Thoughts and Questions from recent SINDy Lit Review

Hybrid dynamical systems (Mangan et al. 2018):

* How do you choose the discretization of models? What counts as the "same" model for the purpose of evaluating frequency?
* Contamination of clusters near discontinuous transition point.
  + Dynamic cluster sizes (smaller cluster sizes near transition point) can help narrow down the window for where the transition is, but no matter how small your cluster is, the cluster at the exact edge of a discontinuous transition point will always have mixed cluster membership. This is the place where you can't "see" a model because of cluster contamination.
  + Alternately, you could probe behavior at transition point more precisely by paying attention to cluster radius, i.e. what space/time domain are your clusters pulling from? Similarly, you could specify clusters by radius instead of by number of members.
* How do you connect discrete state-space models? What sorts of early warning signs can you look for in this context? Can you specify cluster resolution such that the AICc of the identified models gradually get worse as cluster membership transitions from one state to the other across the state-space boundary? This probably depends on the underlying system dynamics.
* What of path trajectories vs. multiple initializations? What of hybrid chaotic systems? How does that interact with this method?
* How do you know you need to do clustered model identification in the first place? Is there a characteristic relationship of relative AICc to cluster size that indicates optimum cluster size? You could then plot AICc characteristic over the state-space to identify heterogeneous regions of optimum cluster size.

Model selection via sparse regression and information criteria (Mangan et al. 2017):

* SINDy can't find the correct model when a redundant variable is included in the input time series. Is there a way to deal with this if you don't know what variables might be redundant? You can test your time series before SINDying, but sufficient noise would obscure the dependence, yes? Is there a way to build in a failure indicator for SINDy so you know that your SINDy identification is not valid?
* Chaos muddies the power of SINDy to identify the correct model (it is not possible to predict beyond a certain time horizon). You can overcome this by cross-validating on a sufficiently short characteristic time scale. But in real life, you wouldn't know what that is or even that your system is chaotic and time scale is important. You need a sensitivity analysis. You could iterate up from a very short time scale for cross-validation time series to get an indicator of characteristic time scale. Look for where the model ranking by information criteria starts to fluctuate as cross-validation time scale increases.