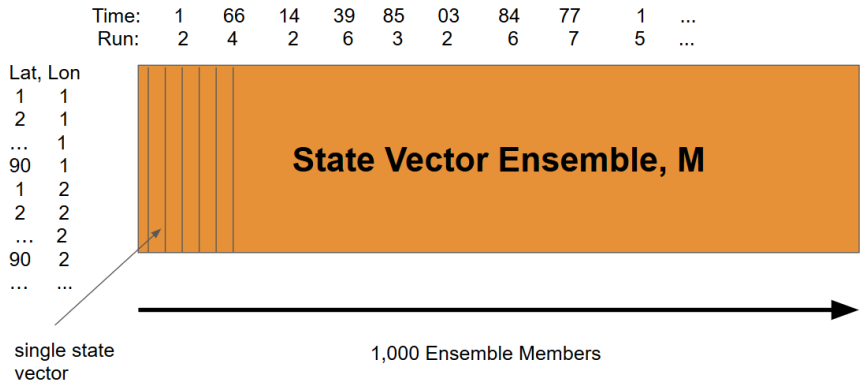
State Vector Design Documentation

1 State and Ensemble Dimensions

For gridded data in Dash, I differentiate between state and ensemble data dimensions. State dimensions are dimensions for which fixed indices are used in every state vector in a state vector ensemble. By contrast, an ensemble dimension is a dimension that uses different indices in every ensemble member in a stater vector ensemble.

For example, say that I have gridded surface temperature data, T. This gridded data has dimensions of lat x lon x time x run (where run is the model ensemble coordinate). Let's say that the time increments are monthly, and that my state vector consists of Northern Hemisphere temperatures in June. In this scenario, lat and lon are state dimensions. If I generate a 1,000 member state vector ensemble, every single state vector will be constructed from the data values at lat ¿ 0, and every lon. The lat and lon values at which we sample data are unchanging, so lat and lon are state dimensions. Separately, we will construct the state vector ensemble by selecting Augusts in random years (time) from random model runs (run). So, the first ensemble member might be June Northern Hemisphere temperature in year 15 of model run 2, while the second ensemble member is from year 77 of model run 5, and the third is from year 800 of run 1, etc. These time and run coordinates are different in each state vector, so time and run are ensemble dimensions.

I find it useful to visualize a state vector ensemble and the dimensional metadata associated with the ensemble.



Any dimension whose metadata is specified on each row of the ensemble is a state dimension. A dimension whose metadata is specified on each column of the ensemble is an ensemble dimension.

An important point is that state and ensemble dimensions are specific to each *variable* in a state vector, and NOT the entire state vector. This point is explored in greater detail in the state design demo.

2 Dimension Indices

There are four types of indices involved in the creation of state vectors in an ensemble. These are state indices, ensemble indices, sequence indices, and mean indices.

2.1 State Indices

State indices are relatively simple. For a variable in a state vector, these are the sampling indices for the state dimension. So, for our example Northern Hemisphere T state vector, this would be all lon indices, and lat indices for lat¿90. Note that different variables can have different state indices.

For example, we might wish to expand our T state vector to also include global precipitation rates in June, P. The P variable would use all lat indices, unlike the T variable, but lat is still a state dimension with state indices for each variable.

2.2 Ensemble Indices

Ensemble indices are the indices from which individual ensemble members are drawn. For our T example, the ensemble indices for the time dimension would be the index of any time point that occurs during June. The ensemble indices for the run dimension would simply be all the run indices.

2.3 Sequence Indices

Now, there may be cases where we want a sequence of data values that follow a trajectory through an ensemble dimension. For example, instead of wanting T in just June, we may want T in June, July, and August. We don't want to select any random July and August, but rather the July and August that immediately follow a particular June. Thus, we need to specify the sequence of indices that follow a particular ensemble draw. For this example, the sequence indices would be [0, 1, and 2], that is the time point 0 indices removed from a random June, the point 1 index removed from the random June (so, July), and the point 2 indices removed from the random June (August).

Note that it would be INCORRECT to simply set the ensemble indices to any index in June, July and August. That syntax would indicate to draw each state vector from the points in either a random June or a random July

or a random August. The sequence indices are required to specify a trajectory through the ensemble dimension.

Note that it would also be incorrect to set the ensemble indices to June, July, and August, and then use a sequence index of [0, 1, 2]. This syntax would indicate to draw a random time point from June, July, or August and then get a 3-month time trajectory. So the possible state vectors would be June-July-August, July-August-September, or August-September-October temperature data. This emphasize an important point:

Ensemble indices should only indicate the *start* of a sequence. They should not specify the collection of allowed values in the sequence.

2.4 Mean indices

Mean indices indicate a sequence of indices over which to take a mean in an ensemble dimension. So, if we wanted the mean of June, July, and August, rather than individual J, J, and A data points, we would specify mean indices as [0, 1, and 2] rather than sequence indices.

Mean indices are applied at each element in a sequence. This allows the construction of state vectors that are sequences of means. For example, instead of wanting mean T during a random JJA sequence, we might want a three year trajectory of JJA mean T. (Perhaps we're looking at a multi-year effect of volcanic eruptions). So, the JJA mean in year 1, JJA mean in year 2, and JJA mean in year 3. To set this up, we would set the ensemble indices to a random June, the sequence indices to [0 12 24], and the mean indices to [0, 1, 2]. This syntax would say:

Draw a random June. Advance 0 time indices past this June (so, year 1, still in June). Take the mean over the time points [0, 1, and 2] indices from this point (so, June, July, and August of year 1). Then, advance 12 time points from the initial draw (so, year 2 June) to get the next sequence member. Take the mean over the time points 0, 1, and 2 indices from this year 2 June (so, June July and August of year 2). Finally, advance 24 time points from the initial draw to get the last sequence member (year 3 June). Take the mean over the time points 0, 1, and 2 indices from this year 3 June (so, June July and August of year 3). Combine the three summer means to create one state vector. Draw a different random June and repeat the process to create multiple state vectors.

3 Coupled Variables

Often, if a state vector has multiple variables, the variables will be realizations of the same climate instance. So, if we had a state vector composed of T in June and P in June, we would want to use T and P data from the same time step in the same run. When variables need to be drawn from the same indices of ensemble dimensions, then the variables are "coupled" variables. Coupled variables ALWAYS share the same ensemble indices and are always drawn from the same ensemble indices in individual state vectors.

To facilitate state vector design, coupled variables can also optionally share the same state, sequence, and mean indices. This functionality is explored in greater depth in the state design demo.

4 User Functions

The following is a summary of functions available to the user for state vector design:

stateDesign: This creates a new state vector design.

addVariable: This adds and initializes a new variable in a state vector design.

editDesign: This edits the state, ensemble, sequence, and/or mean indices of a variable in a state vector design.

coupleVariables: Couples ensemble indices of variables in a state vector design. Optionally couples state, sequence, and/or mean indices.

uncoupleVariables: Uncouples variable indices.