

In the end, it is your own decision which data to use. If you want to use the latest dataset, use the v4 dataset (your first option). If you want to try things out and get started quickly you can use the legacy “v2” dataset, as it is less memory hungry. You could also start with the [example scripts](#). I haven’t tried to run them, but they are at least a source of some basic ideas, probably also which dataset to use.