

Perform_Principal_Component_Analysis

This tutorial will show you how to use Inspect3D in order to perform [Principal Component Analysis](#) (PCA) using data from a [CMO library](#). For a full treatment of waveform-based PCA to find differences in waveform data, see the explanation presented in [the Research Methods in Biomechanics textbook](#).

For this tutorial, we will be comparing the the knee flexion angles between participants with osteoarthritis and the normal control group. Our problem is to provide an explanation for differences in knee flexion angles between osteoarthritic walking versus normal walking. We can accomplish this by defining two groups that meet these signal definitions, performing PCA, and interpreting the results.

Data

This tutorial uses overground walking data from roughly 100 subjects divided into two conditions, normal control and osteoarthritis (moderate to severe). This data set is included in Demo folder of your Inspect3D installation (e.g., C:\Program Files\C-Motion\Inspect3D\Demo).

Set the library path to the data directory

As with previous tutorials, we begin by loading the CMO library and defining the queries relevant to our question.

1. Click  **Load Library** in the [toolbar](#) to open the Load Library dialog.
2. Click **Browse** and select the folder where the data is stored.
3. Click **Load** button to import the data.

Define queries and calculate groups

For this tutorial we will manually create two groups based on tags, one for subjects with osteoarthritis and one for normal control subjects. We begin by defining a query for subjects with the OA tag (indicating osteoarthritis).

1. Click on the  **Query Definitions** icon on the [toolbar](#) and then click on  **Query Definitions** to open the Query Builder dialog.
2.  Add a query, name it OA, and click **Save**.
3.  Add a condition and name it OA.
 - 3.1. **Signals:** Set TYPE - DERIVED, FOLDER - PCA, NAME - RKNEE_ANGLE, COMPONENT - X. This is the only signal in the data set.
 - 3.2. **Events:** There are no events in this data set, so this tab can be skipped.

3.3. **Refinements**: Check the **Refine using tag** checkbox and select the OA tag.

3.4. Click **Save**

Next we will define a query for subjects with the NC tag (indicating Normal Control). In this case we can easily modify our previous query rather than starting from scratch.

1.  Add a query, name it NC, and click **Save**.
2.  Add a condition and name it NC.
3. In the **Refinements** tab, change the selected tag from OA to NC.
4. Click **Save**.

You can verify here that the new NC group has the same signal and event selections as the OA group. Click **Calculate All Queries** and then close the Query Builder dialog.

Exploring the data

There will now be two Groups in the Queried Data subwindow's **Groups** list. Selecting either of these will display the associated workspaces in the **Workspaces** list below. We can verify the queries that we produced in the first section by visualizing our traces.

1. Select all groups and all workspaces.
2. Check only the **Plot Workspace Mean** checkbox.
3. Click **Refresh Plot**.

The plot that is produced will not be very informative if the traces are not coloured by group, which is the comparison we are interested in. If this is the case, open the  **Show Options** dialog from the application tool bar and under **Plotting options** set the **Display Styles from** to "Group".



Once this is done, inspecting the plot shows that although participants from the two groups walk very differently, obvious when watching them in-person, their knee flexion angles are quite similar. Because the traces overlap significantly between the groups throughout the entire gait cycle, conventional statistics will likely not be useful for describing the differences between these two groups.



This is one of the motivations behind PCA: by transforming our original data into a coordinate system based on principal components we will end up with a few dimensions that explain most of the variance in the data set. This, in turn, will help us to explain and detect the differences between the groups.

Running Principal Component Analysis

1. Ensure that all groups and workspaces are selected in the **Groups** and **Workspaces** lists.
2. Open the  **PCA Options** dropdown menu on the [toolbar](#).
3. Set **Number PCs** to 4.
4. Ensure that **Use Workspace Mean** is checked.
5. Click  **Run PCA**.
6. The results of these calculations will automatically populate the PCA graphs. If these aren't already displayed, click  **Show PCA Graphs** in the  **PCA Options** dropdown menu.



Interpreting PCA results

The results from the PCA are described in 6 windows: Variance Explained, Loading Vector, Workspace Scores, Group Scores, Extreme Plot, and PC Reconstruction. We will focus on four specific windows that are enlightening for this tutorial; full descriptions for each window are available [here](#).

Variance Explained

The Variance Explained window, which displays the variance explained by each principal component as well as the cumulative variance for each principal components. It is important to verify that the calculated principal components do explain a significant amount of the data set's variability. A good heuristic to use is that you want enough principal components to explain 95% of the data set's variety, otherwise there will be at least a moderate amount of variation that your analysis has not captured. In this example, our 4 principal components explain 96% of the data set's variability, which is sufficient and we can continue the exploration. If there less than 95% of the data set's variance was explained then we should re-run the analysis with more principal components.



Clicking on a bar will display the exact variance explained by that principal component. In this example, the second principal component explains 17.6% of the variation in the original data set.

Group Scores

Next, the Group Scores window helps us to determine if there is a significant differences between the groups for any individual principal component. This is established by looking at each principal component's mean score and their standard errors, where if the mean scores for a principal component +/- their standard errors do not cross the x-axis then there are significant differences in the scores for this principal component. In this tutorial principal components 1 and 2 show significant differences for the groups and will therefore be the focus of the remaining investigation.



As with the Variance Explained window, hovering over a point will display a tooltip with precise values for that group score.

Workspace Scores



Having identified our principal components of interest, switch to the Workspace Scores window and plot the chosen principal components against one another (i.e., for this tutorial, plot PC1 versus PC2). We are looking to be able to distinguish our groups and in this case there is a pretty clear separation between the groups in this 2D space. Although PC1 may describe more of the data set's variability than PC2, PC2 can discriminate between the groups better (this isn't mathematically shown here, but can be seen if the visual inspection is followed by a cluster analysis). Of course, PC1 and PC2 can discriminate between the groups even better if they are used together instead of separately. If you want to investigate the original data for a particular workspace, click on the workspace's scatter point and the corresponding data will be highlighted in the original Queried Data plot. Similarly, if you click on a trace in the original Queried Data plot then the corresponding workspace is highlighted in the Workspace Score widget. This can be useful to identify outliers.



Extreme Plot

The Extreme Plot widget allows us to animate the principal components to get some further insight into our problem.

For the **Plot Type** drop-down menu, select the Mean and Slider option. Leave the **PC Number** set to 1 and then click **Play**. The mean reconstruction (gray in the figure below) will remain static while the other reconstruction (blue in the figure below) will vary in the range Mean $\pm$ 2 Standard Deviations. Watching this animation demonstrates that PC1 is capturing variability as an offset in the signal. Vertical offsets in gait data can occur for a variety of reasons, including differences in standing calibration posture. Therefore, PC1 may explain a significant amount of variability that is not very helpful for answering our question.



Now select **PC Number** 2 and click **Play** again. This PC shows a different story, with the largest differences in this animation occurring during late stance and mid-swing.

Navigate Tutorials

Previous: [Tutorial 2: Clean your Data](#)

Inspect3D Getting Started Tutorials

Next: [Tutorial 4: Export Results](#)

From:
<https://has-motion.com/wiki/> - **Wiki Documentation Site**

Permanent link:
https://has-motion.com/wiki/doku.php?id=inspect3d:tutorials:perform_principal_component_analysis&rev=1719433140

Last update: **2024/06/26 20:19**

