Predictability, Probability and THORPEX

Leonard Smith

Centre for the Analysis of Time Series,

London School of Economics

Pembroke College, Oxford

Mark Roulston, Devin Kilminster, Kevin Judd,
Liam Clarke, Jochen Broecker

Aims:

The aim here is strategic: how can we make “better” weather forecasts in 2014.

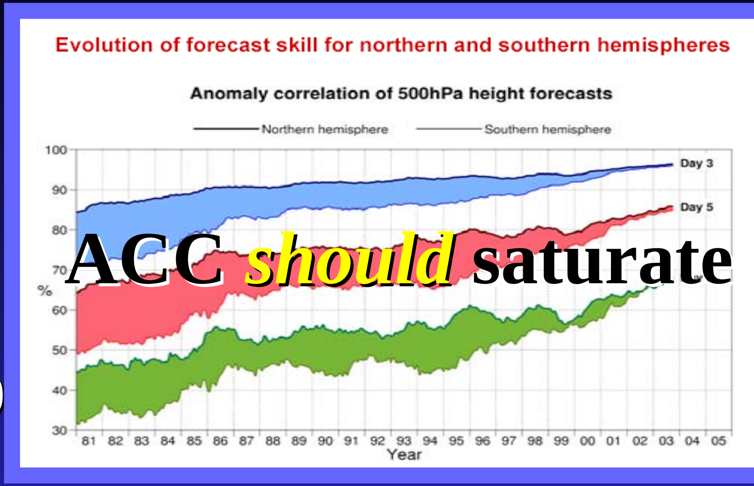
Tactical issues of making better forecasts in 2005 were addressed by Altalo and Gordon already, and by Clarke, Broecker and Kilminster later today.

? Evolution or Revolution?

Take home point:

What are the implications of:

- Taking Ensembles Seriously?
  TIGGE (mIC,mM,mN, ?mA?):
  - **NO**: RMSE, MAE,…(of what?)
  - "Empirically proper" scores
  - empirical verification (for TIGGE?)



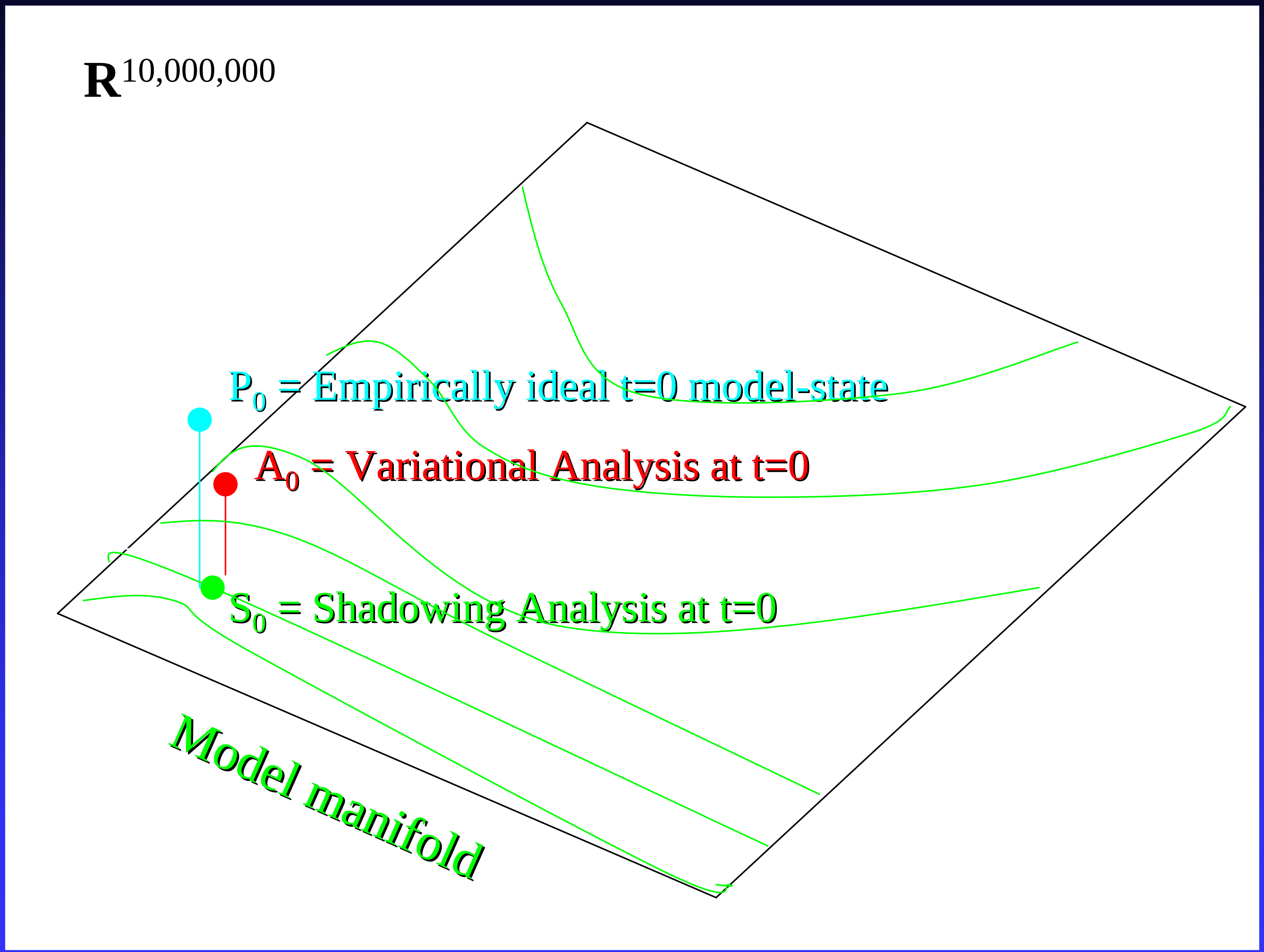
From Holingsworth, et al. 2002

- Taking Model Inadequacy Seriously? *(intrinsic uncertainty)*
  - translating model-simulations into weather-forecasts
    - Information content, not literal interpretation
  - accepting that perfect model studies are misleading
  - preferring initial conditions that "look" worse
- Adaptive observations in the mM, mIC, mN context:
  ? Based on TIGGE?

"WELCOME TO THE REVOLUTION"

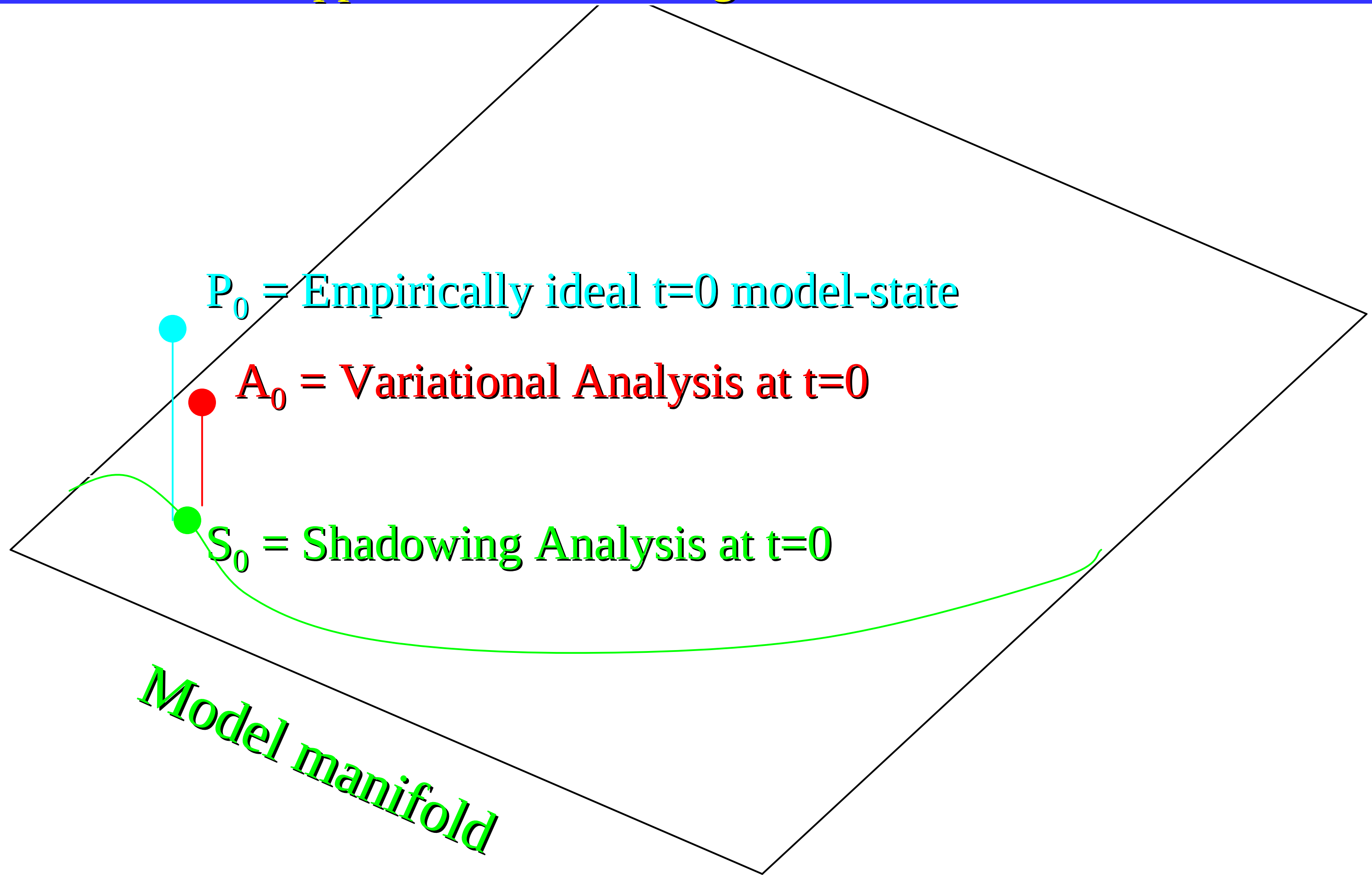
Rage Against the Machine

THORPEX Montreal 2004


$\mathbf{R}^{10,000,000}$
P_0 = Empirically ideal t=0 model-state
A_0 = Variational Analysis at t=0
S_0 = Shadowing Analysis at t=0
Model manifold

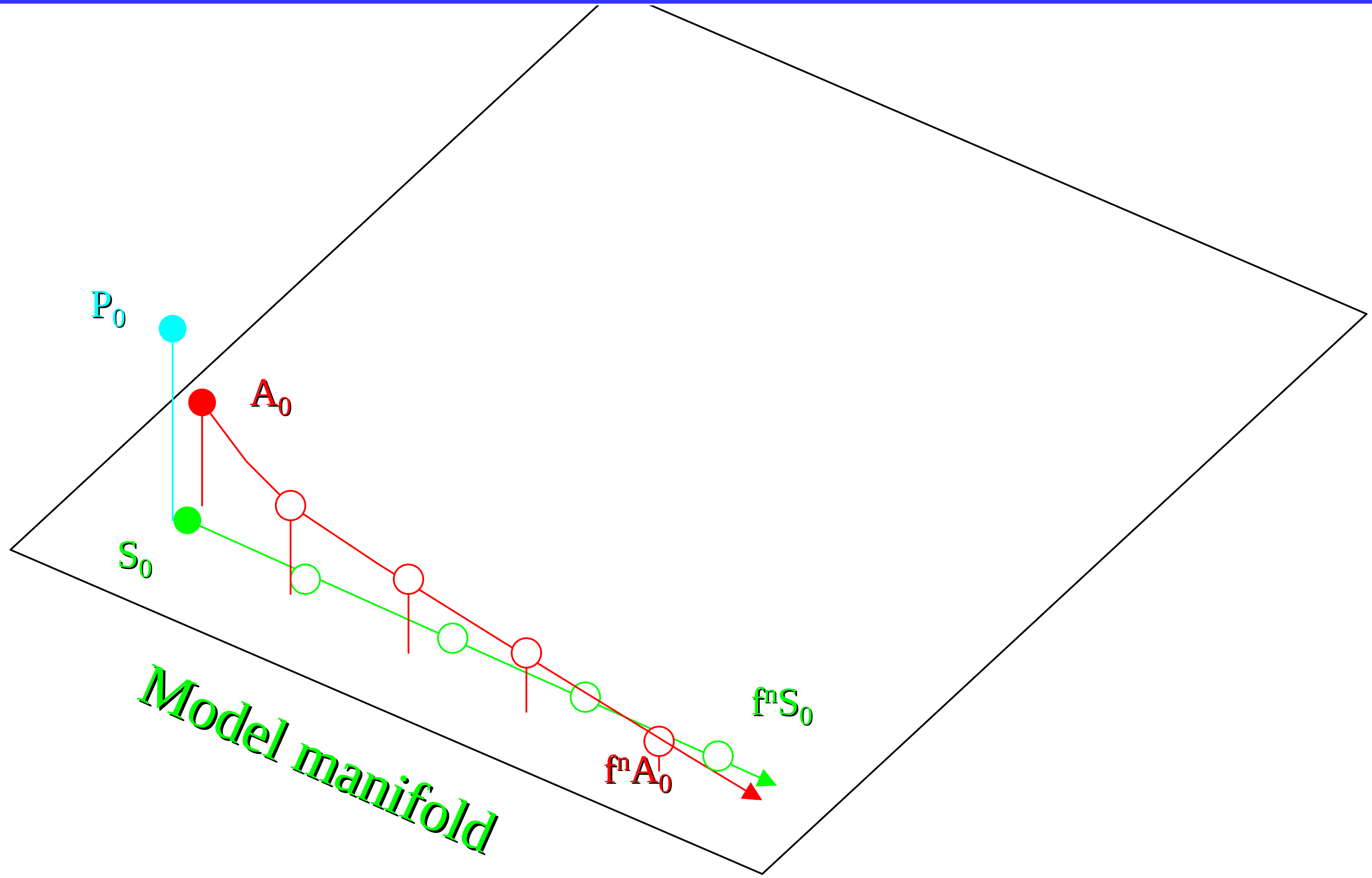
We might keep P_n as a target/verification,
but P_0 is unlikely to provide model-initial condition(s).

Variational Assimilation pulls the initial conditions away from the manifold.
What happens when we "let go" and forecast…

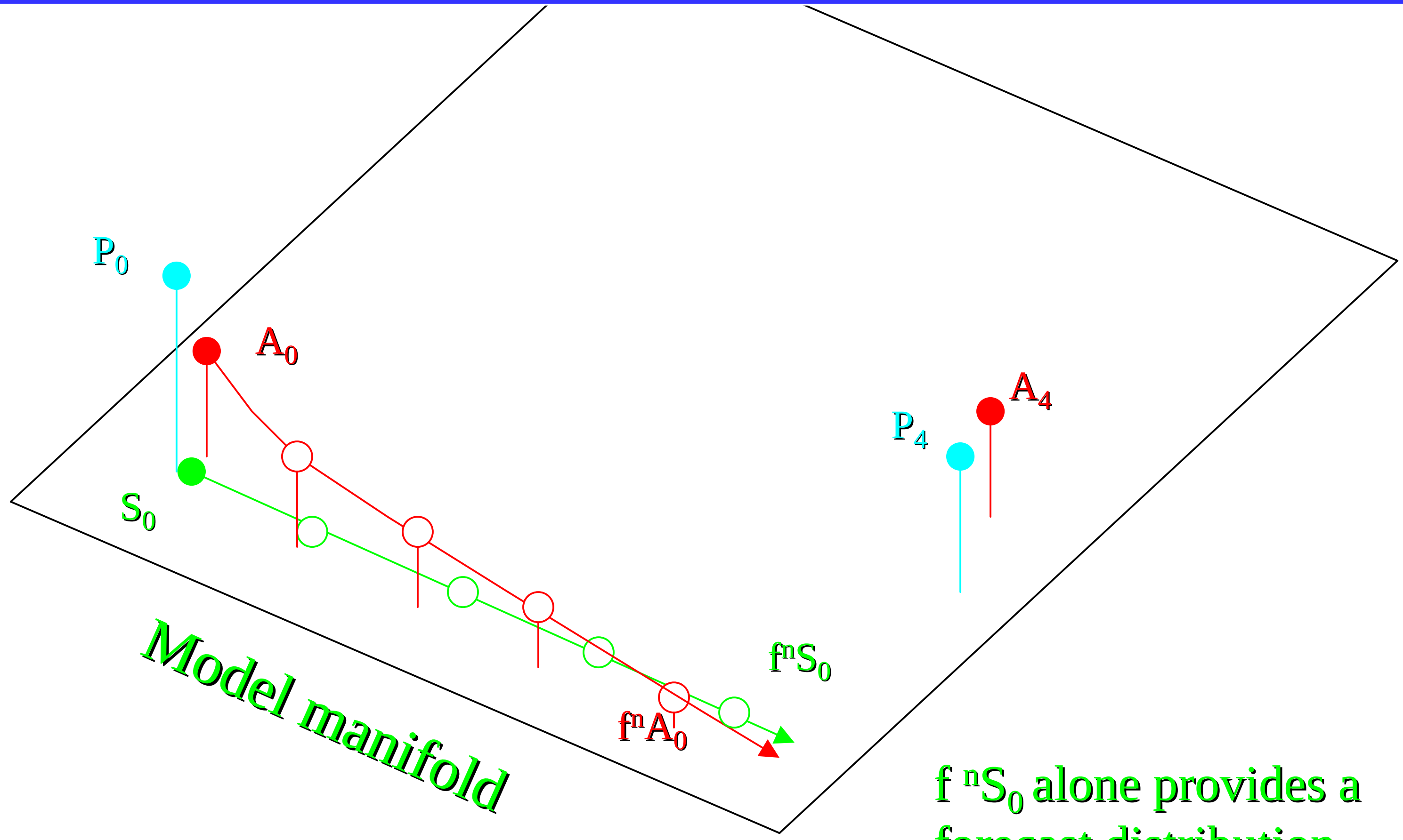


What happens when we "let go" and forecast…

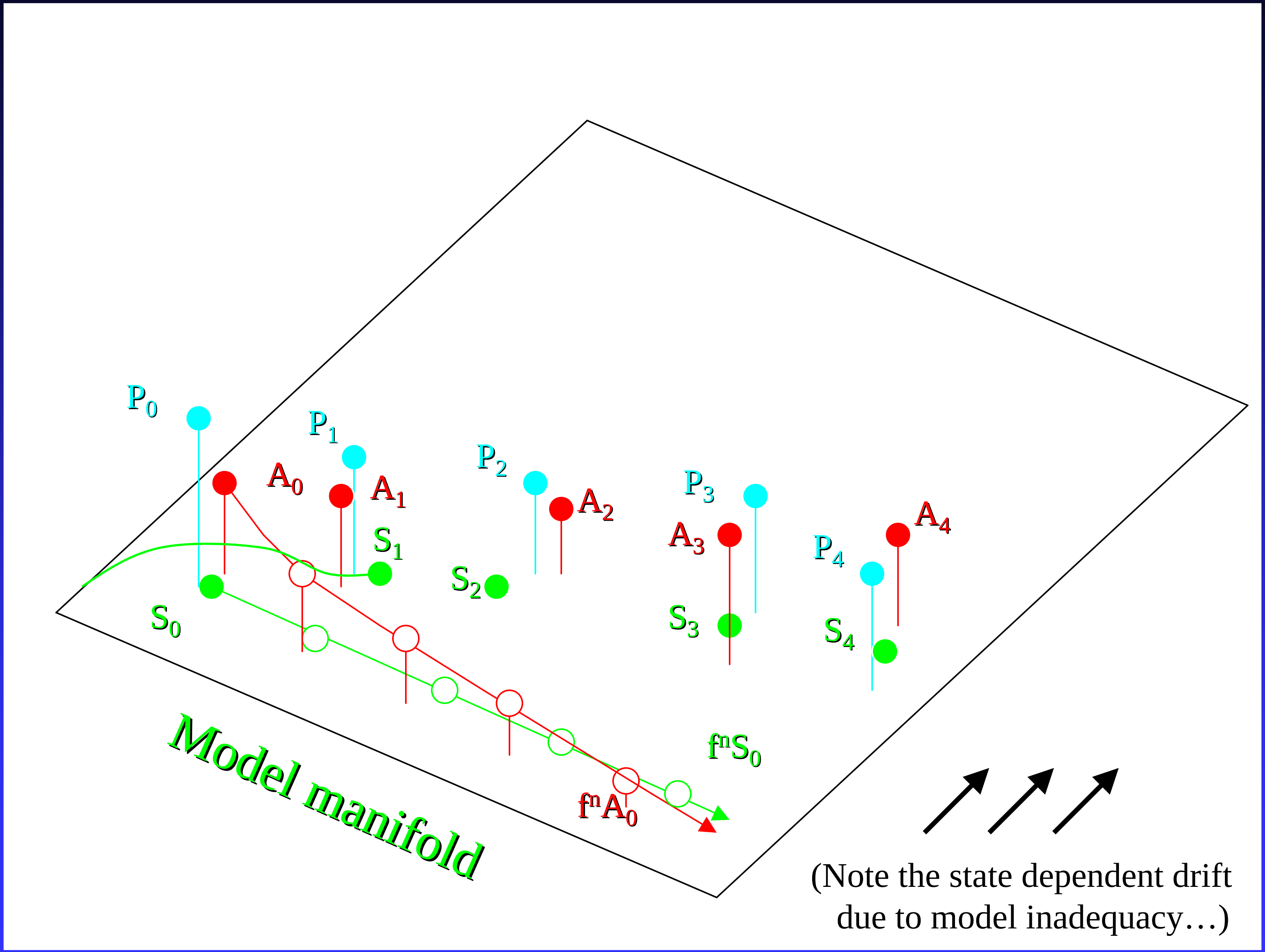
A_0 immediately falls toward *somewhere* on the manifold.



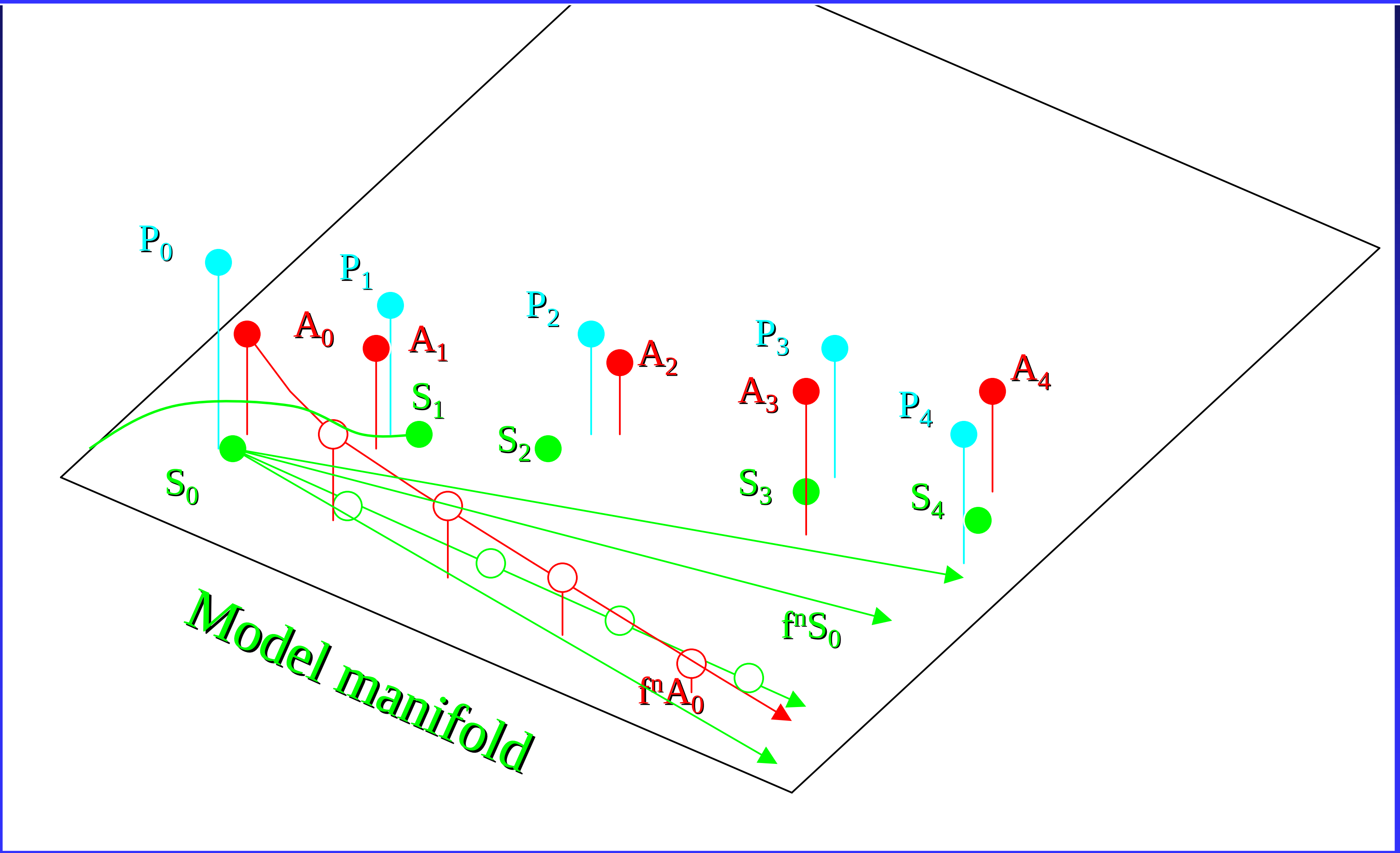
We are allowed a projection operator to map $f^{\,n}S_0$ into a distribution; we take this freedom even if we verify against P!



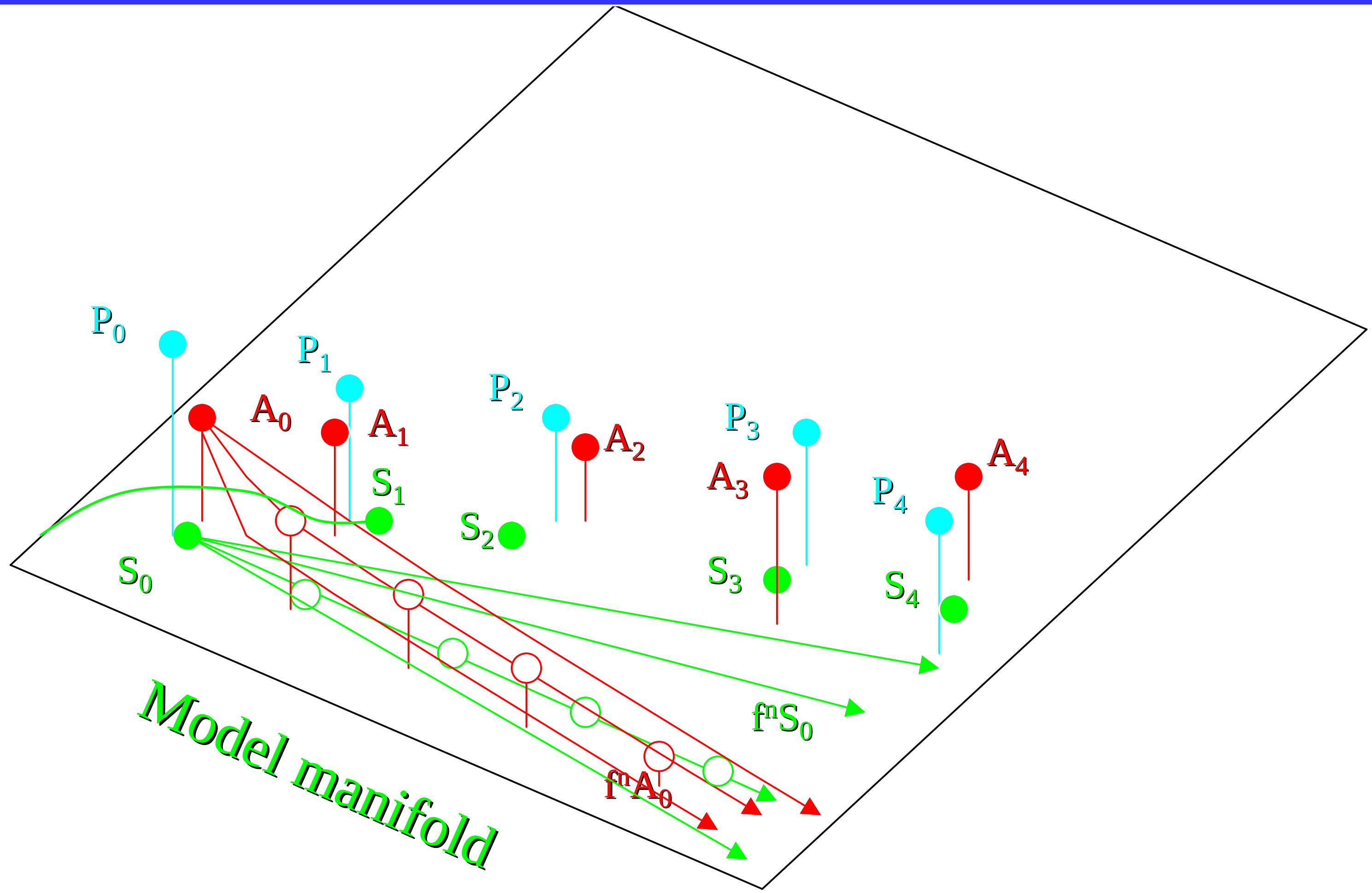
$f^{\,n}S_0$ alone provides a forecast distribution


P0
P1
P2
P3
P4
A0
A1
A2
A3
A4
S0
S1
S2
S3
S4
fnS0
fnA0
Model manifold
(Note the state dependent drift
due to model inadequacy…)

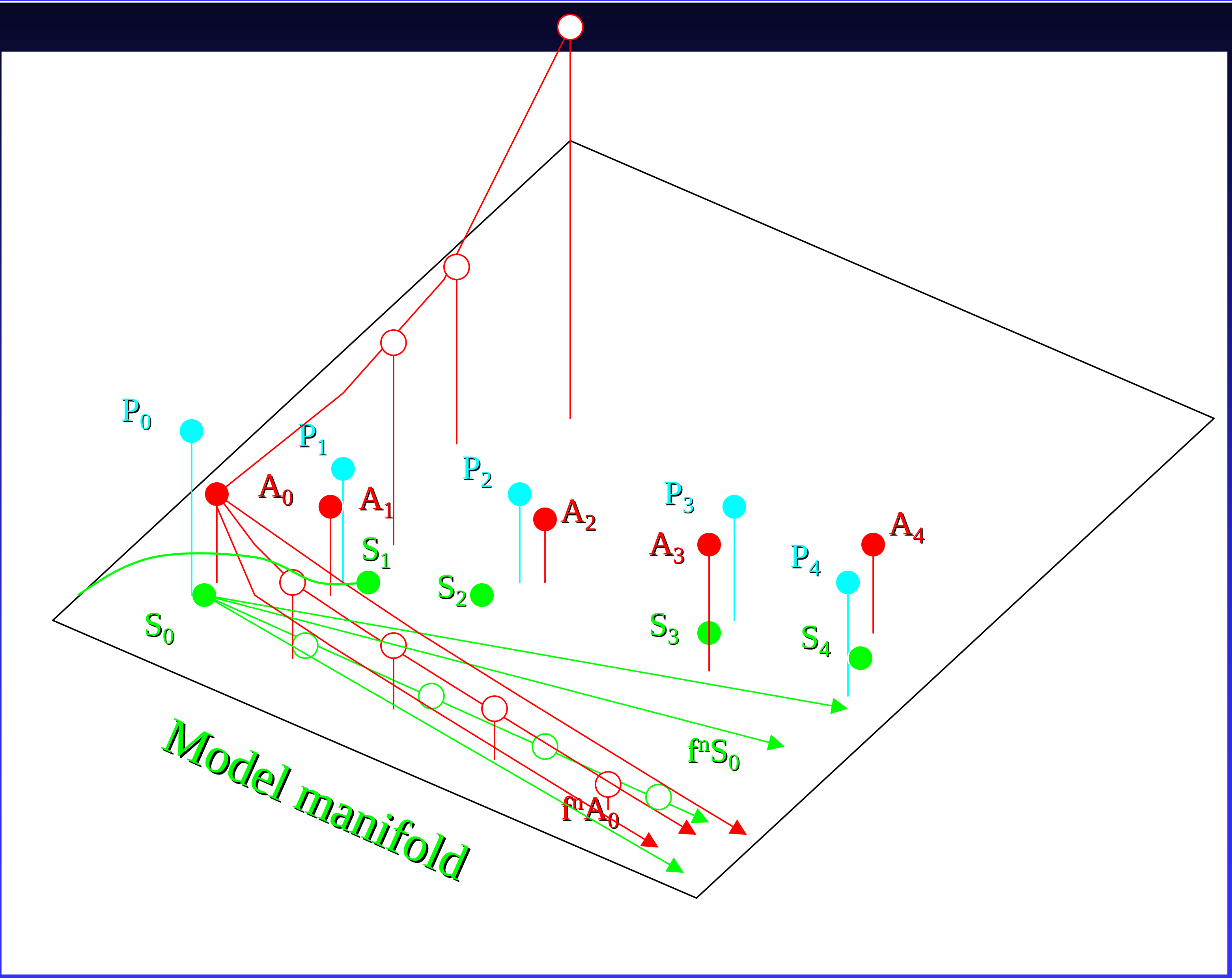
But if we have taken ensembles seriously then
we have an ensemble of simulations from near S.



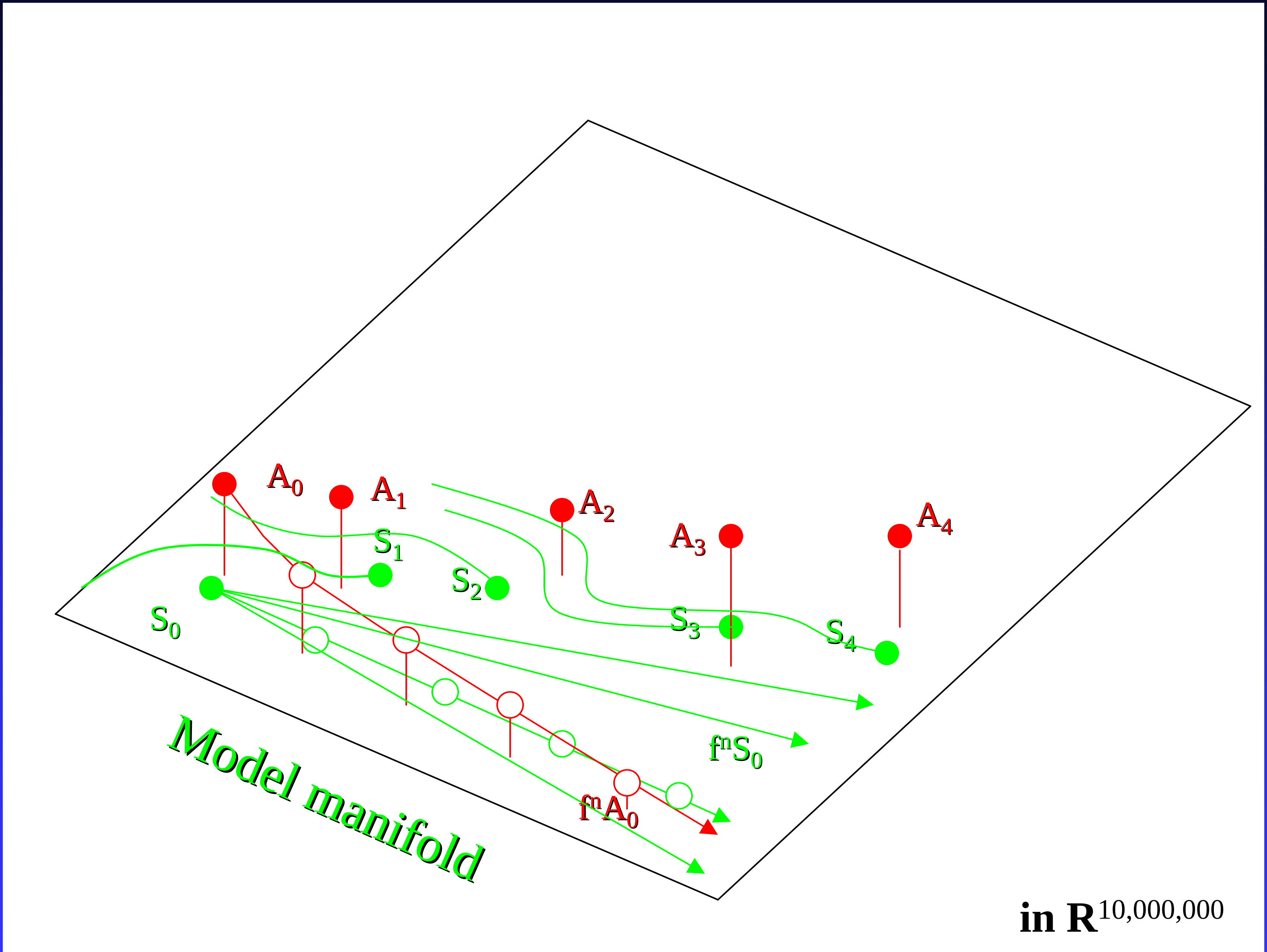
And an ensemble of model simulations from near A.



Of course, points near A can fall onto other bits of the manifold.



What can we know operationally?



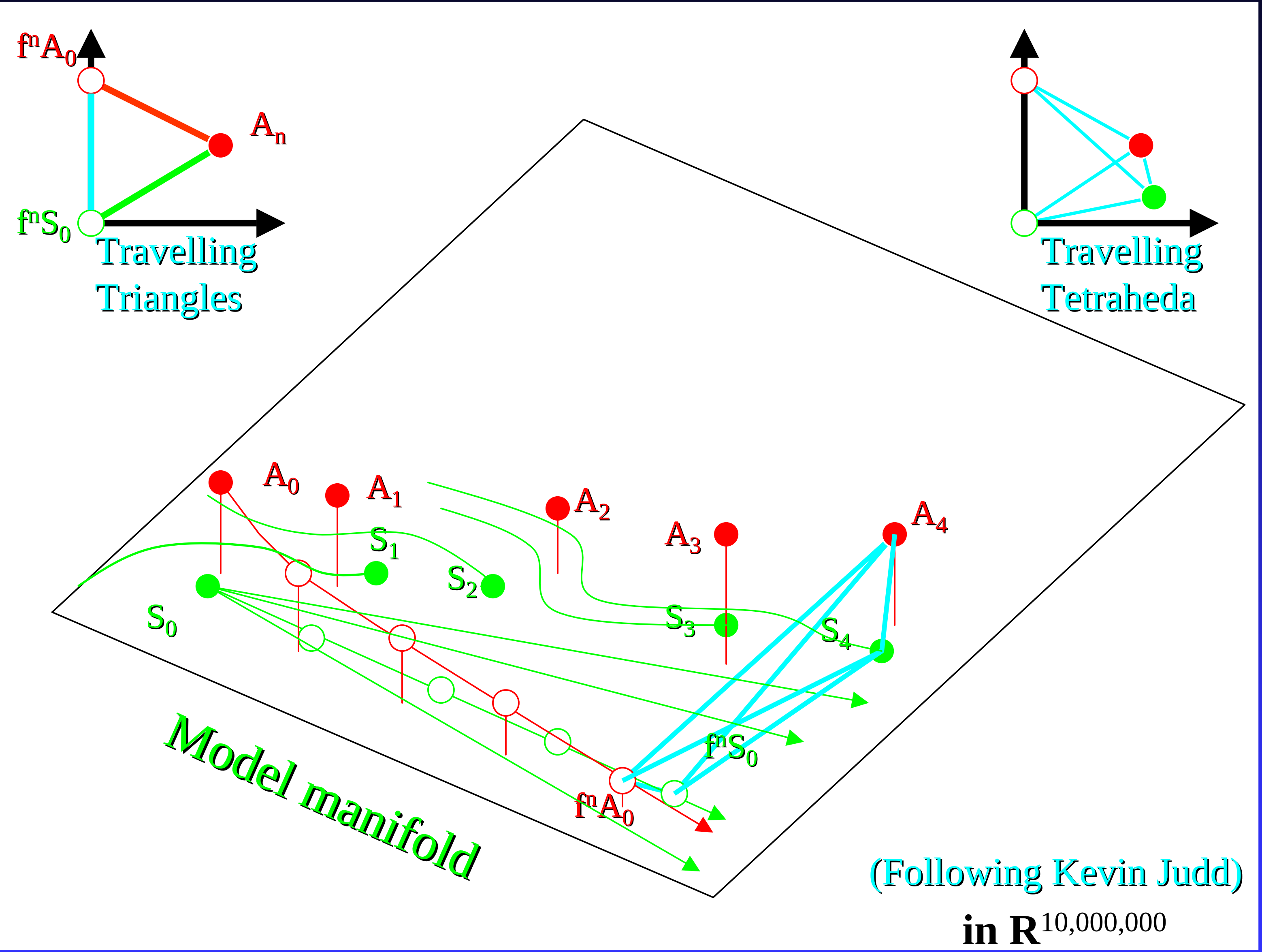
The empirically relevant questions are (include):

Is it better to pull A_0 away from the model manifold or to project S_n back into the obs space?

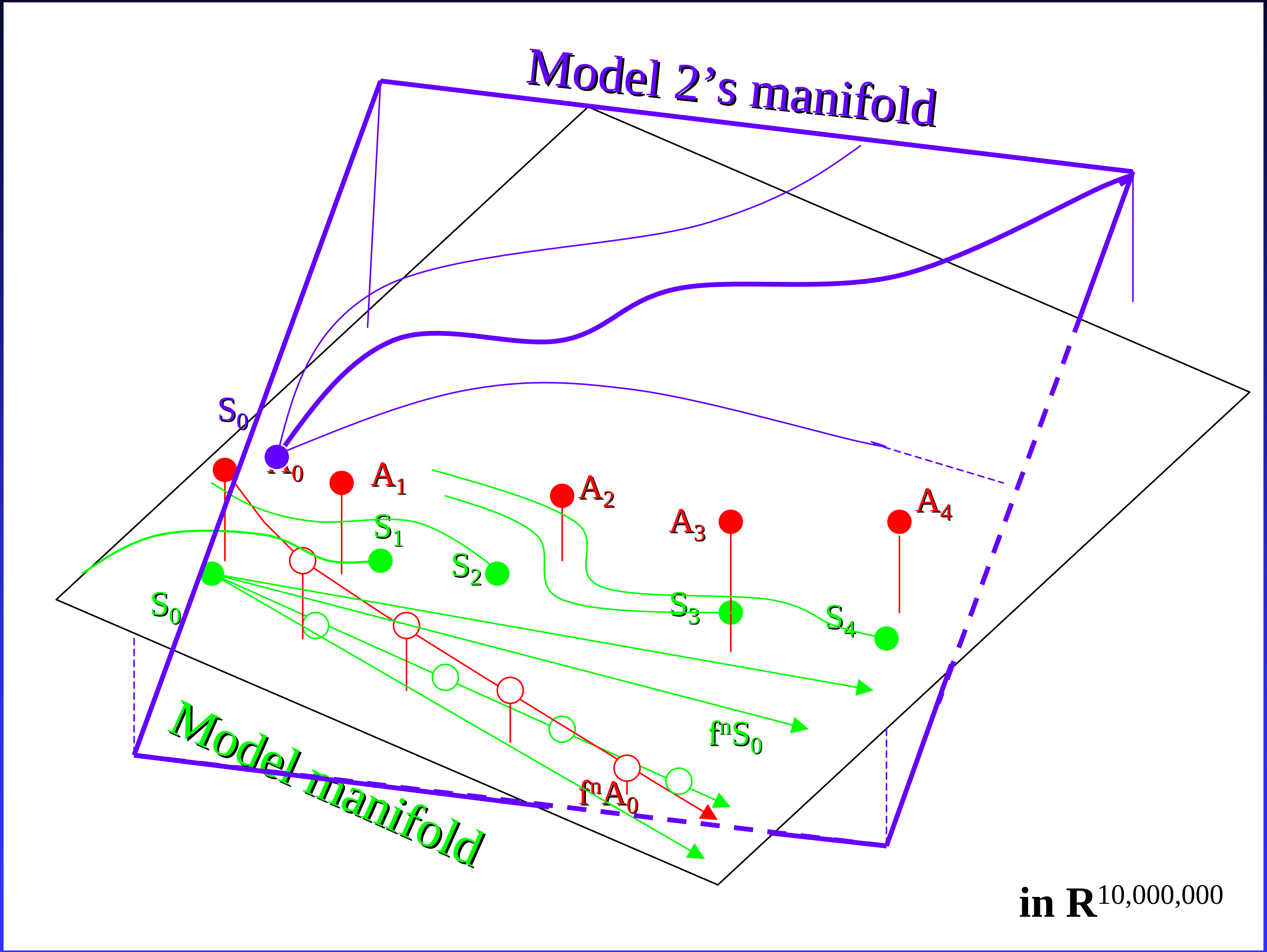
How would ensembles on/near the model manifold compare to ensembles from perturbed variational analyses?

For other physical systems, taking initial conditions on the model manifold seems to be better; for NOGAPS, Judd et al have suggested that this yields no RMS penalty for a single model run after day two:

But could we ever interpret such diagrams operationally?



What does a multiple model (structure) add?



The empirically relevant questions are (include):

Is it better to pull A_0 away from the model manifold or to project S_n back into the obs space?

For other physical systems, taking initial conditions on the model manifold seems to be better; for NOGAPS, Judd et al have suggest that this yields no RMS penalty after day two: how would ensembles on/near the model manifold compare to perturbed variational analyses?

In the multi-model case, we need to look at information content in an empirically meaningful space (obs):
Are better forecasts obtained by interpreting ensemble members as a collection of plausible weather scenarios? or by conditioning on the joint distribution of simulations (multi-IC and multi-model)?

Should we advertise “probability forecasts” or merely “probabilistic forecasts”?

Is it rational to expect probabilities from operational EPS(s)?

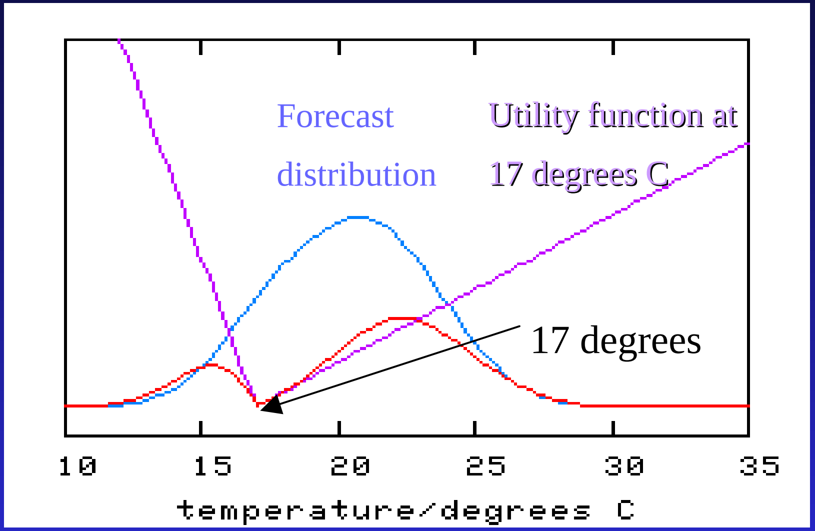
Given a forecast and a cost function, we can calculate expected cost of playing every likely temperature. (here, 17 degrees)

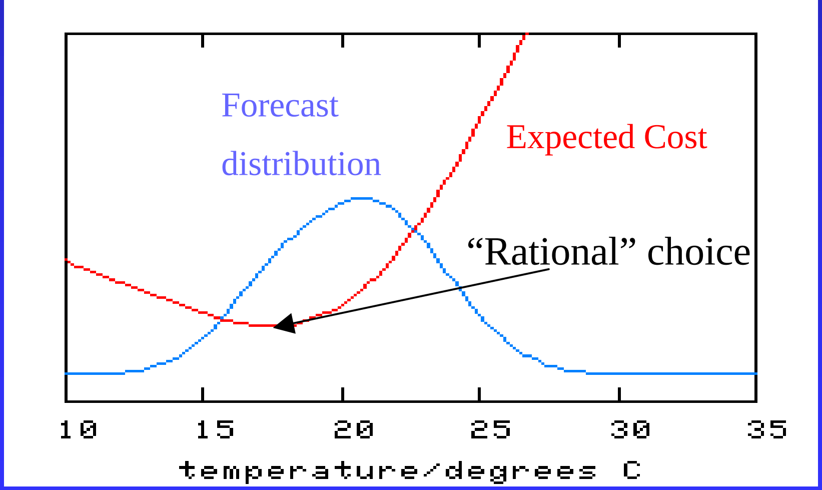
To maximize expected utility, we "should" act on the temperature with smallest expected cost.

For Cal ISO, it proves better to play a empirical quantile: and this is rational, *unless* we insist that the forecast distribution is a probability-DF.

Anyone have a counter-example?

Of course, this can be recast as merely a problem of robust estimation.





The empirically relevant questions are (include):

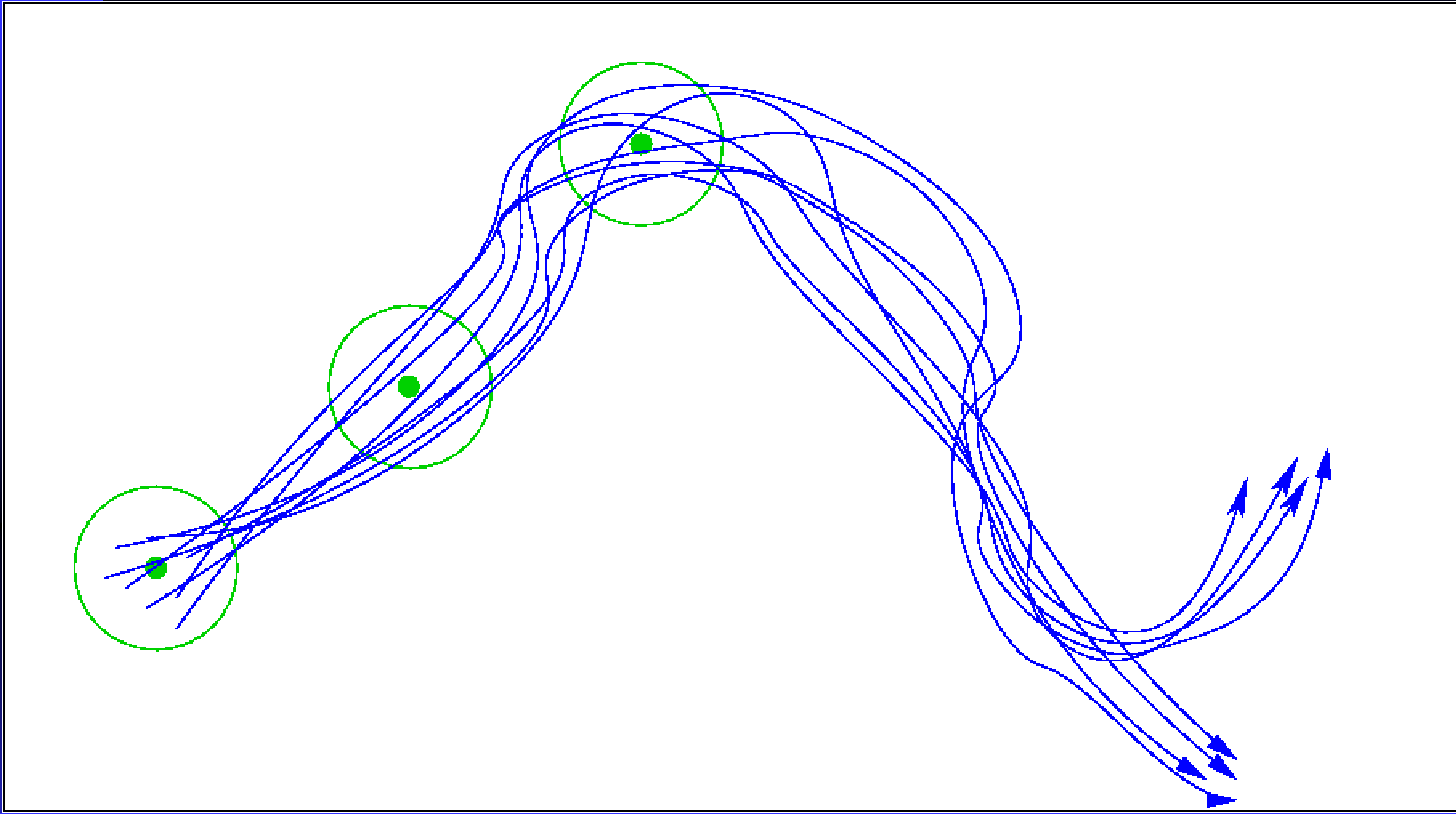
Is it better to pull A_0 away from the model manifold or to project S_n back into the obs space?

For other physical systems, taking initial conditions on the model manifold seems to be better; for NOGAPS, Judd et al have suggest that this yields no RMS penalty after day two: how would ensembles on/near the model manifold compare to perturbed variational analyses?

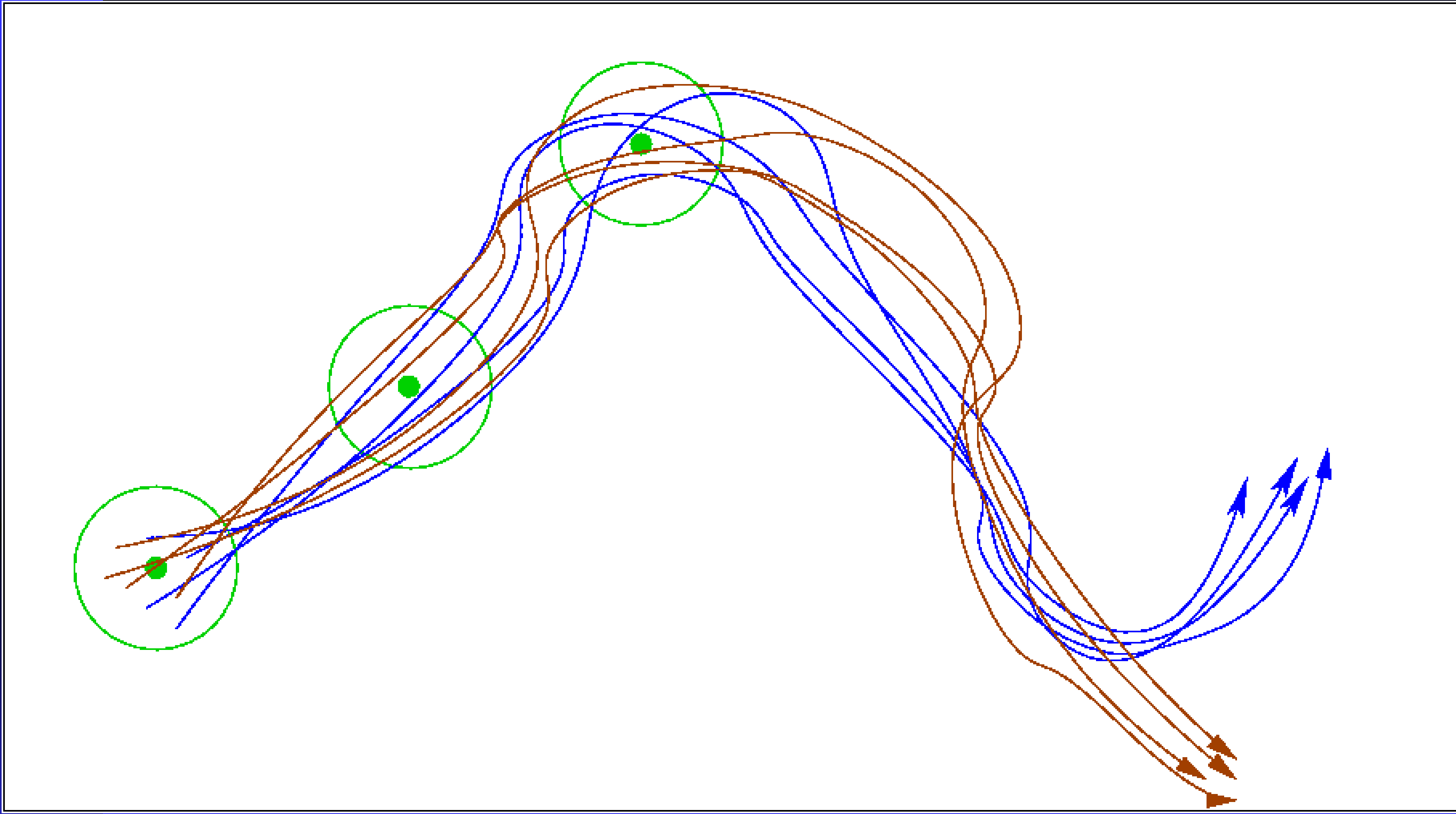
In the multi-model case, we need to look at information content in an empirically meaningful space (obs).

Are better forecasts obtained by interpreting ensemble members as a collection of plausible weather scenarios? or by conditioning on the joint distribution of simulations (multi-IC and multi-model)?

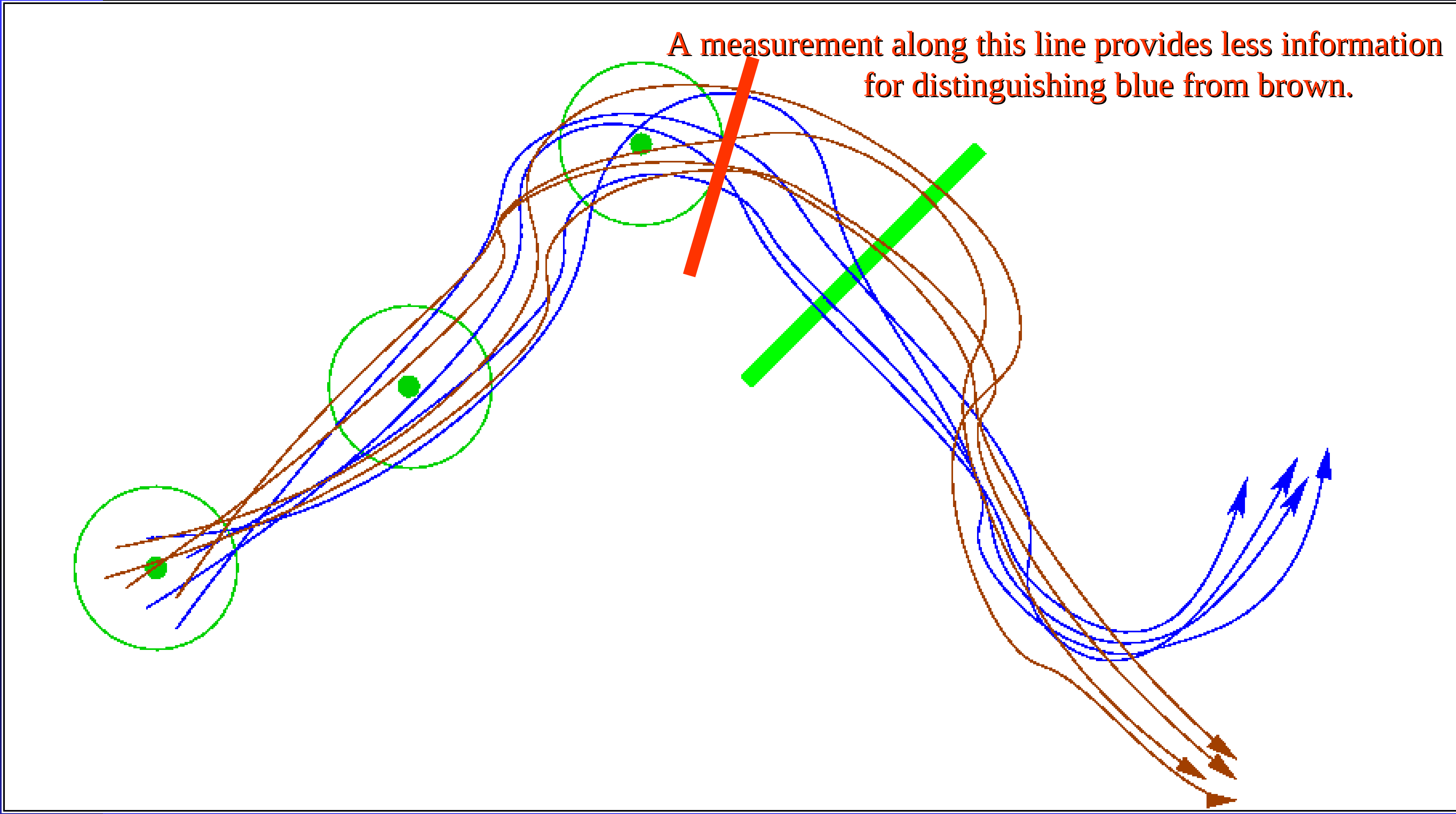
How might adaptive obs be taken in based on TIGGE?



Suppose we wish to distinguish two sets of simulations (say, storm/no storm); in terms of indistinguishable states, the AO question is simply “Which observations are most likely to separate these sets?”



To do this, merely color the trajectories in each set, and determine the observation in space and time (post ‘now’) that is likely to yield the most relevant information.



No linearization,
No implicit perfect model assumption,
And the ability to update the AO in light of scheduled obs without rerunning the simulations.

So (by 2014):

Forecasts may be better *because* the initial conditions look worse.

Forecasts will be improved by better resolving the projection from the simulation(s) to forecast (Moving further from the identity operator toward conditioning on the joint distribution of all simulations).

Probabilistic forecasts may prove more valuable by not providing probabilites.

Data assimilation will produce an ensemble of initial conditions, each “on” the model manifold.

Multi-model adaptive obs will be more straightforward by working in model-state space(s) with large TIGGE-like ensembles.

LA Smith (2003) Predictability Past Predictability Present. ECMWF. soon to be in a CUP book (ed. Palmer).

LA Smith (2000) *Disentangling Uncertainty and Error*, in Nonlinear Dynamics and Statistics (ed A.Mees) Birkhauser.

K Judd and LA Smith (2001) *Indistinguishable States I* , Physica D **151**: 125-151 *(2004) Indistinguishable States II,* **196:** 224-242 .

M. Altalo and LA Smith (2004) Environmental Finance **6** (1) 48-49.

M Roulston *and LA Smith* (2003) MWR **130** (6): 1653-1660.

A Weisheimer, L.A.Smith and K Judd (2004) A New Look at DEMETER forecasts via Bounding Boxes Tellus (to appear).

LA Smith (2002) *What might we learn from climate forecasts?*, Proc. National Acad. Sci. **99**: 2487-2492.

www.lsecats.org lenny@maths.ox.ac.uk

Eight Current Challenges:

Moving Beyond Scenarios
$P(\boldsymbol{x} \mid \text{obs}; X_1, X_2, X_3; Y_1, Y_2; Z_0)$

Dressing individual ensemble members may be useful, but a better (& more Bayesian) approach would be to condition on the joint distribution of our (imperfect) models.

Projection Operator –or- Ensemble "Bias Removal"

We do not really understand how to map (individual) model states to and from observational space, much less ensembles.

(Coelho et al 10:00 Thursday)

Parameter estimation in nonlinear models

Even with Normal input errors, nonlinearity implies non-normal output errors, complicating not only "state" estimation but also parameter selection.

"Recalibration"

Unlikely in meteorology

von Mises (1928)

Current Challenges:

Limited relevance of the Kalman Filter	"Of course, in general these tasks (prediction, separation, detection) may be done better by nonlinear filters." (Kalman, 1960; first substantial footnote)
Use of 4DVar with imperfect model(s)	The target is no longer a max likelihood state, in fact the model may not support the most "realistic" looking states.
Ensemble "spread" and "bias" correction.	Distinguishing "good spread" and "bad spread" given ~ 100 points in a ~10,000,000-dim space.
Interpreting parametric uncertainty in the "one-off" case (climate).	What are "reasonable" parameter ranges? How climate variables differ from weather? Can a prior distribution and a transfer function yield a policy relevant PDF?

Applications in this Context

Model development	resource distribution for utility (not for naïve realism)
Parameter Estimation	relaxed (to within the physical relevance of then parameterisation)
Data Assimilation	allow each model its manifold, assimilate without re-simulating!
(Ensemble) Simulation	perturb as far in the past as possible: do NOT resample
Forecasting:	true eMOS
Informed Decision Making	a PDF, but not as we know it
Model(s) Improvement	evaluation & forecast archive

Aim first for mere internal consistency?