SURGE Project Final Report

Statistical Forecasting of future COVID-19 wave in India

Gaurang dangayach | 200373

# Aim of the Project

Previously for predicting date for the fourth wave of COVID-19 in India we had used the data of the four waves of Zimbabwe. The aim of this project was to mathematically support the intuition of choosing Zimbabwe to predict about next wave of COVID-19 in India. I tried to work on how I can use mathematical tests to comment on the similarity of COVID-19 graphs of India and Zimbabwe with respect to other countries.

# Method Used

For achieving the desired goal, I had used the COVID-19 data of the countries and used Python and Google Colab as the platform to perform necessary operations on the data. Using the libraries of Python (*SciPy.stats* & *similaritymeasures*) I performed some tests which could be used to ascertain the similarity between the graphs of India and other countries.

# Procedure

## Taking Data

The data was taken from **COVID-19 Dataset by Our World in Data**.

I had taken the COVID-19 data of six countries to compare them with India –

1. Zimbabwe
2. South Africa
3. Afghanistan
4. U.S.A.
5. United Kingdom
6. Germany

Out of these six countries, the first three have graph appeared to be similar to that of India and have four or more waves. The graphs of next three countries are not very similar to India but they have been chosen to compare the results of the tests with the first three countries.

## Data Wrangling

#### Part 1

1. For my project, I have only used the columns of ‘date’, ‘number of new cases’ and ‘number of new cases per million’ from the dataset.
2. For each country, data corresponding to the first three waves was chosen, rest of the data was removed.
3. For some countries where the new cases were not reported daily, I first fill the N/A values with zero and then these zeroes were replaced with the average value of the preceding day and the following day.
4. The date column was replaced by the number of the day taking the first day when the data was reported as Day 0.

#### Part B

1. I created 1-D arrays for each country such that the day number appeared in it as many times as there were cases on that day.
2. Using these arrays as input for training the GMM (Gaussian Mixture Model), I found out the weights, means and standard deviations of the gaussian distributions corresponding the three waves.

## Different Transformations

1. **Height Neglected:** (done after Part B) Initially, this transformation was also done earlier by my senior guide. In this, I made each wave of unit height. This would help to consider only the position of each wave and not their respective heights. In this case, the third peak of India was ahead the peak of other countries. This issue was dealt in case of Case .
2. **Height Neglected and India Shifted:** (done after Part B) In this, after transformation as written in part A was performed, the graph of India was padded with extra days on the right so that the peaks of India matches more with other countries.
3. **Height Normalized:** (done after Part B) In this, instead of using the number of new cases as height of wave, the height of each wave was taken to be its weight as obtained in GMM. This was done to reduce the gap between the height of waves of different countries according to their number of cases. In this case, the third peak of India was ahead the peak of other countries. This issue was dealt in case of Case D.
4. **Height & Distance Normalized, India Shifted:** (done after Part B) In this, along with the height of the peaks, the total length of the timeline was also reduced to 500 days so that the three waves of each country falls within the same number of days.
5. **Original Data:** (done after Part A) I tried to run the test on the cleaned data obtained directly after the initial set of operations. I used the columns ‘Date’ and ‘Number of New Cases’. This was done to check if we could comment on the similarity at the initial stages without training models on the data. In this case, there was a lagre gap between the number of cases in different countries. This issue was dealt in case of Case F.
6. **Number of New Cases Per Million:** (done after Part A) Finally, I tried to run the test on the data obtained directly after the initial set of operations. This time, I used the columns ‘Date’ and ‘Number of New Cases per Million’. This was done to normalize the data according to the population of each country. This was also done by my senior guide earlier.

## Tests Used

Two categories of tests were used –

* Dataset Similarity Tests (from SciPy)
  + Kolmogorov-Smirnov Test - The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution functions of two samples.
  + Epps-Singleton Test – This test uses the Empirical Characteristic Function (ECF) which is Fourier Transform of the distribution function.
  + Kruskal-Wallis H-Test (not included) - The Kruskal–Wallis test is just the rank-sum test extended to more than two samples. Think of it informally as testing if the distributions have the same median.
  + Brunner-Munzel Test - This test is used for checking stochastic equality of two samples. Stochastic equality is a similarity measure between two populations, which means that the populations have roughly the same frequency of greater values.
  + Anderson Darling Test (not included) - This test is also based on Empirical Distribution Function of two samples. It is particularly sensitive to tail of the distributions.
  + Mann-Whitney U Test (not included) - This test is used for checking stochastic equality of two samples. In other words, it tells about the difference in medians of the distributions.
  + Pearson Correlation Coefficient - The Pearson correlation coefficient measures the linear relationship between two datasets. The calculation of the p-value relies on the assumption that each dataset is normally distributed.
  + Procrustes Analysis – It standardizes the matrices and then applies the optimal transform to the second matrix (including scaling/dilation, rotations, and reflections) to minimize , or the sum of the squares of the pointwise differences between the two input datasets.
* Curve Matching Tests (from similaritymeasures) - [1]
  + **Fréchet Distance** (not included) - Suppose a man is walking a dog, where the man is constrained to stay on one curve and the dog on another. The man and dog can vary their velocities independently at all times. Both the man and the dog are limited to either moving forward or stopping along their curves, as it is forbidden for them to move backwards. The Fréchet distance reflects the shortest possible leash connecting the man and dog sufficient to complete the walk along the curves.
  + **DTW (Dynamic Time Wrapping)** - DTW first calculates the distance between points of one curve onto the other curve. The goal of DTW is to find the path between curves that minimizes the cumulative distance between points.
  + **PCM (Partial Curve Mapping)** – In this, first the arc-length of the shorter curve is imposed onto a section on the longer curve. Then trapezoids are constructed between the curves, and the areas of the trapezoids are summed. The final PCM value is the minimum area from all attempted arc-length offsets.
  + **Area between Curves** – In this, corresponding points on both the curve are joined to form quadrilaterals and the sum of the areas of all the quadrilaterals is taken and used as a similarity measure.
  + **Curve Length Measure** - A corresponding data point on the numerical model is calculated at the equivalent curve length location of the experimental curve. Squared residual values are then calculated as function of both the dependent and independent variables. The sum of these squared residuals is used to quantify the difference between the two curves.

*[1] (Information taken from Similarity measures for identifying material parameters from hysteresis loops using inverse analysis Charles F. Jekel · Gerhard Venter · Martin P. Venter · Nielen Stander · Raphael T. Haftka)*

# Observations

## Test Results

1. **Height Neglected:**
   1. KS Test:

Ind vs Zim : Ks\_2sampResult(statistic=0.094, pvalue=0.02406614705241275)

Ind vs SA : Ks\_2sampResult(statistic=0.174, pvalue=5.009206937519233e-07)

Ind vs Afg : Ks\_2sampResult(statistic=0.236, pvalue=1.2717086973677652e-12)

Ind vs USA : Ks\_2sampResult(statistic=0.434, pvalue=1.1311548920484481e-42)

Ind vs UK : Ks\_2sampResult(statistic=0.238, pvalue=7.848024594418195e-13)

Ind vs Ger : Ks\_2sampResult(statistic=0.256, pvalue=8.416140722635984e-15)

* 1. Epps Singleton Test:

Ind vs Zim : Epps\_Singleton\_2sampResult(statistic=6.116558375224596, pvalue=0.19061108152019626)

Ind vs SA : Epps\_Singleton\_2sampResult(statistic=14.823573916550323, pvalue=0.005081478447884015)

Ind vs Afg : Epps\_Singleton\_2sampResult(statistic=50.89635753785516, pvalue=2.3463465558306747e-10)

Ind vs USA : Epps\_Singleton\_2sampResult(statistic=271.52022176672796, pvalue=1.5000040271262235e-57)

Ind vs UK : Epps\_Singleton\_2sampResult(statistic=54.019069795372246, pvalue=5.214515137385852e-11)

Ind vs Ger : Epps\_Singleton\_2sampResult(statistic=70.12575693512879, pvalue=2.1352353468740725e-14)

* 1. Brunner Munzel Test:

Ind vs Zim : BrunnerMunzelResult(statistic=-2.342002621671666, pvalue=0.01937708983839647)

Ind vs SA : BrunnerMunzelResult(statistic=3.5147968829065652, pvalue=0.00045993310047842684)

Ind vs Afg : BrunnerMunzelResult(statistic=6.800982800031113, pvalue=1.7927659357042103e-11)

Ind vs USA : BrunnerMunzelResult(statistic=15.273107137417918, pvalue=0.0)

Ind vs UK : BrunnerMunzelResult(statistic=6.90576453366317, pvalue=8.889555758173628e-12)

Ind vs Ger : BrunnerMunzelResult(statistic=7.536099291541836, pvalue=1.0857981180834031e-13)

* 1. Curve Length Measure:

Ind vs Zim : 3.6808012789824054

Ind vs SA : 4.377087400363241

Ind vs Afg : 4.529255206218274

Ind vs USA : 8.513384247159282

Ind vs UK : 11.038448085271447

Ind vs Ger : 14.359887248553608

* 1. PCM:

Ind vs Zim : 19.939858943725604

Ind vs SA : 15.664463677908707

Ind vs Afg : 17.170868471519714

Ind vs USA : 52.542120210853135

Ind vs UK : 28.52212478654105

Ind vs Ger : 48.578658938805454

* 1. DTW:

Ind vs Zim : 135.5672344557471

Ind vs SA : 142.70877187431014

Ind vs Afg : 199.98977663460076

Ind vs USA : 157.08373979846348

Ind vs UK : 208.44007428482777

Ind vs Ger : 224.30856257003208

* 1. Area between Two Curves:

Ind vs Zim : 135.56396981923356

Ind vs SA : 142.70762007988296

Ind vs Afg : 199.98948555940825

Ind vs USA : 157.00023345363127

Ind vs UK : 208.28976001828175

Ind vs Ger : 224.29886389502417

* 1. Pearson Correlation Coefficient:

Ind vs Zim : 0.23123021698269122

Ind vs SA : 0.2633449943100264

Ind vs Afg : -0.11271458850994522

Ind vs USA : 0.509775771630028

Ind vs UK : -0.10671726526241458

Ind vs Ger : -0.18781015639407872

* 1. Procrustes Analysis:

Ind vs Zim : 7.447129277112512e-06

Ind vs SA : 6.9486562411095155e-06

Ind vs Afg : 9.626813519123311e-06

Ind vs USA : 4.817819000863912e-06

Ind vs UK : 1.0012163280861124e-05

Ind vs Ger : 9.037894271675071e-06

1. **Height Neglected and India Shifted:**
   1. KS Test:

Ind vs Zim : Ks\_2sampResult(statistic=0.16, pvalue=0.011843449760085422)

Ind vs SA : Ks\_2sampResult(statistic=0.225, pvalue=7.536785951251345e-05)

Ind vs Afg : Ks\_2sampResult(statistic=0.285, pvalue=1.4639479233432268e-07)

Ind vs USA : Ks\_2sampResult(statistic=0.485, pvalue=1.0971108602139472e-21)

Ind vs UK : Ks\_2sampResult(statistic=0.29, pvalue=8.114129962716335e-08)

Ind vs Ger : Ks\_2sampResult(statistic=0.235, pvalue=2.9607827368482373e-05)

* 1. Epps Singleton Test:

Ind vs Zim : Epps\_Singleton\_2sampResult(statistic=2.9945914080619813, pvalue=0.5587309229413727)

Ind vs SA : Epps\_Singleton\_2sampResult(statistic=15.956021901947254, pvalue=0.0030787469756572163)

Ind vs Afg : Epps\_Singleton\_2sampResult(statistic=35.47834693672445, pvalue=3.704463163928036e-07)

Ind vs USA : Epps\_Singleton\_2sampResult(statistic=139.2086531396313, pvalue=4.1692349447494176e-29)

Ind vs UK : Epps\_Singleton\_2sampResult(statistic=34.953310458622965, pvalue=4.749068423263853e-07)

Ind vs Ger : Epps\_Singleton\_2sampResult(statistic=21.921282719554135, pvalue=0.00020778223219854273)

Brunner Munzel Test:

Ind vs Zim : BrunnerMunzelResult(statistic=0.4065915020210623, pvalue=0.6845363460655052)

Ind vs SA : BrunnerMunzelResult(statistic=3.9187400253333244, pvalue=0.00010562018759552316)

Ind vs Afg : BrunnerMunzelResult(statistic=5.937059863798745, pvalue=6.536261309975089e-09)

Ind vs USA : BrunnerMunzelResult(statistic=10.888341968487454, pvalue=0.0)

Ind vs UK : BrunnerMunzelResult(statistic=6.027275268015134, pvalue=3.9163565723754346e-09)

Ind vs Ger : BrunnerMunzelResult(statistic=4.725734226299594, pvalue=3.2052630911660884e-06)

Curve Length Measure:

Ind vs Zim : 3.105985982420123

Ind vs SA : 3.9160182726061725

Ind vs Afg : 3.324937661842337

Ind vs USA : 7.972687425499658

Ind vs UK : 9.16198245949238

Ind vs Ger : 9.645889228737973

* 1. PCM:

Ind vs Zim : 8.84629431735245

Ind vs SA : 5.682382215135162

Ind vs Afg : 6.031221426978156

Ind vs USA : 23.238514288600907

Ind vs UK : 18.978068209429576

Ind vs Ger : 20.73972401654204

* 1. DTW:

Ind vs Zim : 25.260938538400477

Ind vs SA : 41.91125548977209

Ind vs Afg : 64.56030406857064

Ind vs USA : 69.85005949230562

Ind vs UK : 89.97340767432857

Ind vs Ger : 83.47921891548202

* 1. Area between Two Curves:

Ind vs Zim : 25.25726907974464

Ind vs SA : 41.90971481477514

Ind vs Afg : 64.55974512709209

Ind vs USA : 69.76814292174294

Ind vs UK : 89.82247858325591

Ind vs Ger : 83.47158942504653

* 1. Pearson Correlation Coefficient:

Ind vs Zim : 0.764381555748457

Ind vs SA : 0.5761266843264121

Ind vs Afg : 0.2714693590766768

Ind vs USA : 0.441426928924902

Ind vs UK : -0.20708329703135456

Ind vs Ger : -0.13593625141637994

* 1. Procrustes Analysis:

Ind vs Zim : 1.2882487308206306e-05

Ind vs SA : 1.899548470233755e-05

Ind vs Afg : 4.772778416284285e-05

Ind vs USA : 3.095474717511848e-05

Ind vs UK : 5.435222198488344e-05

Ind vs Ger : 5.890786711599336e-05

1. **Height Normalized:**
   1. KS Test:

Ind vs Zim : Ks\_2sampResult(statistic=0.235, pvalue=2.9607827368482373e-05)

Ind vs SA : Ks\_2sampResult(statistic=0.17, pvalue=0.006094418258803505)

Ind vs Afg : Ks\_2sampResult(statistic=0.24, pvalue=1.8266119303942767e-05)

Ind vs USA : Ks\_2sampResult(statistic=0.19, pvalue=0.0014264303851701861)

Ind vs UK : Ks\_2sampResult(statistic=0.23, pvalue=4.7487878961137165e-05)

Ind vs Ger : Ks\_2sampResult(statistic=0.26, pvalue=2.3797467071606504e-06)

* 1. Epps Singleton Test:

Ind vs Zim : Epps\_Singleton\_2sampResult(statistic=9.001717846640668, pvalue=0.0610565573069297)

Ind vs SA : Epps\_Singleton\_2sampResult(statistic=44.3190707909314, pvalue=5.507644483695521e-09)

Ind vs Afg : Epps\_Singleton\_2sampResult(statistic=11.701937467173732, pvalue=0.01971099513310235)

Ind vs USA : Epps\_Singleton\_2sampResult(statistic=30.97792028125325, pvalue=3.093309309957793e-06)

Ind vs UK : Epps\_Singleton\_2sampResult(statistic=50.16703107483428, pvalue=3.3322205670627225e-10)

Ind vs Ger : Epps\_Singleton\_2sampResult(statistic=57.60129212954842, pvalue=9.252604916343866e-12)

* 1. Brunner Munzel Test:

Ind vs Zim : BrunnerMunzelResult(statistic=2.27778227801853, pvalue=0.023376561058046086)

Ind vs SA : BrunnerMunzelResult(statistic=2.1525488920196225, pvalue=0.03196913190699524)

Ind vs Afg : BrunnerMunzelResult(statistic=3.636462174133348, pvalue=0.00031424814396863177)

Ind vs USA : BrunnerMunzelResult(statistic=2.556180280412185, pvalue=0.010967459360987863)

Ind vs UK : BrunnerMunzelResult(statistic=3.839531820832304, pvalue=0.0001437782738258342)

Ind vs Ger : BrunnerMunzelResult(statistic=4.88145625144065, pvalue=1.554773049861069e-06)

* 1. Curve Length Measure:

Ind vs Zim : 17.092722957957683

Ind vs SA : 10.051409918353704

Ind vs Afg : 13.584781246881409

Ind vs USA : 7.472384849255539

Ind vs UK : 9.109232711617318

Ind vs Ger : 9.688520685854355

* 1. PCM:

Ind vs Zim : 29.362710359836523

Ind vs SA : 15.993210412163275

Ind vs Afg : 18.646362095691277

Ind vs USA : 7.435984506544252

Ind vs UK : 13.865450137201755

Ind vs Ger : 12.9263277868091

* 1. DTW:

Ind vs Zim : 41.00386966380471

Ind vs SA : 28.818021291114995

Ind vs Afg : 38.85407139167689

Ind vs USA : 19.434967461231043

Ind vs UK : 41.074958874121926

Ind vs Ger : 43.51139381428193

* 1. Area between Two Curves:

Ind vs Zim : 41.00303605733892

Ind vs SA : 28.817178812269134

Ind vs Afg : 38.85312140539196

Ind vs USA : 19.433996369472396

Ind vs UK : 41.07404369685129

Ind vs Ger : 43.510533868489155

* 1. Pearson Correlation Coefficient:

Ind vs Zim : -0.02123400422750987

Ind vs SA : 0.3047502602275829

Ind vs Afg : -0.11174741141863946

Ind vs USA : 0.70395437700505

Ind vs UK : 0.010414504727562616

Ind vs Ger : -0.0548814749020985

* 1. Procrustes Analysis:

Ind vs Zim : 5.0051584034216326e-05

Ind vs SA : 1.8092255837595475e-05

Ind vs Afg : 2.7792776469891497e-05

Ind vs USA : 7.743603507786905e-06

Ind vs UK : 2.6227665955038238e-05

Ind vs Ger : 2.6915160683133597e-05

1. **Height & Distance Normalized, India Shifted:**
   1. KS Test:

Ind vs Zim : Ks\_2sampResult(statistic=0.165, pvalue=0.008539483949831865)

Ind vs SA : Ks\_2sampResult(statistic=0.21, pvalue=0.0002829222008740643)

Ind vs Afg : Ks\_2sampResult(statistic=0.18, pvalue=0.0030098470801667434)

Ind vs USA : Ks\_2sampResult(statistic=0.26, pvalue=2.3797467071606504e-06)

Ind vs UK : Ks\_2sampResult(statistic=0.21, pvalue=0.0002829222008740643)

Ind vs Ger : Ks\_2sampResult(statistic=0.21, pvalue=0.0002829222008740643)

* 1. Epps Singleton Test:

Ind vs Zim : Epps\_Singleton\_2sampResult(statistic=12.063216728654554, pvalue=0.016887308781309875)

Ind vs SA : Epps\_Singleton\_2sampResult(statistic=58.65672858627302, pvalue=5.555232146179389e-12)

Ind vs Afg : Epps\_Singleton\_2sampResult(statistic=11.729048358678122, pvalue=0.019484082016560565)

Ind vs USA : Epps\_Singleton\_2sampResult(statistic=52.70461023049576, pvalue=9.825015300886363e-11)

Ind vs UK : Epps\_Singleton\_2sampResult(statistic=67.53915690769202, pvalue=7.503395640668419e-14)

Ind vs Ger : Epps\_Singleton\_2sampResult(statistic=73.35553504833153, pvalue=4.437446375795505e-15)

* 1. Brunner Munzel Test:

Ind vs Zim : BrunnerMunzelResult(statistic=-0.6987722262270907, pvalue=0.4851144331624424)

Ind vs SA : BrunnerMunzelResult(statistic=2.8161007081863954, pvalue=0.00510300829769017)

Ind vs Afg : BrunnerMunzelResult(statistic=2.0271512535577725, pvalue=0.043317385528601315)

Ind vs USA : BrunnerMunzelResult(statistic=4.822693340157932, pvalue=2.0574030925590847e-06)

Ind vs UK : BrunnerMunzelResult(statistic=1.510346447876205, pvalue=0.1317587976819421)

Ind vs Ger : BrunnerMunzelResult(statistic=1.254372993193683, pvalue=0.21045376255221382)

* 1. Curve Length Measure:

Ind vs Zim : 16.89807387405243

Ind vs SA : 9.559694762315296

Ind vs Afg : 13.441832578484153

Ind vs USA : 7.719719496625525

Ind vs UK : 7.994465263758347

Ind vs Ger : 8.447290960833062

* 1. PCM:

Ind vs Zim : 22.5974433047991

Ind vs SA : 20.336063485971454

Ind vs Afg : 21.379681969941334

Ind vs USA : 9.103824337834768

Ind vs UK : 11.675717773591197

Ind vs Ger : 14.535072685224332

* 1. DTW:

Ind vs Zim : 39.47192853450411

Ind vs SA : 29.26265419765714

Ind vs Afg : 41.48023968182025

Ind vs USA : 25.88734716018074

Ind vs UK : 42.42571488814012

Ind vs Ger : 54.00309421098887

* 1. Area between Two Curves:

Ind vs Zim : 39.463006934488256

Ind vs SA : 29.260340104981488

Ind vs Afg : 41.4741115306052

Ind vs USA : 25.883924909991563

Ind vs UK : 42.421054280416726

Ind vs Ger : 53.96711799031851

* 1. Pearson Correlation Coefficient:

Ind vs Zim : 0.46540663594410975

Ind vs SA : 0.536955496882833

Ind vs Afg : 0.09126279831448246

Ind vs USA : 0.6889250572305813

Ind vs UK : 0.17588086151485613

Ind vs Ger : -0.2669249380843316

* 1. Procrustes Analysis:

Ind vs Zim : 2.9290845785140078e-05

Ind vs SA : 1.3236852970485278e-05

Ind vs Afg : 3.110789593662247e-05

Ind vs USA : 9.102321214695903e-06

Ind vs UK : 2.4331004146744625e-05

Ind vs Ger : 1.825263220378163e-05

1. **Original Data:**
   1. KS Test:

Ind vs Zim : Ks\_2sampResult(statistic=0.9648093841642229, pvalue=2.43875125747567e-243)

Ind vs SA : Ks\_2sampResult(statistic=0.6874991145678383, pvalue=7.240429943278472e-134)

Ind vs Afg : Ks\_2sampResult(statistic=0.9688628926006205, pvalue=3.631657620071078e-249)

Ind vs USA : Ks\_2sampResult(statistic=0.1892444367776436, pvalue=4.0313086202559134e-11)

Ind vs UK : Ks\_2sampResult(statistic=0.5185812518960461, pvalue=1.2186587691436046e-73)

Ind vs Ger : Ks\_2sampResult(statistic=0.5749560117302053, pvalue=7.571784922581526e-88)

* 1. Epps Singleton Test:

Ind vs Zim : Epps\_Singleton\_2sampResult(statistic=2058.847987149311, pvalue=0.0)

Ind vs SA : Epps\_Singleton\_2sampResult(statistic=1250.7702790951691, pvalue=1.5686508028033668e-269)

Ind vs Afg : Epps\_Singleton\_2sampResult(statistic=2048.8567975989627, pvalue=0.0)

Ind vs USA : Epps\_Singleton\_2sampResult(statistic=122.18607091086771, pvalue=1.8225301019172178e-25)

Ind vs UK : Epps\_Singleton\_2sampResult(statistic=421.65507154546344, pvalue=5.817642027996323e-90)

Ind vs Ger : Epps\_Singleton\_2sampResult(statistic=873.280864945461, pvalue=1.024677121915286e-187)

* 1. Brunner Munzel Test:

Ind vs Zim : BrunnerMunzelResult(statistic=-1094.85877529649, pvalue=0.0)

Ind vs SA : BrunnerMunzelResult(statistic=-66.15386705384478, pvalue=0.0)

Ind vs Afg : BrunnerMunzelResult(statistic=-1127.2496647723322, pvalue=0.0)

Ind vs USA : BrunnerMunzelResult(statistic=6.3545624062729615, pvalue=2.846529678635079e-10)

Ind vs UK : BrunnerMunzelResult(statistic=-29.230964363770152, pvalue=3.217255531583181e-142)

Ind vs Ger : BrunnerMunzelResult(statistic=-47.12942019241196, pvalue=2.9684768097852375e-277)

* 1. Curve Length Measure:

Ind vs Zim : 19.595217860038353

Ind vs SA : 18.435262473335904

Ind vs Afg : 20.674682861708636

Ind vs USA : 27.271068022341815

Ind vs UK : 17.734699761085636

Ind vs Ger : 18.751915439840708

* 1. PCM:

Ind vs Zim : 68.20879901093912

Ind vs SA : 78.21203810039448

Ind vs Afg : 92.19773473328874

Ind vs USA : 45.28915581508065

Ind vs UK : 62.4032254174066

Ind vs Ger : 69.94745640220015

* 1. DTW:

Ind vs Zim : 68.20879901093912

Ind vs SA : 78.21203810039448

Ind vs Afg : 92.19773473328874

Ind vs USA : 45.28915581508065

Ind vs UK : 62.4032254174066

Ind vs Ger : 69.94745640220015

* 1. Area between Two Curves:

Ind vs Zim : 149721826.34375

Ind vs SA : 173072786.25

Ind vs Afg : 221980418.53125

Ind vs USA : 569082144.5

Ind vs UK : 268963463.5

Ind vs Ger : 300858528.5625

* 1. Pearson Correlation Coefficient: Can’t run because of different sizes of arrays
  2. Procrustes Analysis: Can’t run because of different sizes of arrays

1. **Number of New Cases Per Million:**
   1. KS Test:

Ind vs Zim : Ks\_2sampResult(statistic=0.4917839414145055, pvalue=1.5422589661749226e-63)

Ind vs SA : Ks\_2sampResult(statistic=0.21511515340407347, pvalue=1.1302070390684094e-13)

Ind vs Afg : Ks\_2sampResult(statistic=0.5994836159632794, pvalue=1.1756366227255798e-95)

Ind vs USA : Ks\_2sampResult(statistic=0.6518759703294807, pvalue=1.1102230246251565e-15)

Ind vs UK : Ks\_2sampResult(statistic=0.3654413995348367, pvalue=8.765837585260372e-37)

Ind vs Ger : Ks\_2sampResult(statistic=0.24580645161290324, pvalue=1.1102230246251565e-16)

* 1. Epps Singleton Test:

Ind vs Zim : Epps\_Singleton\_2sampResult(statistic=321.2060559597268, pvalue=2.880319504745488e-68)

Ind vs SA : Epps\_Singleton\_2sampResult(statistic=80.3526917161715, pvalue=1.4665014404049837e-16)

Ind vs Afg : Epps\_Singleton\_2sampResult(statistic=527.1339240487978, pvalue=9.054373202033149e-113)

Ind vs USA : Epps\_Singleton\_2sampResult(statistic=811.3268217409196, pvalue=2.7030510498285885e-174)

Ind vs UK : Epps\_Singleton\_2sampResult(statistic=199.60825382378798, pvalue=4.561384930488744e-42)

Ind vs Ger : Epps\_Singleton\_2sampResult(statistic=130.86642902410333, pvalue=2.5415652941404608e-27)

* 1. Brunner Munzel Test:

Ind vs Zim : BrunnerMunzelResult(statistic=-21.129041449387497, pvalue=4.3074593926969305e-80)

Ind vs SA : BrunnerMunzelResult(statistic=6.887537863848542, pvalue=9.315437310419838e-12)

Ind vs Afg : BrunnerMunzelResult(statistic=-34.77393031112259, pvalue=2.3189351641475156e-178)

Ind vs USA : BrunnerMunzelResult(statistic=28.016469717730935, pvalue=0.0)

Ind vs UK : BrunnerMunzelResult(statistic=13.630113994706633, pvalue=0.0)

Ind vs Ger : BrunnerMunzelResult(statistic=2.91337085919573, pvalue=0.003672341350113406)

* 1. Curve Length Measure:

Ind vs Zim : 23.423620060131952

Ind vs SA : 33.94465381780811

Ind vs Afg : 20.122722803870232

Ind vs USA : 54.932786473116145

Ind vs UK : 44.40496547287163

Ind vs Ger : 34.77852061752734

* 1. PCM:

Ind vs Zim : 66.55424839499061

Ind vs SA : 80.77363756779376

Ind vs Afg : 91.00664075038819

Ind vs USA : 134.01755049624742

Ind vs UK : 147.4408162767735

Ind vs Ger : 70.24342706496962

* 1. DTW:

Ind vs Zim : 28295.55531585147

Ind vs SA : 45196.761444422424

Ind vs Afg : 38796.79113910178

Ind vs USA : 117514.68990969757

Ind vs UK : 108848.44149120555

Ind vs Ger : 57478.9718794882

* 1. Area between Two Curves:

Ind vs Zim : 218731.53125

Ind vs SA : 468024.125

Ind vs Afg : 235872.5

Ind vs USA : 1435255.25

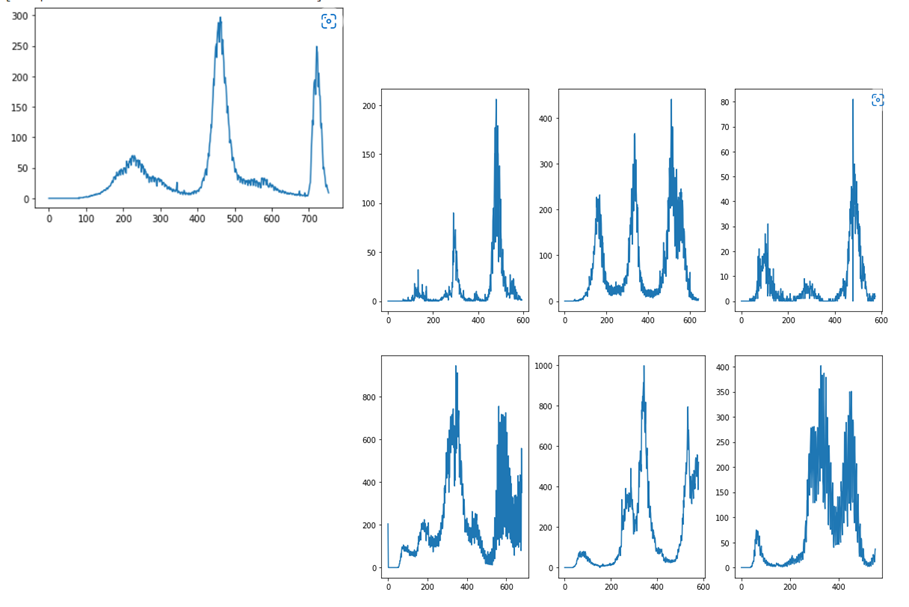
Ind vs UK : 887281.6875

Ind vs Ger : 904083.75

* 1. Pearson Correlation Coefficient: Can’t run because of different sizes of arrays
  2. Procrustes Analysis: Can’t run because of different sizes of arrays

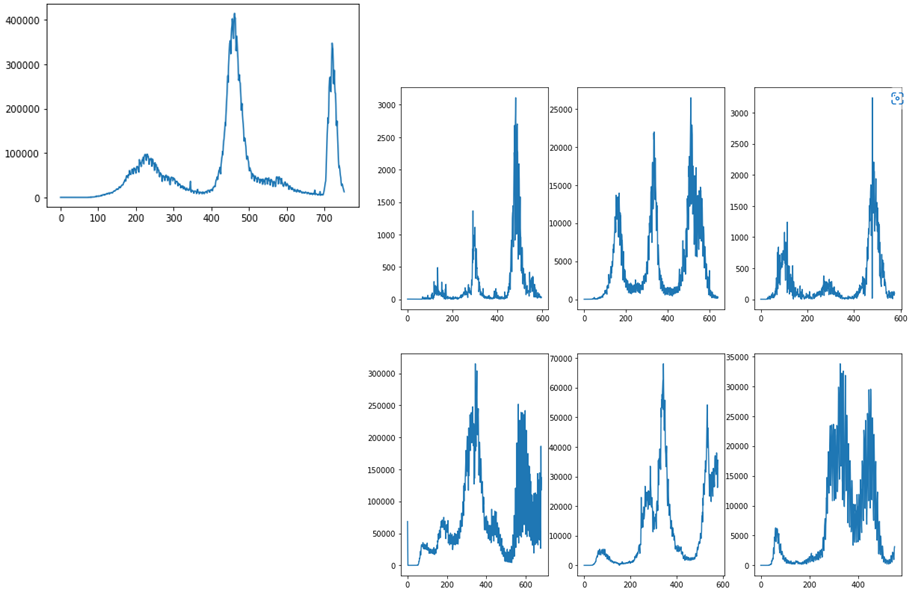
## Summary of Results & Graphs

I have taken the test results obtained in the **Case F : Number of Cases per Million** as the basis to comment on the similarity of the graphs of India and Zimbabwe with respect to other countries. While the Dataset Similarity Tests from SciPy didn’t gave satisfactory results, later when I tried Curve Matching Tests from similaritymeasures, all of them worked well on the data and the results for Zimbabwe were best in case of PCM, DTW & Area method and satisfactorily good in case of Curve Length Measure.



For other cases:

In **Case E : Original Data**, the results were not satisfactory in most of the tests. One factor which is creating the issue maybe the large gap between the number of cases in different countries. To overcome this issue, in Case F, I had taken Number of New Cases per Million instead of Number of New Cases.



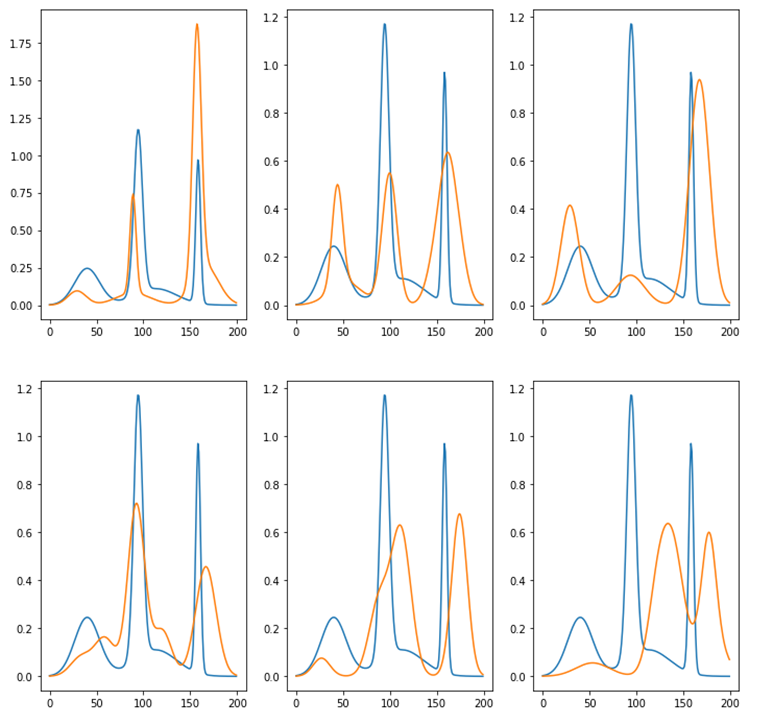
In **Case C: Height Normalized,** the results in case of Dataset Similarity Tests from SciPy were acceptable but not in the favour of Zimbabwe, in case of Curve Matching Tests from similaritymeasures, the results were not good and favoured USA.

Chart, histogram

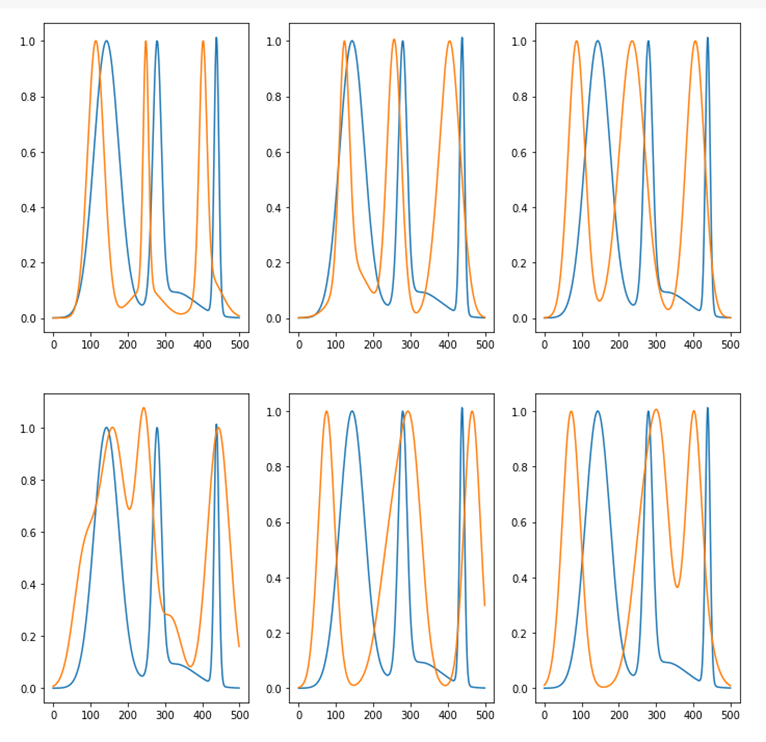
Description automatically generated

In **Case D: Height & Distance Normalized & India Shifted**, the results in case of Dataset Similarity Tests from SciPy were overall more in the favour of Zimbabwe, but in case of Curve Matching Tests from similaritymeasures, the results were still not good and favoured USA.

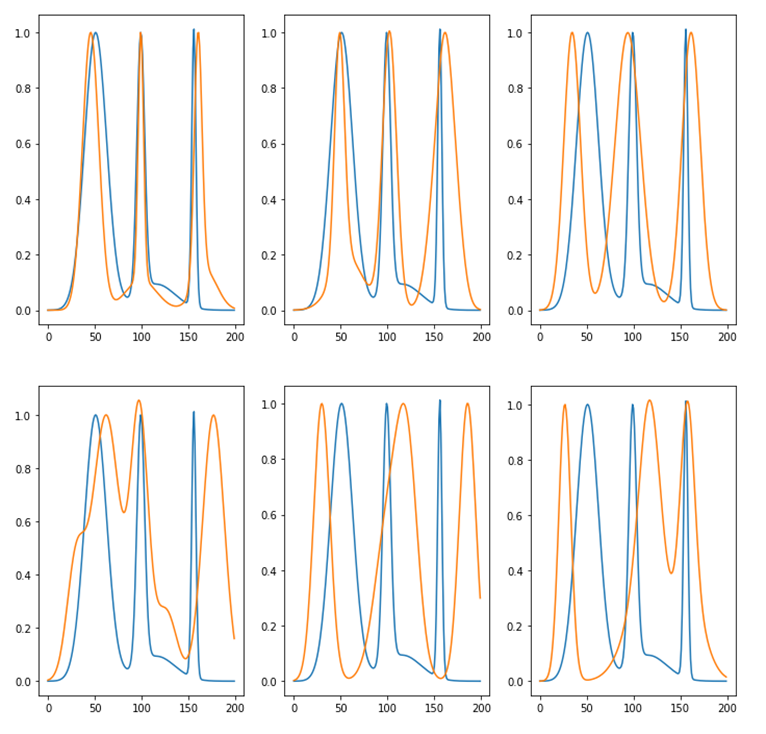
The reason for this is that the graph of USA is not much similar to India but the position of the second wave of both the countries is very much same and their shape is also matching.

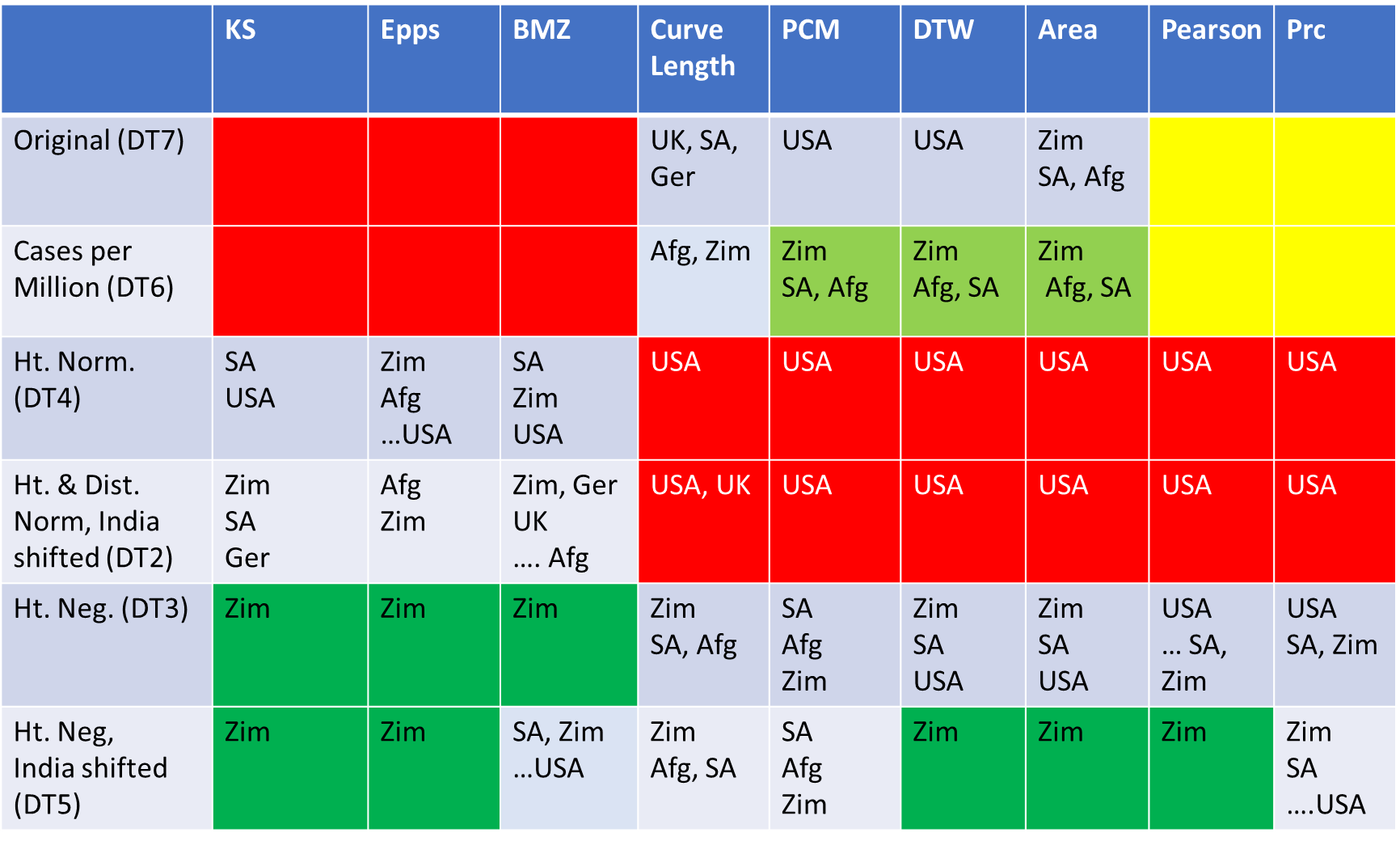


In **Case B: Height Neglected**, the results in case of Dataset Similarity Tests from SciPy were very much in favour of Zimbabwe, while in case of Curve Matching Tests from similaritymeasures, the results were overall not very decisive.



In **Case C: Height Neglected and India Shifted**, the result of Brunner Munzel Test became less in favour of Zimbabwe, while in case of Curve Matching Tests from similaritymeasures, the results were overall much better and in favour of Zimbabwe.





# Results

On the basis of results obtained by the Curve Matching Tests from similaritymeasures library of Python for Case F: Number of New Cases per Million, the graph of Zimbabwe is similar to India with respect to other countries and hence Zimbabwe is a good choice to use as model for India.