Out-of-distribution generalisation presentation

The objective of this task/presentation is to create a regression dataset from a composition of predefined functions and use it to analyse the performance of different regressors in an out-of-distribution setup. The dataset should be generated using the following functions:

$$g_0(x_0) = \sin(x_0), g_1(x_0, x_1) = x_0 x_1, g_2(x, y, r) = \begin{cases} x & \text{if } r > 0, \\ y & \text{if } r < 0. \end{cases}, g_3(x_0, x_1) = x_0 + x_1,$$

You will need to perform the following tasks to prepare a presentation:

1. Create a function that is a composition of the g_i functions above, and make it as complex or simple as you think is needed. As an example you can use, $F_0(x_0, x_1, x_2, x_3) = g_2(g_0(x_0), g_1(x_1, x_2), x_3)$, but feel free to use anything that makes sense, and make liberal use of g_3.

2. Visualise the generated function using randomely generated inputs.

3. Create input space bias by excluding specific regions of the feature space (i.e. your x_i) from the training data. Use these excluded regions as test data to evaluate the compositional generalisation. Visualise appropriately.

4. Using any supervised learning (or otherwise) method of your choice, approximate the compositional function you created above. Assume that you do not have access to any of the g_i functions during learning (i.e., you cannot perform symbolic regression) and evaluate in the test region.

Prepare a 10 to 15-minute presentation summarising your results and explain your approach and methodology. Conclude by discussing how you would evaluate your setup against other methods on benchmark datasets.