

ShapeLME: Longitudinal statistical analysis tutorial

Deformetrica gives us a way to estimate smooth subject-specific shape models. These personalized models remove extraneous, unwanted variability and noise we find in structural segmentations extracted from images scanned over time. But given a longitudinal data set, a natural question is to understand the representative behavior of a group of subjects belonging to a certain class, say, subjects at risk of Huntington’s disease (HD). Another quest is to distinguish between two different groups of subjects, say, healthy vs disease. In contrast to subject-specific modeling, the optimization problem for diffeomorphic longitudinal modeling is challenging and existing approaches haven’t yet provided a consistent way to compare two different trajectories.

To address this, we use **ShapeLME** to compute linear-mixed-effects longitudinal models [2] on outputs of Deformetrica, i.e., we take the Deformetrica estimated personalized model for each subject and evaluate the model-predicted regressed shapes at observed time-points. We use the particle optimization framework for longitudinal shape analysis developed in [1] to summarize the behavior of a specific group.

Compute shape sequences at observed time points

- **reg_path**: path to deformetrica regression directory - **regression**
- **output_base_path**: path to new directory with regressed shape files at observed time points - **reg_at_obs_time_pts**

Run the script:

- **python copy_regression_time_points_scripts.py**

Diffeomorphic trajectories are good at capturing smooth anatomical shape changes, while the particle optimization framework excels at finding compact statistical shape representations with increased statistical power. The idea here is to leverage the strengths of both approaches, to provide an integrated solution, characterized by improved statistical performance in the analysis of both scalar and shape trajectory data derived from noisy segmentations. Details of this approach can be found in [3].

Note that

- Each shape observation is in longitudinal correspondence with every other shape. This was set up in the Deformetrica preprocessing pipeline.

- Deformetrica estimates diffeomorphic models for entire shape complexes (such as the striatal complex of left/right caudate/putamen), whereas here we separate out each structure, center them, align each structure to a template structure, and estimate longitudinal models per structure. This doesn't have to be the case. Here too, one could consider the whole shape complex, with an appropriate alignment, and estimate models that way as long as there is longitudinal correspondence across subjects and time-points. This will be incorporated in future iterations of this software, but for now, we estimate models for each structure separately.
- Each shape observation for each subject is stored as ***.vtk** with information about mesh vertices, polygonal connectivity, surface normals. To do longitudinal statistical modeling, we extract just the point correspondence positions, i.e., the mesh vertices, and store it in **lpts** format.

From shape meshes (*.vtk) to point correspondences (*.lpts)

- **reg_path:** path to regressed observed shape meshes - `reg_at_obs_time_pts`
- **output_base_path:** path to new folder with extracted point correspondence files - `shape_lpts_for_stats`

Run the script:

- **python write_vtk_to_lpts_script.py**

The linear mixed-effects model, originally developed by Laird and Ware [2] has the ability to incorporate multiple simultaneous relationships. For the PREDICT-HD dataset we work with in this tutorial, covariates such as risk of onset classes (Ctrl, Low, Med, High), sex, genetic scores, etc., are among few of the confounders one can control for. This model was extended to study longitudinal shape by Datar, et al. [1] which has been extended further to study relationships between shape evolution and multiple covariates.

As a toy example for this tutorial, we consider 7 Ctrl and 7 High risk female subjects. Therefore, the only other covariate aside from age is the risk group covariate (Ctrl - 0, High - 1) which we model as a boolean variable as specified.

Estimating longitudinal trajectories

Along with longitudinal image scans, the PREDICT-HD database consists of associated information such as subject id, age, group membership, sex, volumes of structures, etc.. To do longitudinal shape analysis with multiple covariates, we first prepare a parameter file consisting of the relevant covariate information along with extracted shape point files in correspondence. This is shown in Figure 1.

We highlight the meaning of each column in this file and how you can use/modify it for your own data analysis.

Subject Number	Group	Sex	Age	Shape File Path
50015	1	0	61.66461328	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50015/left_caudate_time 00.1
50015	1	0	63.18685832	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50015/left_caudate_time 01.1
50015	1	0	64.15605749	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50015/left_caudate_time 02.1
50352	1	0	61.87268994	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50352/left_caudate_time 00.1
50352	1	0	63.93976728	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50352/left_caudate_time 01.1
50567	1	0	58.09445585	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50567/left_caudate_time 00.1
50567	1	0	60.31485284	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50567/left_caudate_time 01.1
50855	0	0	59.56468172	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50855/left_caudate_time 00.1
50855	0	0	60.73100616	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50855/left_caudate_time 01.1
50855	0	0	61.5578371	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50855/left_caudate_time 02.1
50855	0	0	63.55099247	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50855/left_caudate_time 03.1
50855	0	0	64.44900753	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50855/left_caudate_time 04.1
50855	0	0	65.65639973	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50855/left_caudate_time 05.1
50983	0	0	55.2991102	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50983/left_caudate_time 00.1
50983	0	0	56.25188227	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50983/left_caudate_time 01.1
50983	0	0	57.25119781	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50983/left_caudate_time 02.1
50983	0	0	58.30527036	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50983/left_caudate_time 03.1
50983	0	0	59.41409993	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/50983/left_caudate_time 04.1
51034	1	0	56.11225188	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51034/left_caudate_time 00.1
51034	1	0	57.27036277	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51034/left_caudate_time 01.1
51034	1	0	58.05886379	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51034/left_caudate_time 02.1
51211	0	0	56.10403833	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51211/left_caudate_time 00.1
51211	0	0	58.0971937	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51211/left_caudate_time 01.1
51211	0	0	59.05270363	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51211/left_caudate_time 02.1
51211	0	0	60.07118412	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51211/left_caudate_time 03.1
51706	1	0	55.84668036	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51706/left_caudate_time 00.1
51706	1	0	58.001718	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51706/left_caudate_time 01.1
51706	1	0	58.88560473	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51706/left_caudate_time 02.1
51706	1	0	59.70466119	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51706/left_caudate_time 03.1
51855	1	0	58.89664613	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51855/left_caudate_time 00.1
51855	1	0	59.89322382	/home/prasanna/work/git/PHD Shape/tutorial/shape_lpts_for_stats/51855/left_caudate_time 01.1

Figure 1: Screenshot of parameter file needed for estimation of longitudinal trajectories. In chronological order, the columns are: (1) subject number (2) group (ctrl - 0, hd - 1) (3) sex (female - 0, male - 1) (4) age (5) shape file path

- For this tutorial, you can find these parameter files for each structure in the directory **param_files**.
- Column 1 represents the subject number as specified in the PREDICT-HD dataset. See that each successive instance of the same subject is a new time point whose associated independent variable (age) is in column 4.
- Column 2 represents group membership and is boolean for the tutorial example and as seen in the figure above. This is because our data only belongs to two classes - 0 (Ctrl), 1 (High). The larger PREDICT-HD corpus has 4 classes (Ctrl, Low, Med, High). Column 2 can be edited to represent all these classes as - 0 (Ctrl), 1 (Low), 2 (Med), 3 (High).
- Column 3 represents the sex variable. In the figure above, it is a string of zeros since we only have female subjects for the tutorial. In the general case for this data-set, we would have both males and females making this variable boolean - 0 (female) or 1 (male).
- Column 4 represents the independent regressor variable. For the tutorial, we choose age to be this variable. You could choose to have alternate regressors such as derived genetic measures like CAP score.
- Column 5 is the path to shape files in longitudinal correspondence obtained from running the script **write_vtk_to_lpts_script.py**. Make sure to edit these paths in the parameter file depending on where the shape point files are on your system.

Script arguments Ensure these paths are right for your system

- **paramfile_path**: the directory containing parameter files - we call it **param_files**
- **template_path**: path to a common template file for each structure - we choose one of the control subjects
- **output_path**: path to directory containing estimated fixed and random effects files for each structure - we call it **mixed_effects_results**
- **me_test_path**: path to executable which computes mixed-effects parameters to longitudinal data

Run the script:

- **python estimate_mixed_effects_script.py**

Once the script is run, output directories for each structure (left/right caudate/putamen) are created at the specified **output_path** and fixed and random effects estimates are stored in files called by the same name.

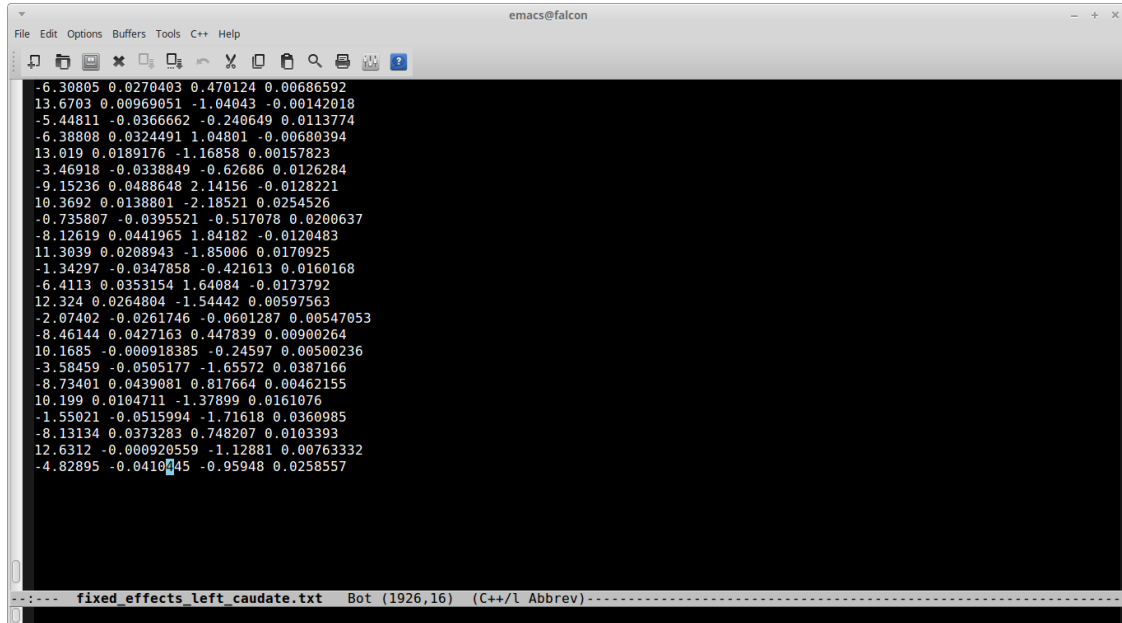


Figure 2: Screenshot of fixed effects parameter file estimated from given longitudinal data.

In Figure 2, see that there are 1926 rows. These rows correspond to each coordinate of each particle. A shape is represented by particles, which for us is 642 in number with x, y, z locations. These amount to $642 * 3 = 1926$ parameters. Also, see that there are 4 columns of numbers. These are associated to the number of categories our covariates determine. For example, we have two groups (Ctrl, High) of female subjects as our dataset. Since this is a linear mixed-effects model, we compute intercept and slope associated to both groups. So in all there are 4 types of parameters (2 for control, 2 for high). More elaborately,

- Column 1: Intercept α_0 for control group (this is the reference shape parameter at time 0)
- Column 2: Slope α_1 for control group (this is the reference evolution parameter depicting how the control group changes over time)
- Column 3: Difference α_2 between intercept parameter for control and hd groups. Adding α_2 to α_0 gives us the intercept for high risk group.
- Column 4: Difference α_3 between slope parameter for control and hd groups. Adding α_3 to α_1 gives us the slope for high risk group.

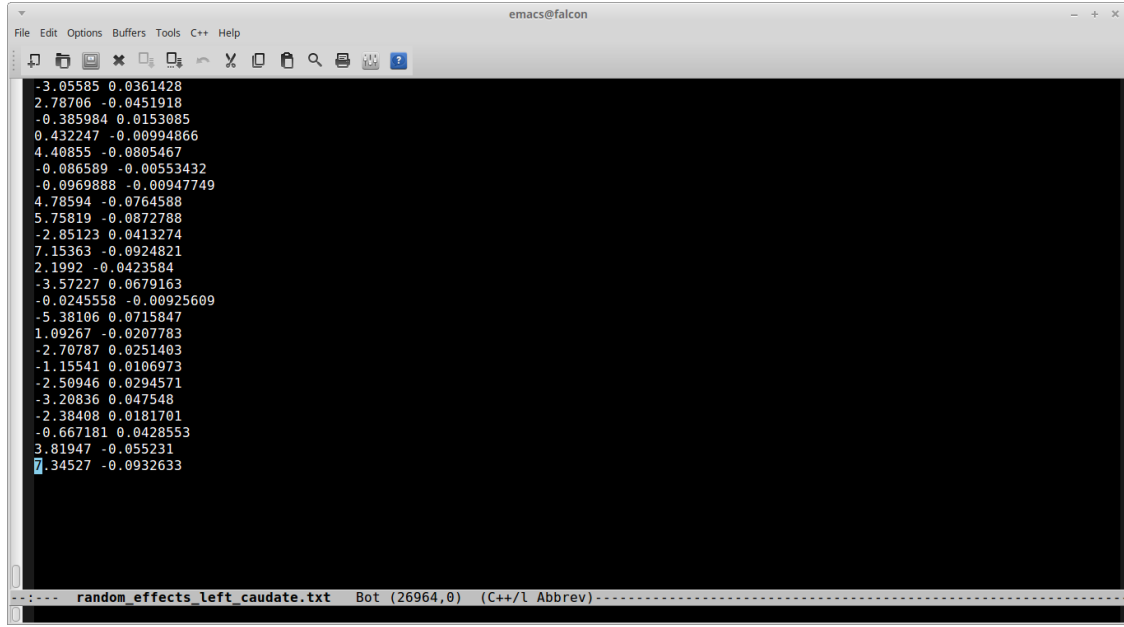


Figure 3: Screenshot of random effects parameter file estimated from given longitudinal data.

Figure 3 shows a screenshot of random effects parameters for this experiment. The estimates for random effects represent perturbation of each **individual's** shape model from a reference linear model for shape (in our case, control females). The two columns depict perturbations in intercept and slope respectively. The rows are $642 * 3 * 14$ in number, where $642 * 3$ is every coordinate of each particle and 14 is the total number of subjects.

Alternate scenarios and models:

- If suppose we had male subjects in addition to female subjects in our data, we would add on another covariate (which is boolean) in our model for the sex variable (therefore in the parameter file). This would add 2 more columns in fixed-effects parameters - difference parameter from reference (females) in intercept to model male subjects and a difference parameter for slope.
- If the data had more categories in terms of group membership (say group = *Ctrl, Low, Med, High* now has 4 options instead of 2), we would need 4 extra columns - two columns that

represent difference of intercept for low and med from reference, and two columns for difference in slope.

- Higher order polynomials instead of just intercept and slope.

So depending on input data information and model choice, you can set up your own parameter file which leads to a certain design for mixed-effects modeling, giving different group summaries in terms of fixed and random-effects parameters.

Visualizing shapes and trajectories

We visualize the estimated longitudinal trajectories as shape sequences evaluated at different time points in Paraview. The time axis is taken to be between min and max of the given age range. To do this, we first convert the estimated intercept-slope vectorized parameters to represent actual shapes and shape evolutions and store them in vtk format. When generating vtk files, we maintain the neighborhood and connectivity information of the original Deformetrica mesh.

- **input_path**: path to directory consisting fixed effects parameters - **mixed_effects_results**
- **vtk_template_path**: Choose a vtk template file to maintain the same connectivity.
- **output_path**: path to new output directory where we generate shape sequences for each structure - our folder: **output_shape_seq-vtk**

Run the script:

- **python create_shape_sequences_script.py**

The script also computes a local expansion/contraction measure to indicate how each structure evolves locally. This is computed as a dot product of the slope parameter and the outward normal at each point location of the shape. Here's a visualization of two different views of the left caudate

Create your own visualizations in Paraview

- To create a similar visualization as in 4, load the vtk files in Paraview created at the **output_path** location. This will be denoted by ctrl.0.vtk, hd.0.vtk.
- Load up two windows and link their cameras so as to see both groups in the same perspective. Turn on the "Expansion" scalar value to have local expansion / contraction measures overlaid.
- If you load up each category as two groups of shape sequences say Ctrl and High groups, you can see them as a video over time.

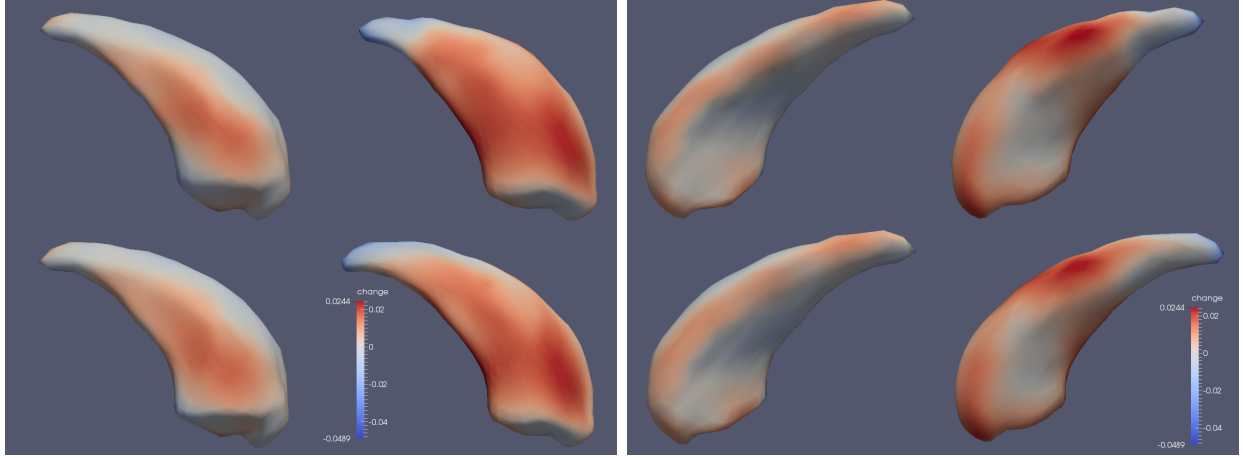


Figure 4: Two views (left and right) of left caudate fixed-effects estimates: CTRL (top), HIGH (bottom). Colormap indicates local expansion (red) and contraction (blue)

Hypothesis testing

We finally test the hypothesis for significant differences between Ctrl and High risk groups by a non-parametric permutation test using the Hotelling t^2 statistic. The input to this are the point correspondence files, and the output is a p-value. The tutorial does two different hypothesis tests

- Are there significant differences in intercept shape between Ctrl and High groups, and
- Are there significant differences in shape change (slope) between Ctrl and High groups over time.

As explained before, the first two columns of the estimated fixed-effects parameters represent intercept and slope of a reference category - Ctrl females in our case. The next two columns represent the difference between Ctrl and High groups in intercept and slope respectively. The question then becomes are the estimated parameters of columns 3 & 4 significant? The non-parametric permutation tests with the Hotelling t^2 statistic is explained in Section 3.4 in [1].

The script arguments are

- **paramfile_path**: the directory containing parameter files - we call it **param_files**
- **template_path**: path to a common template file for each structure - we choose one of the control subjects
- **output_path**: text file listing hypothesis results - we store it in **mixed_effects_results**
- **hypothesis_test_path**: path to executable which computes group hypothesis testing
- **outfile_path**: text file listing results of hypothesis

Run the script:

- `python group_hypothesis_testing_script.py`

The result of this experiment gives us a p-value indicating whether there is a significant difference between two groups of longitudinal data based on the intercept parameter and on the slope parameter. If $p < 0.05$, we deem it significant. Turns out for our tutorial example, we obtain significance between shape evolutions (slope parameter) of the right putamen. No other structure gives us significance. This could be attributed to low sample size since we only consider 7 subjects each.

References

- [1] M. Datar, P. Muralidharan, A. Kumar, S. Gouttard, J. Piven, G. Gerig, R.T. Whitaker, and P.T. Fletcher. Mixed-effects shape models for estimating longitudinal changes in anatomy. In *MICCAI STIA*, volume 7570, pages 76–87, 2012.
- [2] Nan Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.
- [3] P. Muralidharan, J. Fishbaugh, H.J. Johnson, S. Durrleman, J.S. Paulsen, G. Gerig, and P.T. Fletcher. Diffeomorphic shape trajectories for improved longitudinal segmentation and statistics. In *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2014.