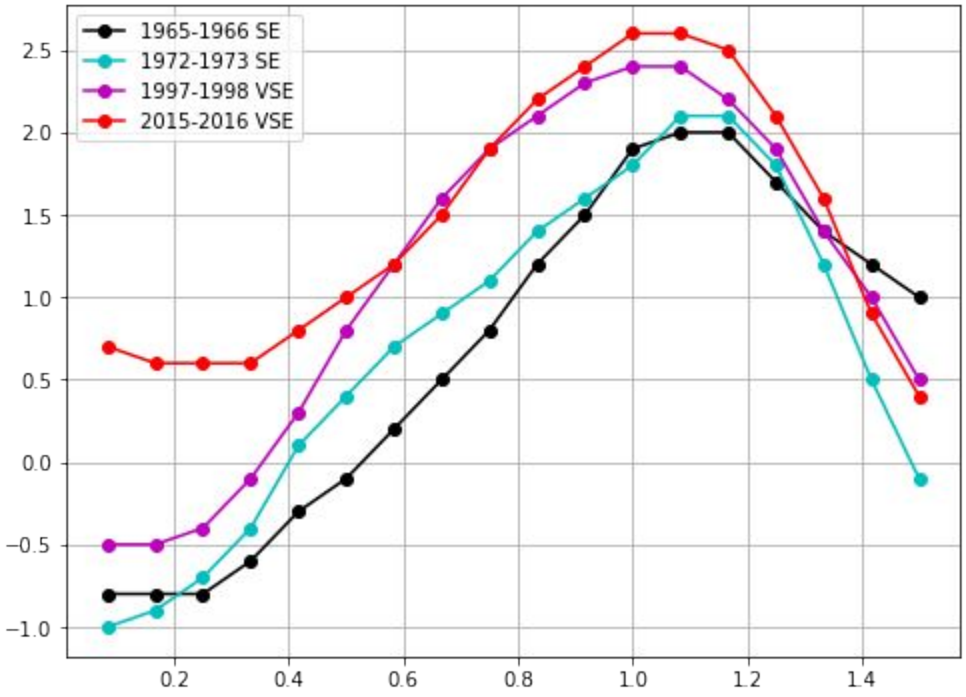


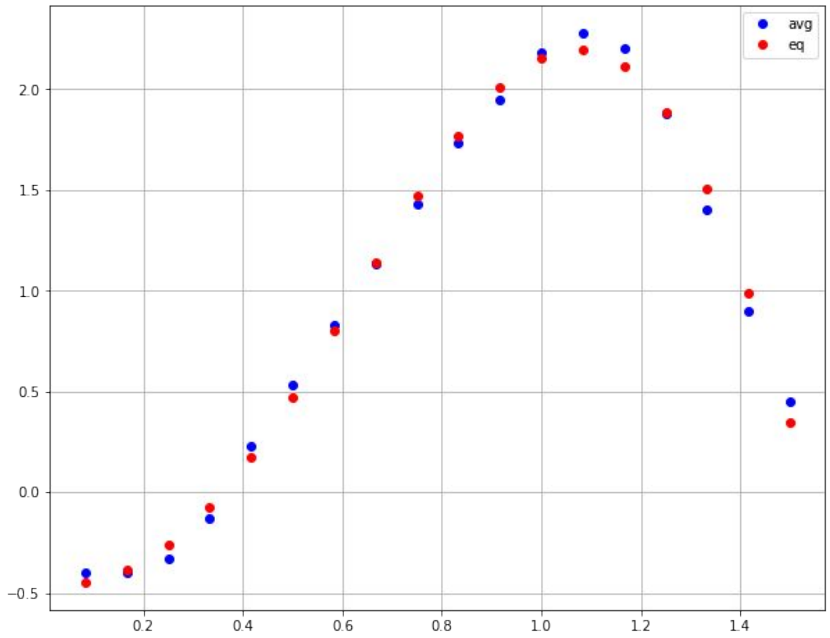
- eq5 = sin(2.892047 * x + 4.724232)



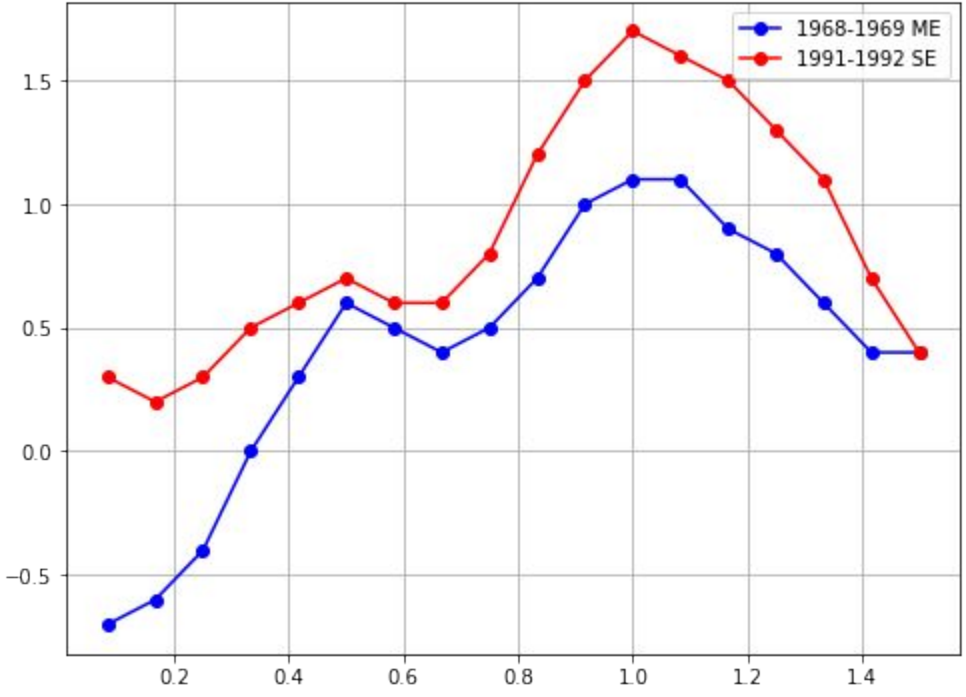


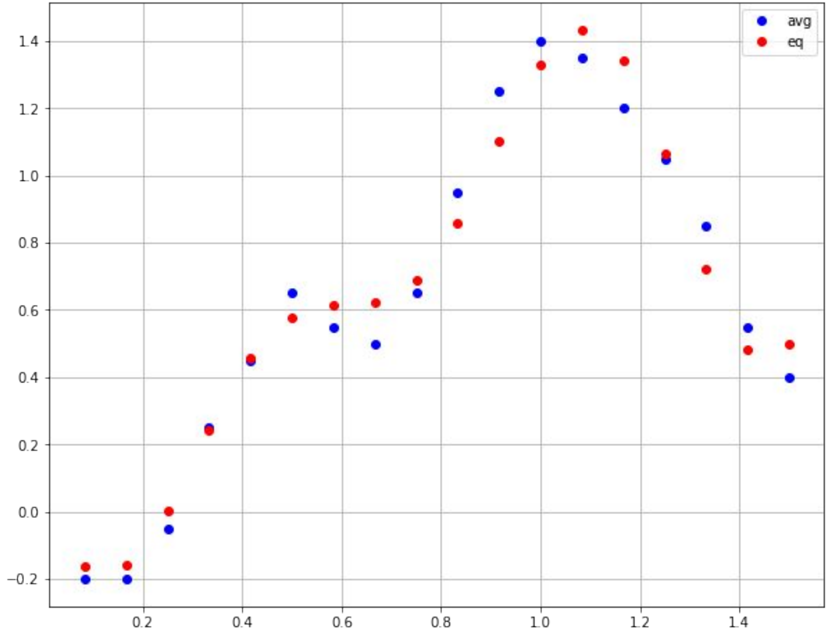
- $eq6 = x + (-0.497150 - 3.264897 * x) * \sin(2.908535 * x) + 5.213310 * x * \sin(2.648657 * x) - 0.441066$



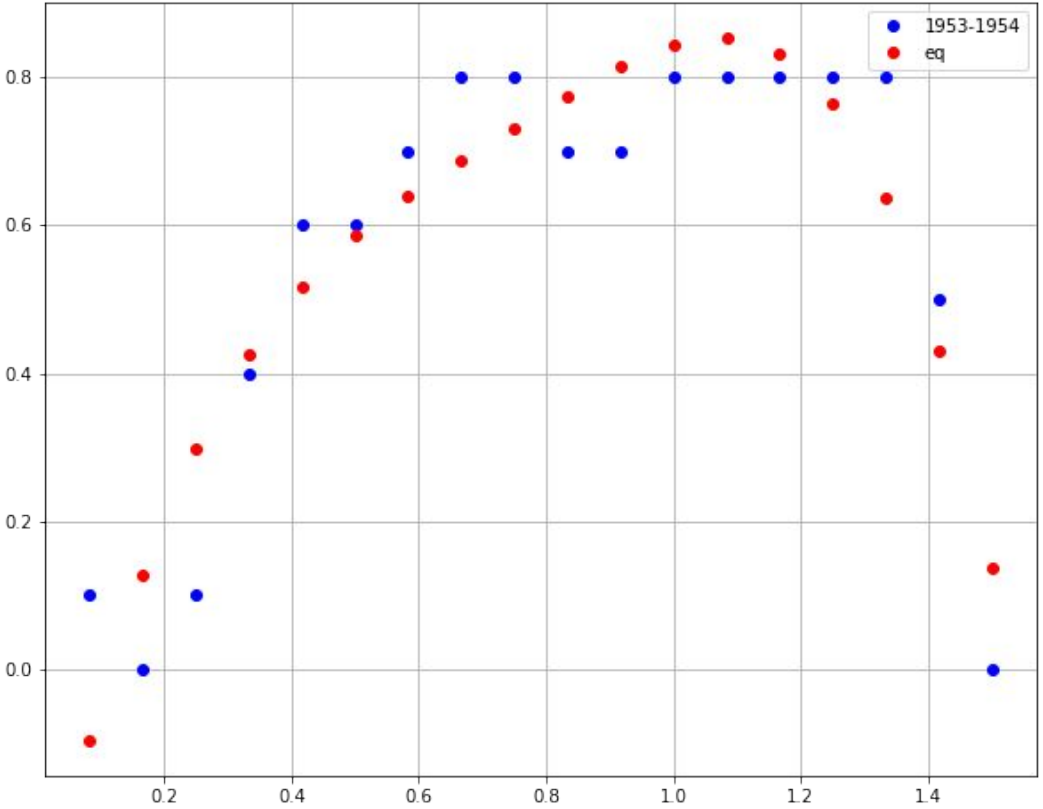


- $eq7 = 1.270450 * \sin(0.869869 * x)+(0.873001 * x-0.504249) * \sin(7.622553 * x)$

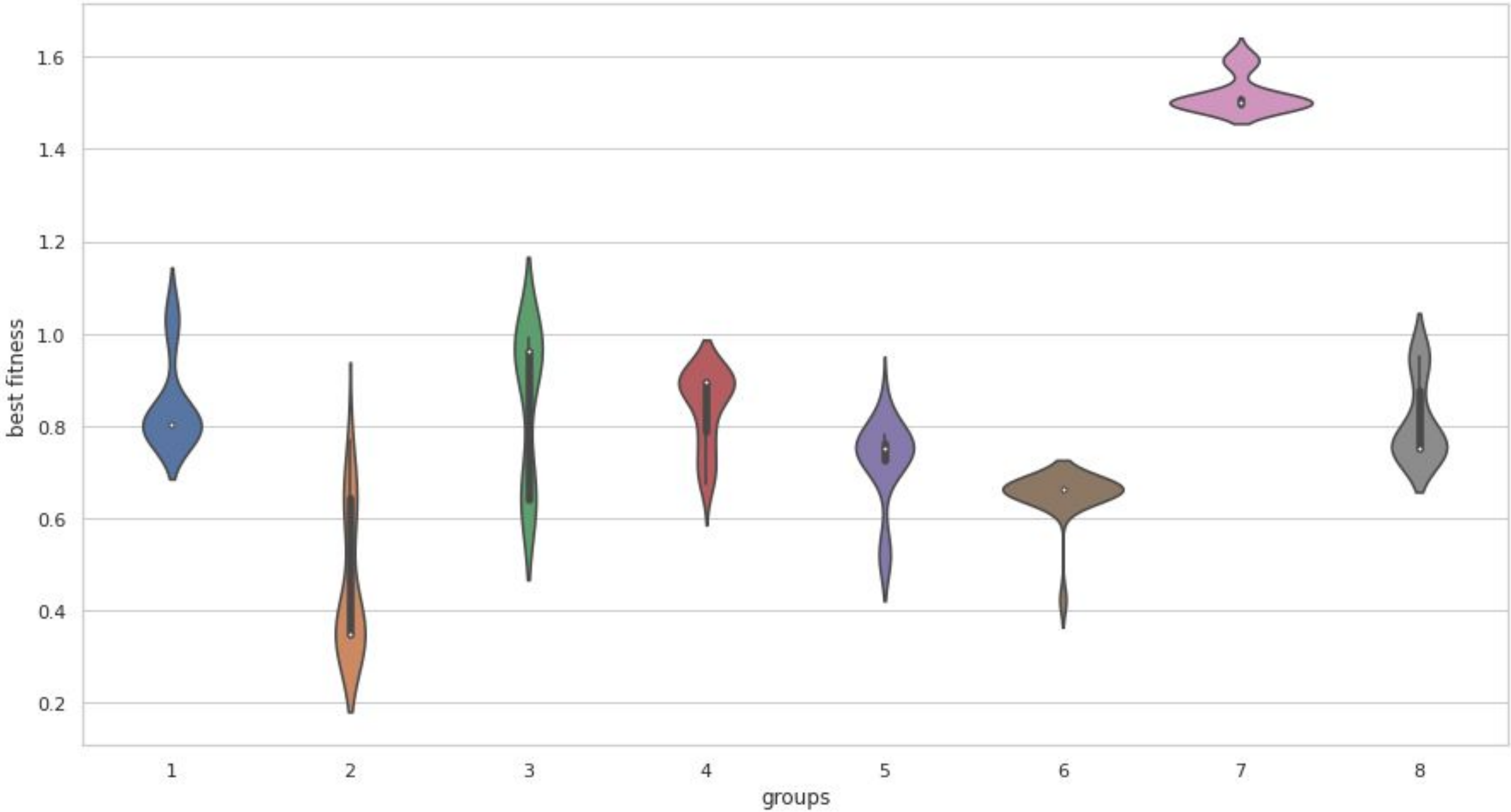




- eq8 = −0.513958 * x − (4.298025 * x − 0.378910) * sin(x − 1.612762) + (x − 2.151642) * x * sin(2.249024 * x)

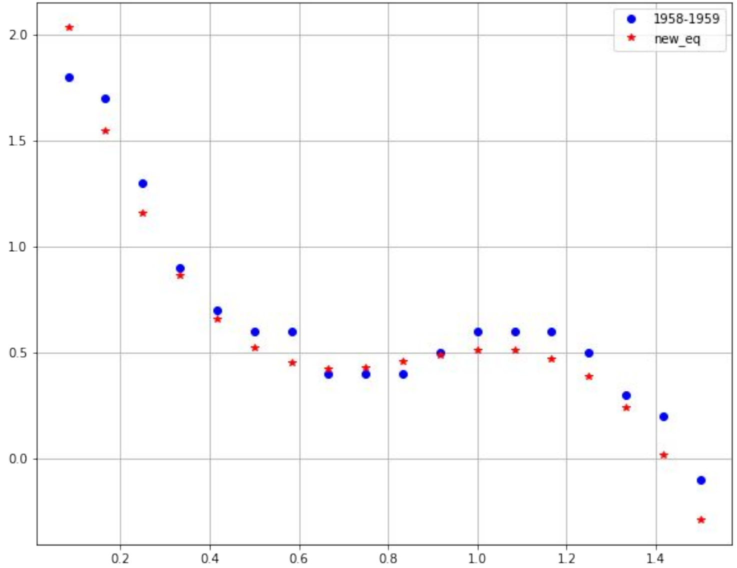


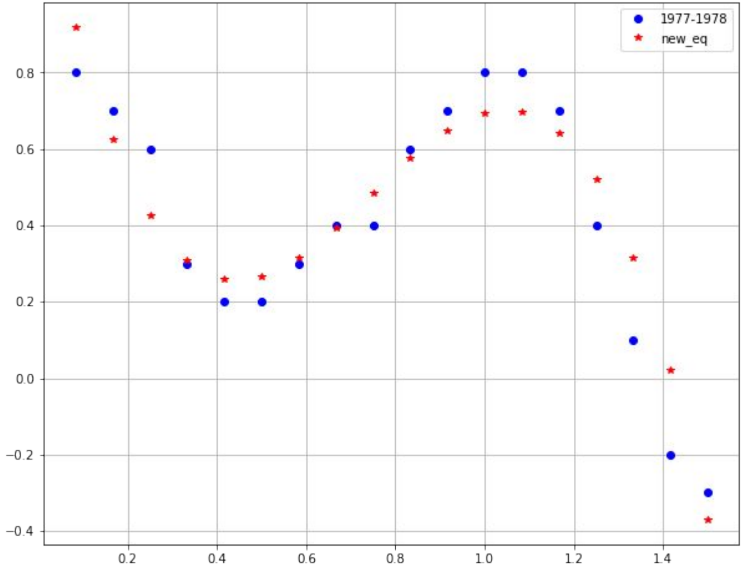
Violin Plots



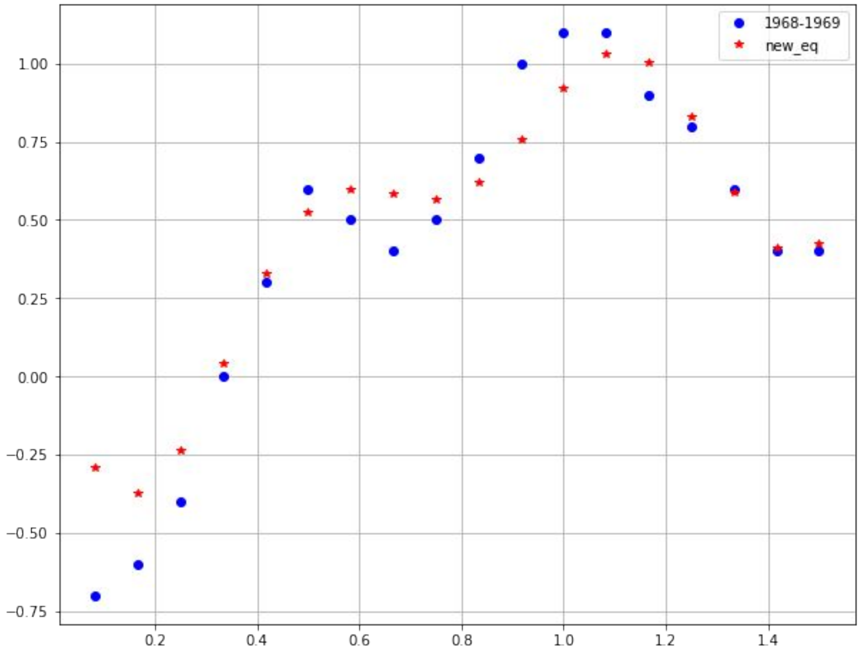
Finding equation of each shape in the each group using RVGA

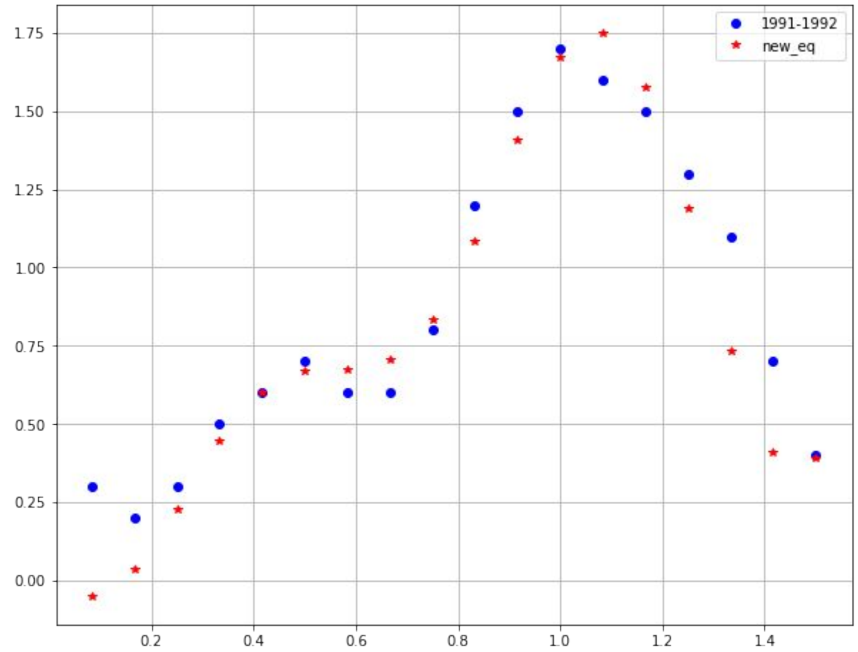
event	function	best fitness
avg	(1.4221796-x)*(2*x*(x-0.995554)+(2*x-0.64533426)*sin(x-0.694731)+1.0)	0.642469
1958-59 WE	(1.422180-x)*(2*x*(x-0.995554)+(2*x-**1.339720**)*sin(x-0.694731)+1.0)	11.8577
1977-78 WE	(1.422180-x)*(2*x*(x-0.995554)+(2*x-**(-0.114362)**)*sin(x-0.694731)+1.0)	10.795051



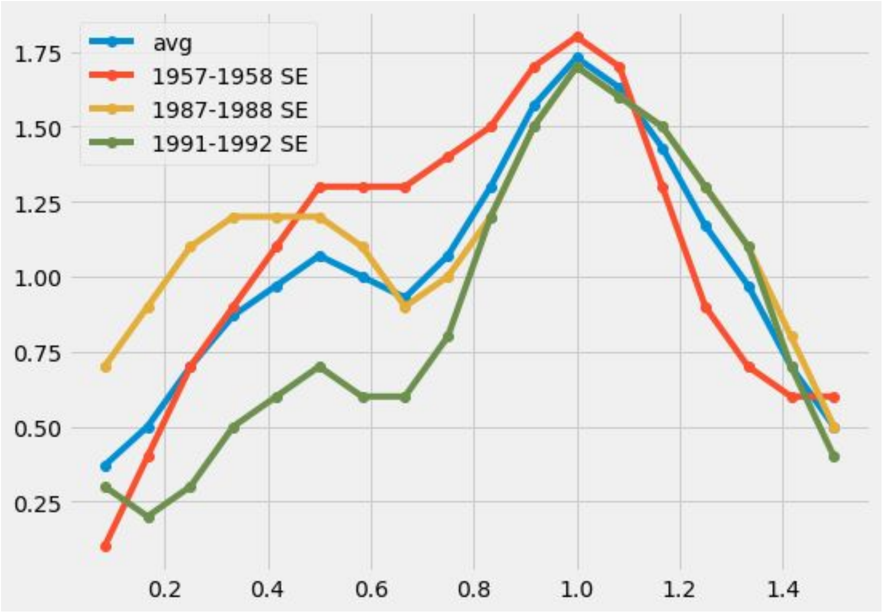


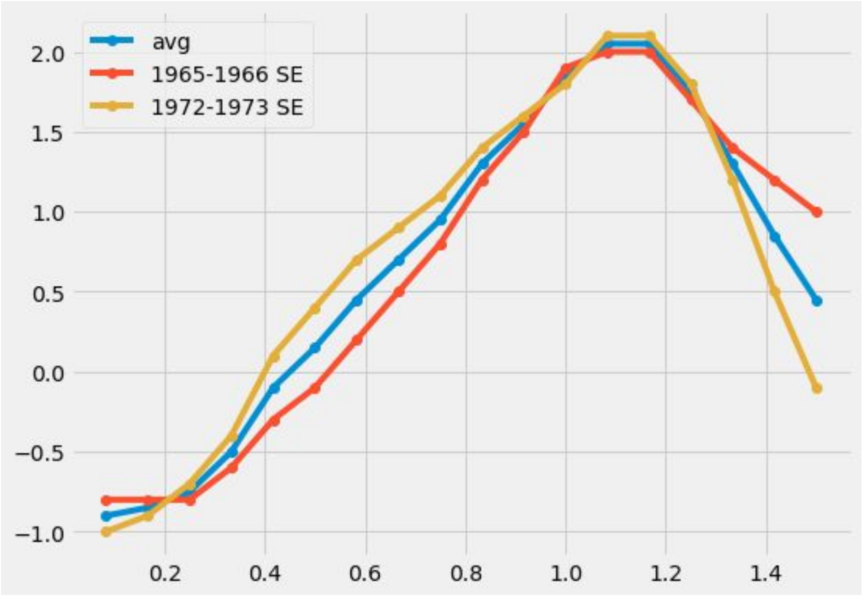
event	**function**	**best fitness**
avg	1.270450*sin(0.869869*x)+(0.873001*x-0.504249)*sin(7.622553*x)	1.49723
1968-69 ME	1.270450*sin(**0.607653***x)+(0.873001*x-**0.671472**)*sin(7.622553*x)	15.224091
1991-92 SE	1.270450*sin(**1.187490***x)+(0.873001*x-**0.368266**)*sin(7.622550*x)	15.917866

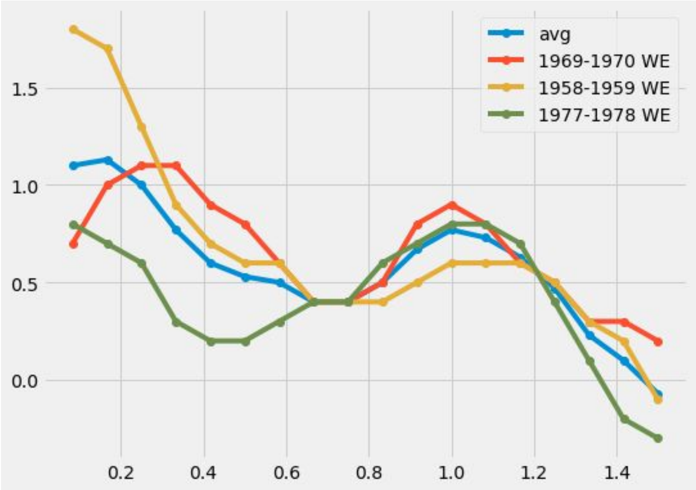


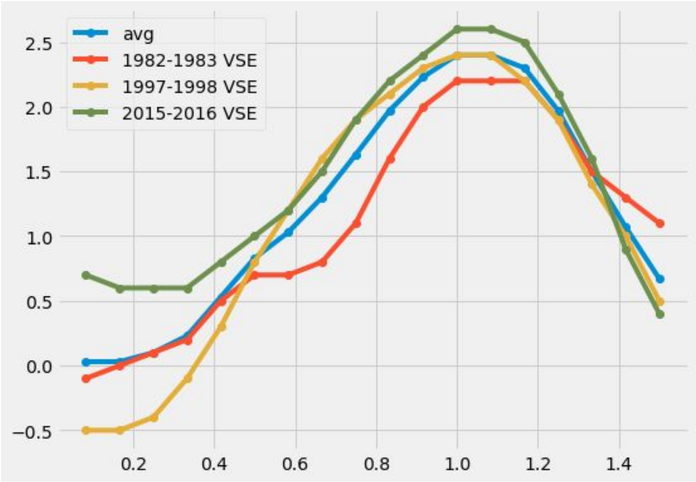


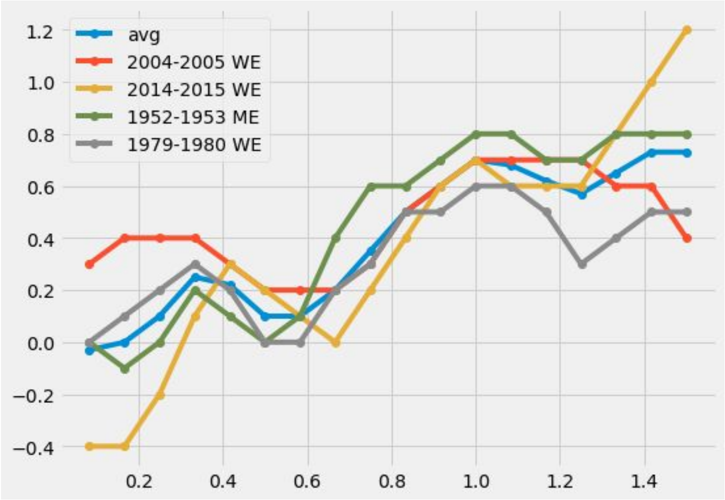
Grouped El Nino events according to shapes automatically by ML

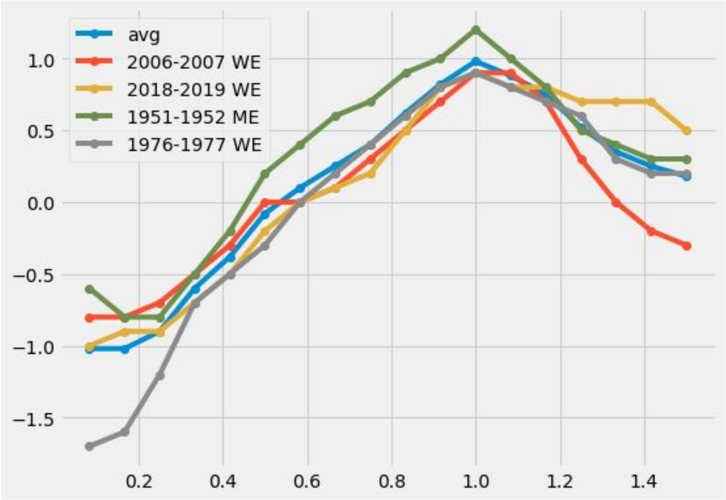


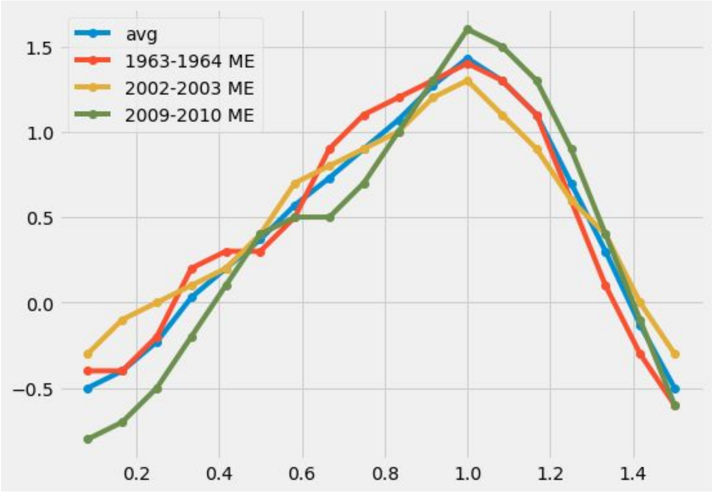



avg
1969-1970 WE
1958-1959 WE
1977-1978 WE
1.5
1.0
0.5
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

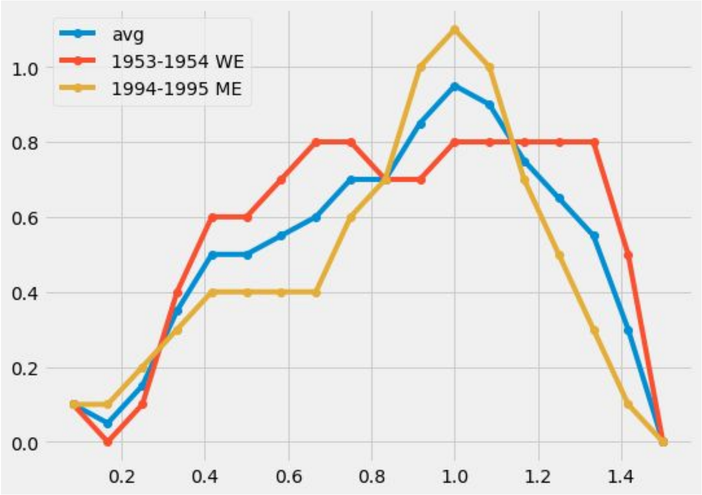

avg
1982-1983 VSE
1997-1998 VSE
2015-2016 VSE
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0.2
0.4
0.6
0.8
1.0
1.2
1.4


avg
2004-2005 WE
2014-2015 WE
1952-1953 ME
1979-1980 WE
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.2
0.4
0.6
0.8
1.0
1.2
1.4


avg
2006-2007 WE
2018-2019 WE
1951-1952 ME
1976-1977 WE
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.2
0.4
0.6
0.8
1.0
1.2
1.4

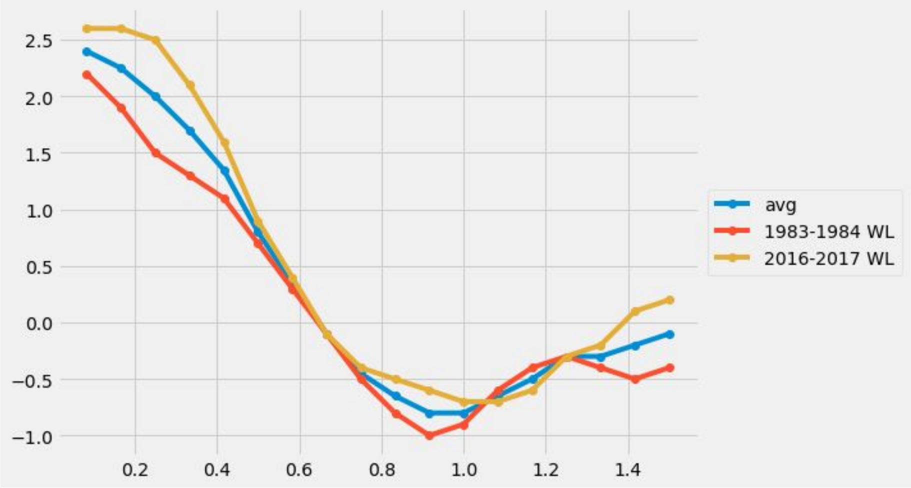

avg
1963-1964 ME
2002-2003 ME
2009-2010 ME

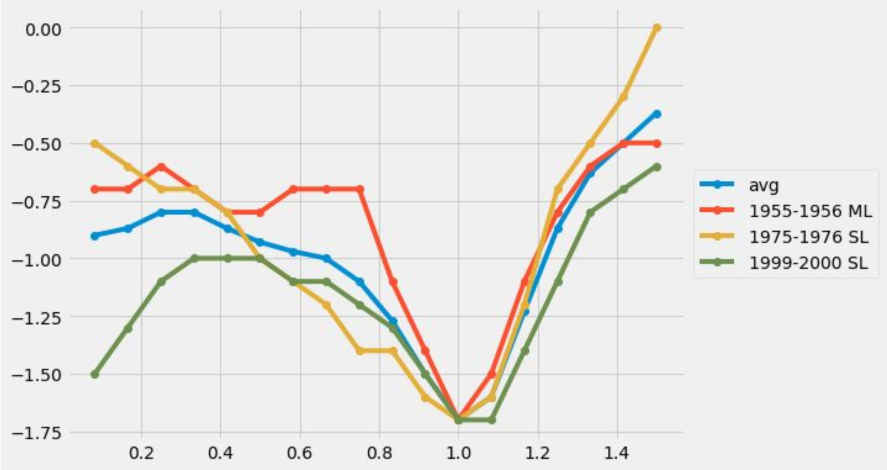

avg
1968-1969 ME
1986-1987 ME

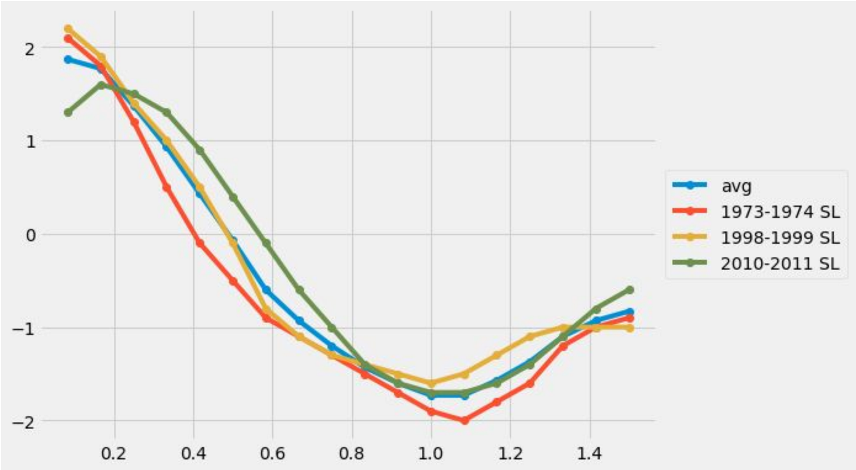

avg
1953-1954 WE
1994-1995 ME

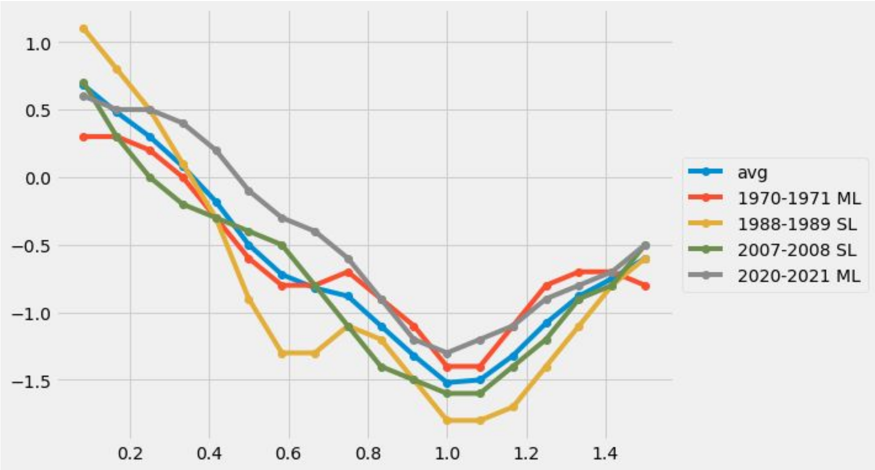
Grouped by eye		Grouped by ML	
Groups	Events	Groups	Events
1	52-53 WE, 14-15 WE	1	57-58 SE, 87-88 SE, 91-92 SE
2	69-70 WE, 79-80 WE, 87-88 SE 04-05 WE	2	65-66 SE, 72-73 SE
3	58-59 WE, 77-78 WE	3	69-70 WE, 58-59 WE, 77-78 WE
4	57-58 SE, 63-64 ME, 82-83 VSE 94-95 ME, 02-03 ME, 06-07 WE 09-10 ME	4	82-83 VSE, 97-98 VSE, 15-16 VSE
5	51-52 ME, 76-77 WE, 86-87 ME 18-19 WE	5	52-53 ME, 79-80 WE, 04-05 WE 14-15 WE
6	65-66 SE, 72-73 SE, 97-98 VSE 15-16 VSE	6	51-52 ME, 76-77 WE, 06-07 WE 18-19 WE
7	68-69 ME, 91-92 SE	7	63-64 ME, 02-03 ME, 09-10 ME
8	53-54 WE	8	68-69 ME, 86-87 ME
		9	53-54 WE, 94-95 ME

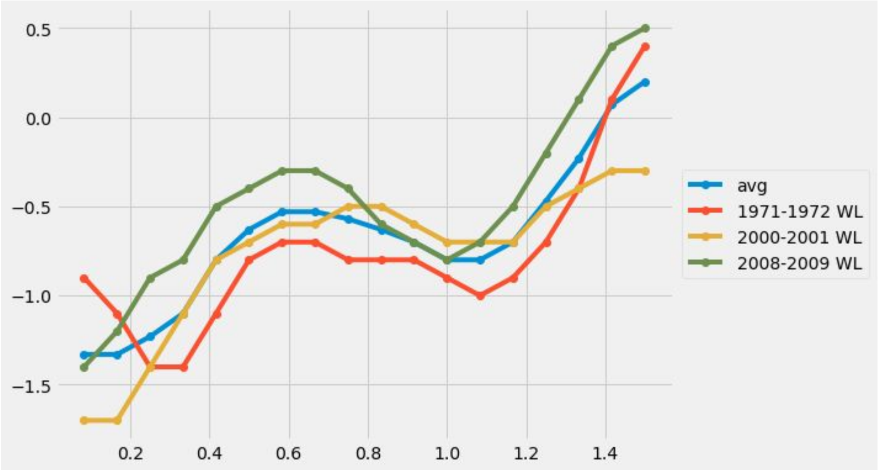
Grouped La Nina events according to shapes automatically by ML

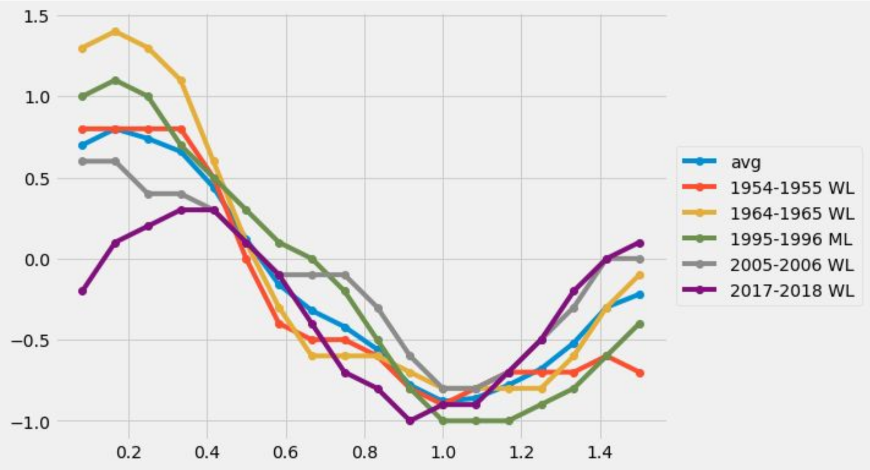




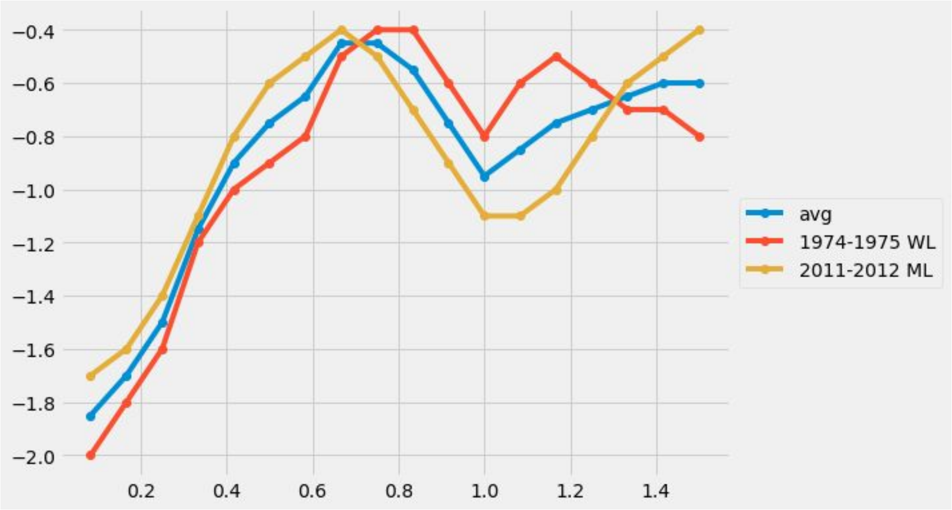

avg
1973-1974 SL
1998-1999 SL
2010-2011 SL


avg
1970-1971 ML
1988-1989 SL
2007-2008 SL
2020-2021 ML


avg
1971-1972 WL
2000-2001 WL
2008-2009 WL


avg
1954-1955 WL
1964-1965 WL
1995-1996 ML
2005-2006 WL
2017-2018 WL

−0.2
−0.4
−0.6
−0.8
−1.0
−1.2
0.2
0.4
0.6
0.8
1.0
1.2
1.4
avg
1984-1985 WL
2021-2022 ML



−0.4
−0.6
−0.8
−1.0
−1.2
−1.4
−1.6
−1.8
−2.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
avg
1974-1975 WL
2011-2012 ML

Discussion and Conclusion

- Classification of El Nino and La Nina events in the literature was confirmed using ML methods.
- New classification of El Nino and La Nina events were defined by shape.
- Simple equations were obtained using genetic programming to describe the behaviour of the events in each group.

Future perspectives

- To work on ML classified groups to find equations for the events.
- To redefine the strength of the events.
- To use these equations for predicting future events.

Thanks for your attention