

Tissue Spatial Geometrics
Laboratory



Lawrence Berkeley
National Laboratory

How to Apply Pavlidis Template Matching (PTM) to Data From the Planarized Globe Plot (PGP) Projection

David H. Nguyen^{1,2}, Jun Ha Cho³, Manfred Auer³

1. Tissue Spatial Geometrics Laboratory; 2. Stanford University;
3. Lawrence Berkeley National Laboratory

Contact: david.hh.n@gmail.com

May 25, 2019

Part 1:
Explanation of
Pavlidis Template Matching
(PTM)

Pavlidis Template Matching (PTM)

Figure 1a. Examples of the PTM Coefficient for data that is known to exhibit high or low correlation. This is just to prove that the automated script works at the most basic level.

All sample data is available in the GitHub folder.

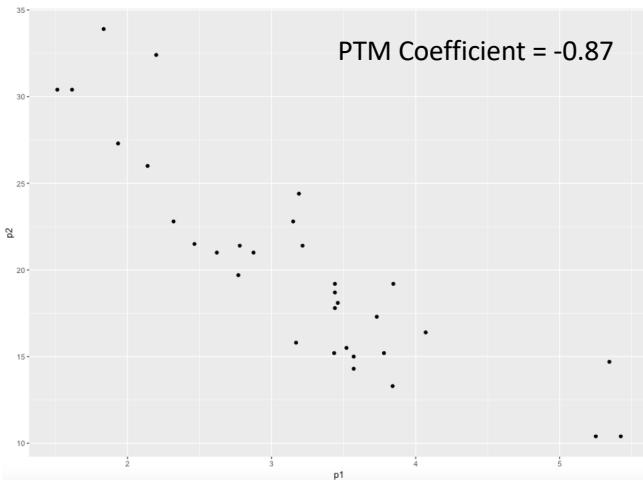
The Pearson Correlation Coefficient

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

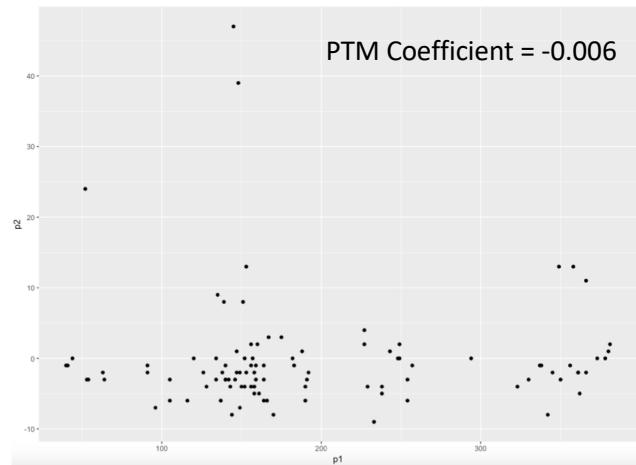
\bar{X} is the mean of group X.

\bar{Y} is the mean of group Y.

Sample Data 1



Sample Data 2



Reference: Pavlidis & Noble (2001). "Analysis of strain and regional variation in gene expression in mouse brain." *Genome Biology*. 2(10):research0042.1–0042.15

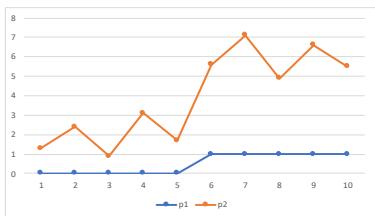
Note: In their paper, their correlation equation did not use square roots for the two terms in the denominator, which may have been a typo.

Figure 2. Examples of PTM Coefficients resulting from correlating data to a template that consists of binary values (0's and 1's). Data are discrete points; lines were drawn between points for ease of viewing.

A

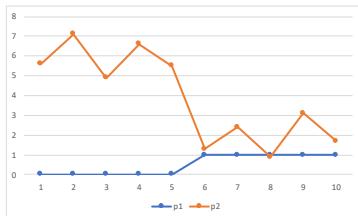
Sample Data 3

PTM Coefficient = 0.93



Sample Data 4

PTM Coefficient = -0.93

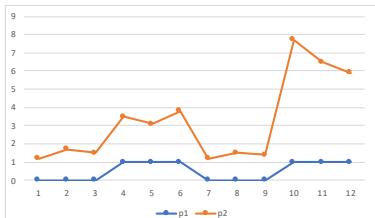


The PTM coefficient quantifies similarity in patterns. The orange (top) lines in Sample 3 and 4 are mirror images of each other. They have the opposite PTM coefficient when compared to the same blue (bottom) line.

B

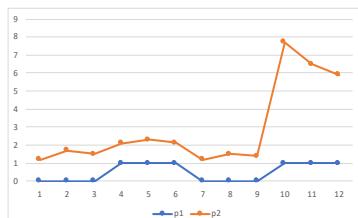
Sample Data 5

PTM Coefficient = 0.83



Sample Data 6

PTM Coefficient = 0.68

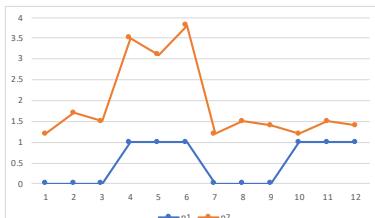


The orange (top) lines in sample 5 and 6 only differ at positions 4-6 on the horizontal axis. The bump at these locations is less pronounced in sample 6, thus it has a weaker PTM coefficient.

C

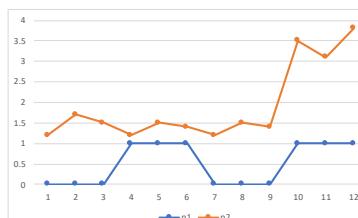
Sample Data 7

PTM Coefficient = 0.55



Sample Data 8

PTM Coefficient = 0.55

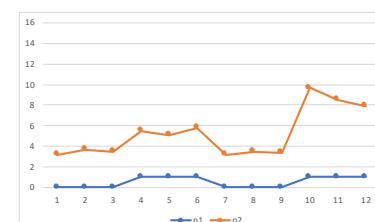


The orange (top) lines in sample 7 and 8 are the same except for swapping values at locations 4-6 and 10-12 on the horizontal axis. Both samples have the same PTM coefficient when compared to the blue (bottom) line.

D

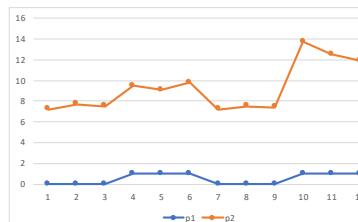
Sample Data 9

PTM Coefficient = 0.83



Sample Data 10

PTM Coefficient = 0.83



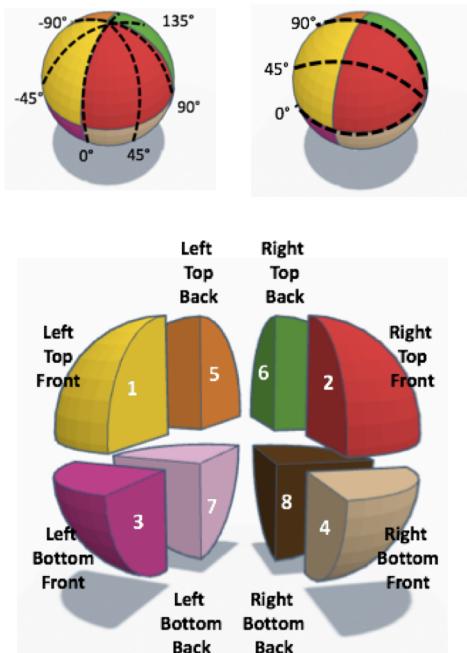
PTM prioritizes pattern over scale. The orange (top) lines in sample 9 and 10 are the same pattern, but the values in sample 10 are larger by a value of 4 compared to sample 9. Both samples have the same PTM coefficient.

Part 2:
Applying PTM to Data from the
Planarized Globe Plot (PGP)
Projection

The Structure of PGP Data

- PGP data has at least 24 numerical values (8 sectors x 3 coordinates = 24 values). See figure below.
 - See next page for a higher resolution PGP system.
- Thus, in order to compare two PGP profiles via the PTM method, each profile **MUST** have the same number of values so that the two profiles pair evenly.
- Remember, how you number the sectors is arbitrary, but you must keep the numbering consistent between samples.

Figure 3a. Explanation of the PGP Projection. Each of the 8 sectors will contain points by which a 3D centroid can be calculated (mean of x, mean of y, mean of z) for each sector. Therefore, an 8-sector grid results in a PTM profile of 24 values (8x3=24). Sectors that have no points in them will have a centroid of (0,0,0); it is important that empty sectors also have a numerical value for their centroids.



A 2D System to Visualize a 3D System

- 1st Dimension: X-axis (Vertical Coordinate)
- 2nd Dimension: Y-axis (Horizontal Coordinate)
- 3rd Dimension: Color intensity or marker size.

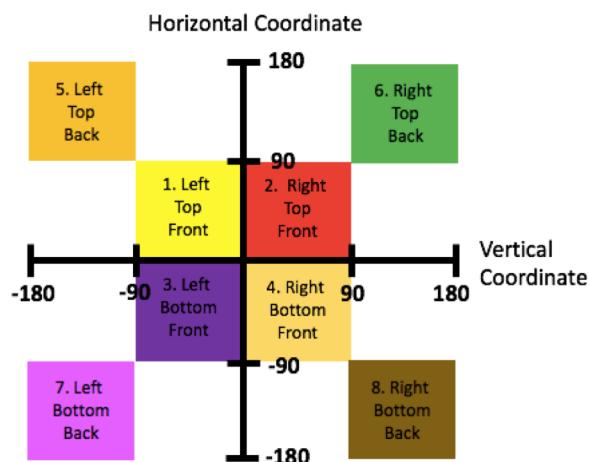
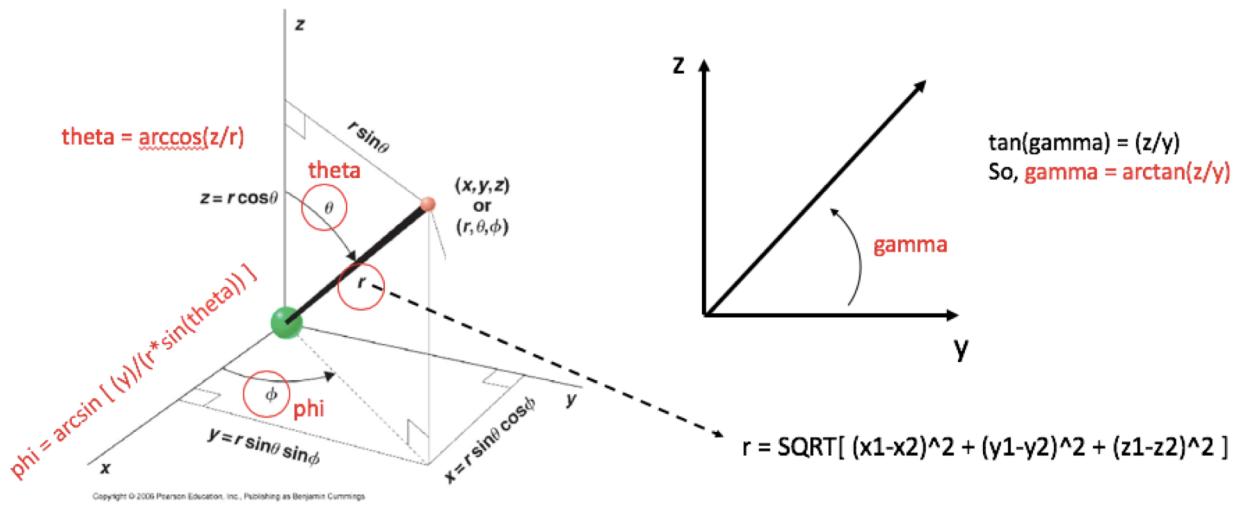


Figure 3b. How (x, y, z) Coordinates Are Converted to the PGP Projection. The PGP projection uses the angles phi and gamma as the x and y coordinates, respectively. The magnitude of the line extending from the origin (r) is the z coordinate.



An 8-unit grid or 32-unit grid?

Figure 4. A 32-Unit Projection System Provides Higher Resolution Than an 8-Unit Projection System. The PGP can be divided into 32 units instead of 8 units. This is useful if a higher resolution grid might be more useful in capturing dense data. Since each unit still has 3 coordinates (the phi angle, the gamma angle, the vector magnitude), a 32-unit grid will result in a 96-value PTM profile ($32 \times 3 = 96$).

The numbering of each unit is arbitrary, but the same numbering system should be used for all samples. The R script that accompanies this file uses the below numbering system.

17	18					21	22
19	20					23	24
1	2	5	6				
3	4	7	8				
9	10	13	14				
11	12	15	16				
25	26					29	30
27	28					31	32

Figure 5. Quantifying a model of Epithelial-to-Mesenchymal Transition (EMT) via the PTM analysis of PGP Projections. Blue dots and red dots can represent distinct cellular organelles or points on a membrane fold. The PGP projection was done with a 32-unit grid (explained in Figure 4).

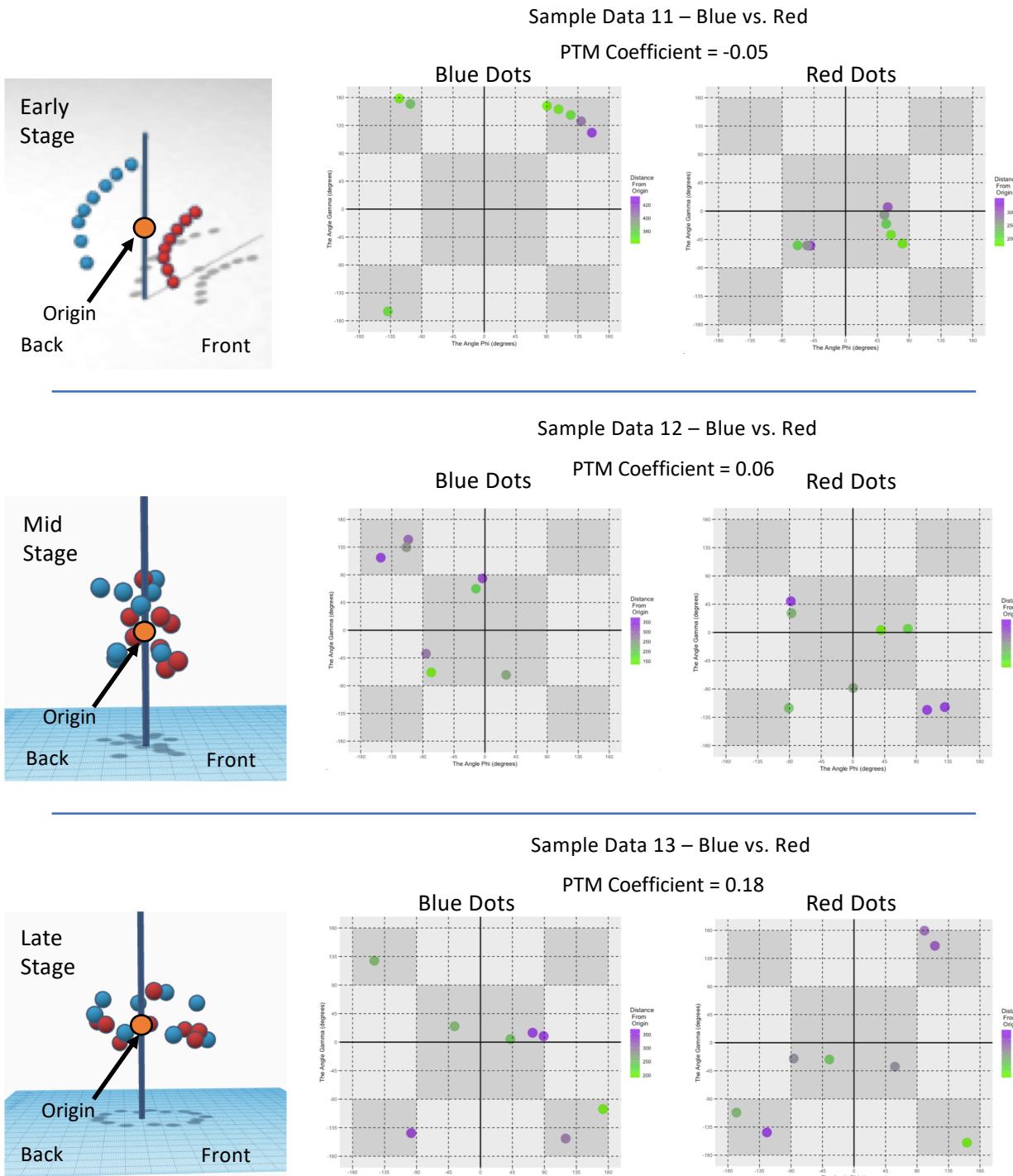
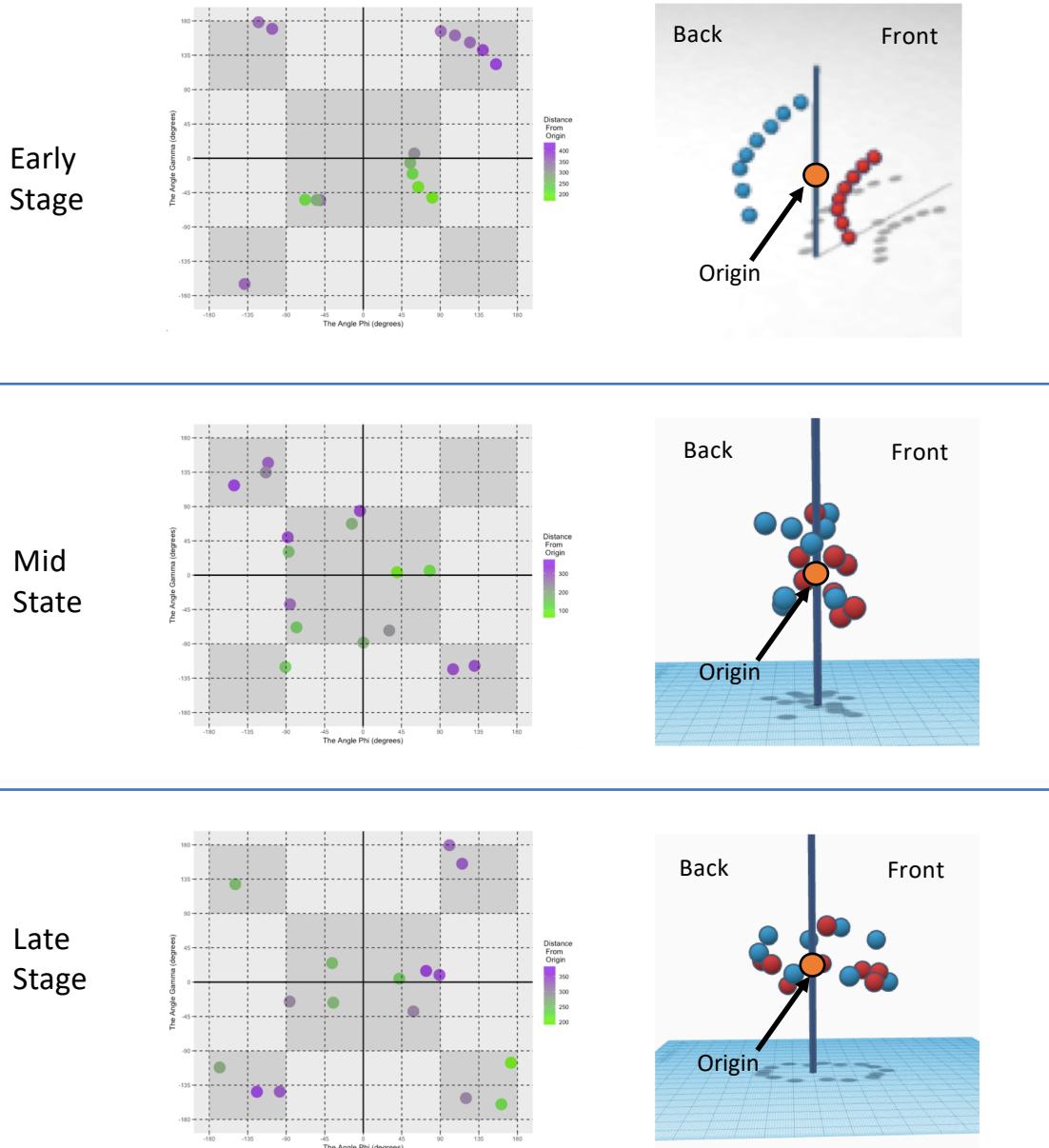
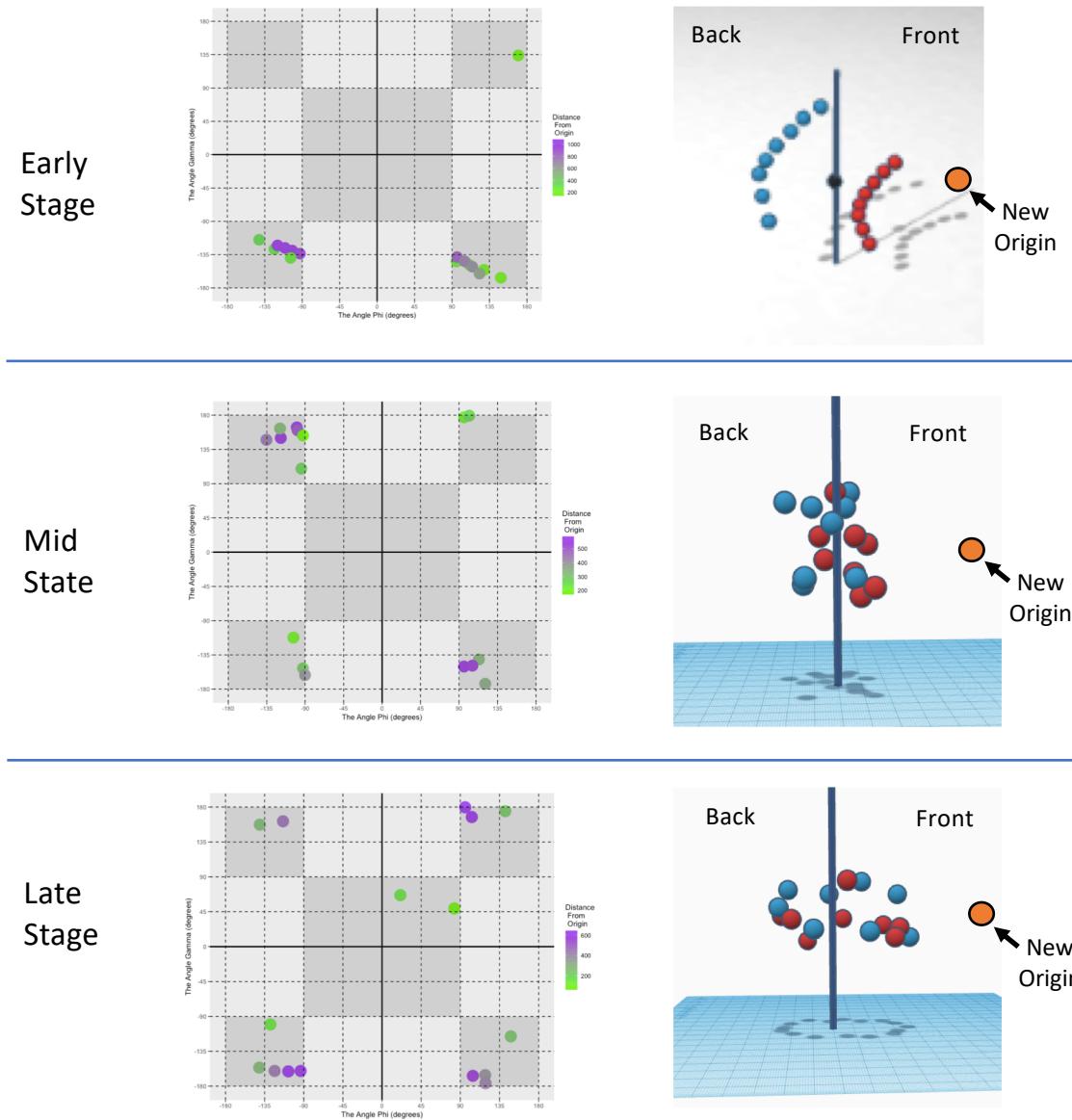


Figure 6a. Placing the Origin as Center May Not Be Optimal for Capturing Spatial Differences. Placement of the origin is arbitrary and must be defined for each system/question asked, but must be kept consistent across samples. Compare to Figure 6b.



Comparison (Ignore Blue vs Red)	PTM Coefficient	Sample Data
Early Stage vs Mid Stage	0.14	Sample Data 14
Early Stage vs Late Stage	0.19	Sample Data 15
Mid Stage vs Late Stage	0.18	Sample Data 16

Figure 6b. A caveat to applying PTM analysis to PGP data is that the location of the origin is arbitrary and can alter the resulting PTM coefficient. The below data is based on defining the origin in each stage as outside and in front of the dots. This dramatically changes the resulting PTM values. Thus, defining the location of the origin is not a one-size-fits-all decision for all applications. Compare this to Figure 6a.



Comparison (Combine Blue & Red)	PTM Coefficient	Sample Data
Early Stage vs Mid Stage	0.43	Sample Data 17
Early Stage vs Late Stage	0.34	Sample Data 18
Mid Stage vs Late Stage	0.78	Sample Data 19

Summary of Key Points

1. The number of grid units can be expanded for higher resolution. The automated script on GitHub that accompanies this document is set to create a 32-unit grid.
2. The placement of the origin of a 3D system is arbitrary. This has pros and cons.
 - The pro is that it allows the researcher to define the origin at a point that is most scientifically meaningful.
 - For example, quantifying the spatial information in apoptosis requires defining an origin that is independent of cellular markers, since these organelles dissipate as apoptosis progresses.
 - The con is that different origins can drastically change the PTM results.
3. Be consistent about the following across all samples and between experiments:
 - The number of grid units in the PGP projection.
 - The numbering of the grid units in the PGP projection.
 - The placement of origin in the 3D object/system.