Improved subseasonal prediction of South Asian monsoon rainfall using data-driven forecasts of oscillatory modes

Eviatar Bach, V. Krishnamurthy, Safa Mote, Jagadish Shukla, A. Surjalal Sharma, Eugenia Kalnay, and Michael Ghil

7 June 2024

California Institute of Technology

E. Bach, V. Krishnamurthy, S. Mote, J. Shukla, A. S. Sharma, E. Kalnay, and M. Ghil (2024). **"Improved Subseasonal Prediction of South Asian Monsoon Rainfall Using Data-Driven Forecasts of Oscillatory Modes".** *Proceedings of the National Academy of Sciences*

E. Bach, S. Mote, V. Krishnamurthy, A. S. Sharma, M. Ghil, and E. Kalnay (2021). **"Ensemble Oscillation Correction (EnOC): Leveraging Oscillatory Modes to Improve Forecasts of Chaotic Systems".** *Journal of Climate*

Introduction

Motivation

There are oscillatory modes in the climate system important on subseasonal-to-seasonal timescales, such as the monsoon intraseasonal oscillation (MISO) and the Madden–Julian oscillation (MJO).

Motivation

There are oscillatory modes in the climate system important on subseasonal-to-seasonal timescales, such as the monsoon intraseasonal oscillation (MISO) and the Madden–Julian oscillation (MJO).

Often predictable from data beyond current dynamical models.

Motivation

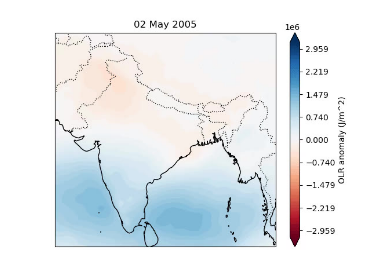
There are oscillatory modes in the climate system important on subseasonal-to-seasonal timescales, such as the monsoon intraseasonal oscillation (MISO) and the Madden–Julian oscillation (MJO).

Often predictable from data beyond current dynamical models.

How to use data-driven forecasts of these modes to improve overall forecasts?

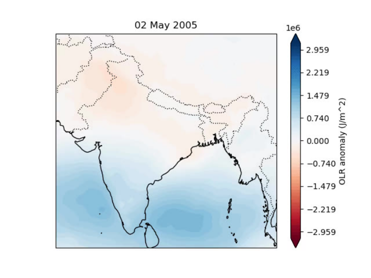
Oscillatory modes in Asian monsoon rainfall

The monsoon intraseasonal oscillation (MISO) is a northward-propagating rainfall mode of period $\sim$45 days.



Oscillatory modes in Asian monsoon rainfall

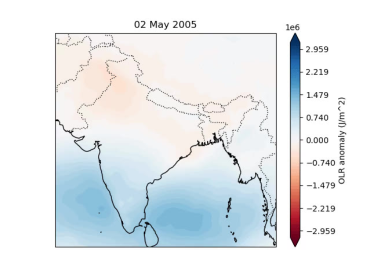
The monsoon intraseasonal oscillation (MISO) is a northward-propagating rainfall mode of period $\sim$45 days.



Controls active and break phases and regional distribution of monsoon rainfall.

Oscillatory modes in Asian monsoon rainfall

The monsoon intraseasonal oscillation (MISO) is a northward-propagating rainfall mode of period $\sim$45 days.



Controls active and break phases and regional distribution of monsoon rainfall.

Essential for subseasonal-to-seasonal prediction with relevance to agriculture, flooding, and water availability.

Combining ML forecasts of oscillatory modes with full-field physical forecasts

The full field admits a modal decomposition; i.e., through singular spectrum analysis (Ghil et al. 2002):

$$P(\mathrm{x}, t) = \sum_i P_i(\mathrm{x}, t) \tag{1}$$

Combining ML forecasts of oscillatory modes with full-field physical forecasts

The full field admits a modal decomposition; i.e., through singular spectrum analysis (Ghil et al. 2002):

$$P(\mathsf{x}, t) = \sum_i P_i(\mathsf{x}, t) \tag{1}$$

The intraseasonal oscillations can be identified as a subset of these modes, defining a *reduced subspace*.

Combining ML forecasts of oscillatory modes with full-field physical forecasts

The full field admits a modal decomposition; i.e., through singular spectrum analysis (Ghil et al. 2002):

$$P(\mathrm{x}, t) = \sum_i P_i(\mathrm{x}, t) \tag{1}$$

The intraseasonal oscillations can be identified as a subset of these modes, defining a *reduced subspace*.

The dynamics in this subspace are *effectively low-dimensional*: can be learned well by ML (e.g., Alexander et al. 2017; Chen et al. 2018; Krishnamurthy and Sharma 2017).

Combining ML forecasts of oscillatory modes with full-field physical forecasts

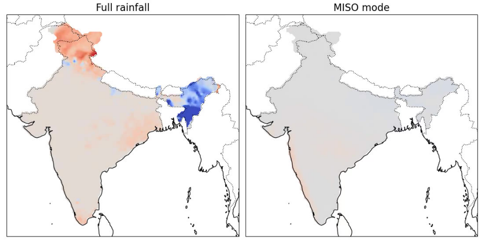
The full field admits a modal decomposition; i.e., through singular spectrum analysis (Ghil et al. 2002):

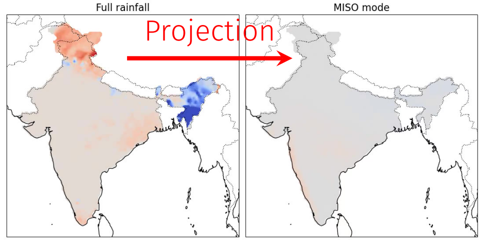
$$P(\mathrm{x}, t) = \sum_i P_i(\mathrm{x}, t) \tag{1}$$

The intraseasonal oscillations can be identified as a subset of these modes, defining a *reduced subspace*.

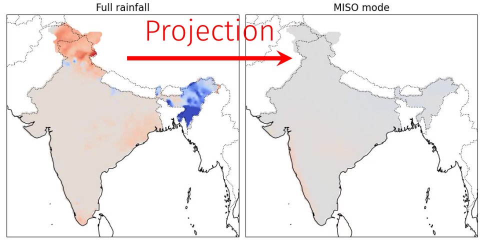
The dynamics in this subspace are *effectively low-dimensional*: can be learned well by ML (e.g., Alexander et al. 2017; Chen et al. 2018; Krishnamurthy and Sharma 2017).

However, these predictions are not useful by themselves, since they only predict a *fraction of the total variance* of the full field.


Full rainfall
MISO mode



Mapping from full phase space to reduced subspace is *non-invertible*.

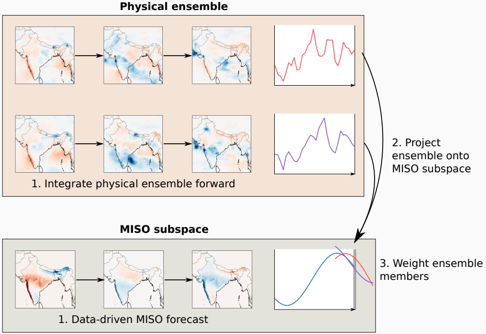


Mapping from full phase space to reduced subspace is *non-invertible.*

Tools from data assimilation can be used to inform the full phase space state from ML forecasts in the reduced subspace!

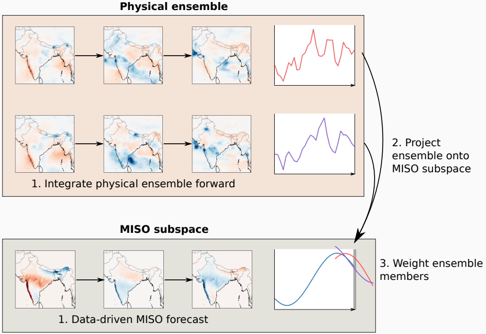
Methods and data

Ensemble Oscillation Correction



Idea (Bach et al. 2021): weight ensemble members of a physical model by their distance from an ML forecast in the corresponding subspace.

Ensemble Oscillation Correction



Idea (Bach et al. 2021): weight ensemble members of a physical model by their distance from an ML forecast in the corresponding subspace.

Similar idea to importance sampling in particle filters: give more weight to ensemble members most likely to result in a predicted MISO pattern.

Methods

Extracted MISO from India Meteorological Department 0.25° gridded rainfall observations since 1901 using multi-channel singular spectrum analysis (M-SSA).

Methods

Extracted MISO from India Meteorological Department 0.25° gridded rainfall observations since 1901 using multi-channel singular spectrum analysis (M-SSA).

Projection into oscillation subspace approximated using a neural network.

Methods

Extracted MISO from India Meteorological Department 0.25° gridded rainfall observations since 1901 using multi-channel singular spectrum analysis (M-SSA).

Projection into oscillation subspace approximated using a neural network.

Analog forecasts in reduced subspace.

Methods

Extracted MISO from India Meteorological Department 0.25° gridded rainfall observations since 1901 using multi-channel singular spectrum analysis (M-SSA).

Projection into oscillation subspace approximated using a neural network.

Analog forecasts in reduced subspace.

Ensemble members weighted using EnOC.

Integrated Forecasting System

We use bias-corrected Integrated Forecasting System (IFS) hindcasts from the ECMWF, started on the first of July, August, and September in the years 2008–2016.

Integrated Forecasting System

We use bias-corrected Integrated Forecasting System (IFS) hindcasts from the ECMWF, started on the first of July, August, and September in the years 2008–2016.

Coupled model with 25 ensemble members, $\sim$36 km resolution and 91 vertical levels.

Integrated Forecasting System

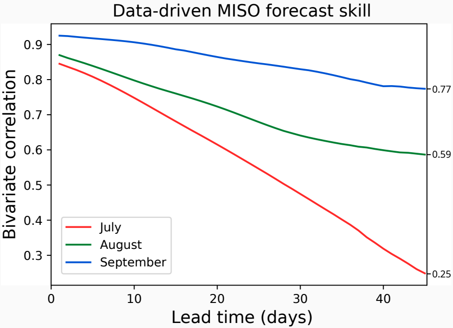
We use bias-corrected Integrated Forecasting System (IFS) hindcasts from the ECMWF, started on the first of July, August, and September in the years 2008–2016.

Coupled model with 25 ensemble members, $\sim$36 km resolution and 91 vertical levels.

IFS has been shown to be state-of-the-art in subseasonal monsoon prediction. It has outperformed all other models to which it has been compared for this task (Jie et al. 2017; Vigaud et al. 2017).

Results

ML MISO forecasts are skillful for over a month



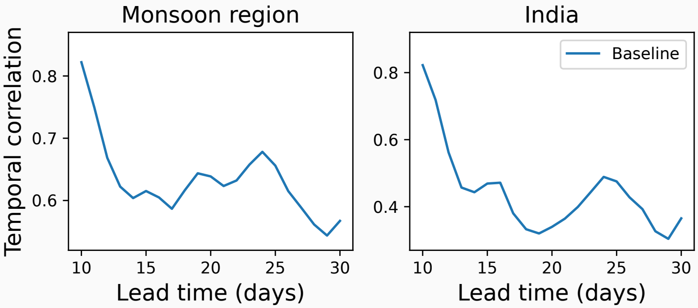
Correlation between ML MISO forecasts and MISO extracted from observations

Skill improvement by using ML forecasts of MISO

We combine a dynamical ensemble of precipitation over South Asia with a data-driven forecast of MISO.

Skill improvement by using ML forecasts of MISO

We combine a dynamical ensemble of precipitation over South Asia with a data-driven forecast of MISO.

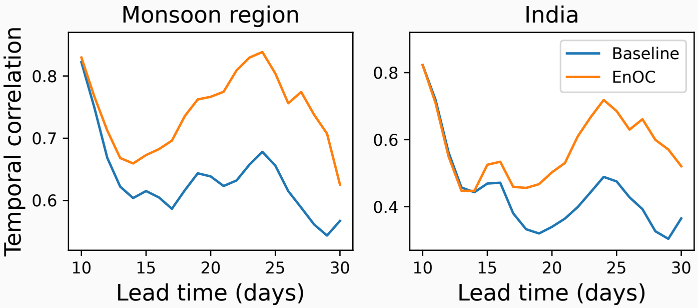


15-day mean rainfall

Baseline: ECMWF bias-corrected ensemble rainfall forecasts

Skill improvement by using ML forecasts of MISO

We combine a dynamical ensemble of precipitation over South Asia with a data-driven forecast of MISO.

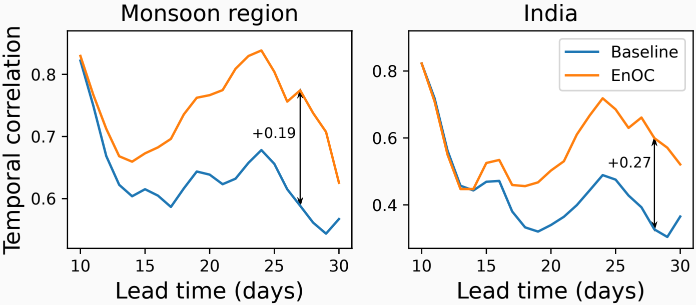


15-day mean rainfall

Baseline: ECMWF bias-corrected ensemble rainfall forecasts

Skill improvement by using ML forecasts of MISO

We combine a dynamical ensemble of precipitation over South Asia with a data-driven forecast of MISO.

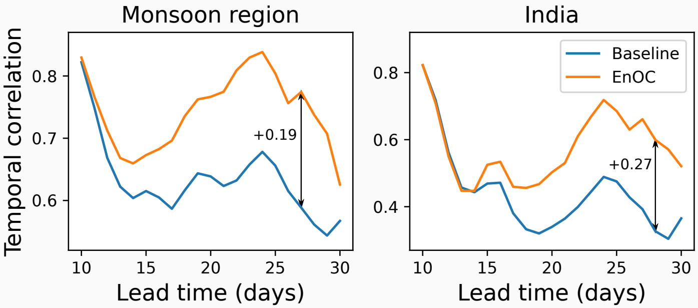


15-day mean rainfall

Baseline: ECMWF bias-corrected ensemble rainfall forecasts

Skill improvement by using ML forecasts of MISO

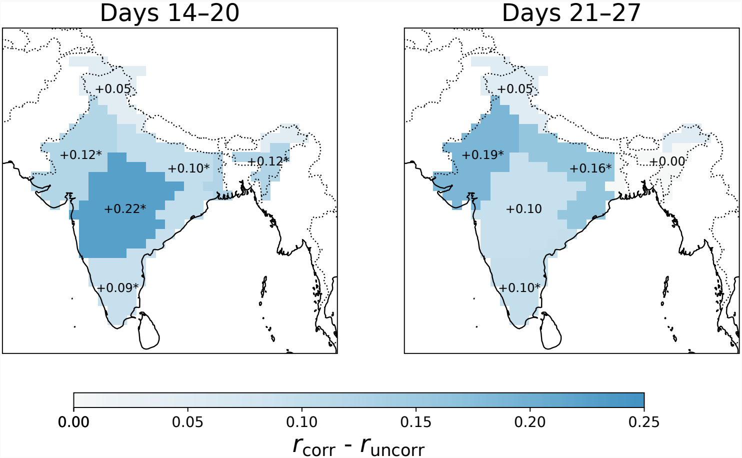
We combine a dynamical ensemble of precipitation over South Asia with a data-driven forecast of MISO.



15-day mean rainfall

Baseline: ECMWF bias-corrected ensemble rainfall forecasts

Regional improvements in skill



Conclusions

Conclusions

By applying EnOC to the South Asian monsoon, we can outperform state-of-the-art forecasts on 10–30 day lead times.

Conclusions

By applying EnOC to the South Asian monsoon, we can outperform state-of-the-art forecasts on 10–30 day lead times.

A demonstration that data-driven forecasts can significantly improve physical model forecasts when combined.

Conclusions

By applying EnOC to the South Asian monsoon, we can outperform state-of-the-art forecasts on 10–30 day lead times.

A demonstration that data-driven forecasts can significantly improve physical model forecasts when combined.

Future work: Application to other important modes of climate variability, in particular the Madden–Julian Oscillation and El Niño.

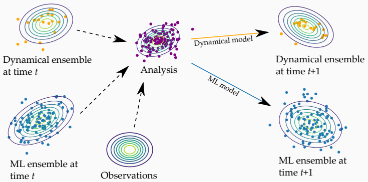
Conclusions

By applying EnOC to the South Asian monsoon, we can outperform state-of-the-art forecasts on 10–30 day lead times.

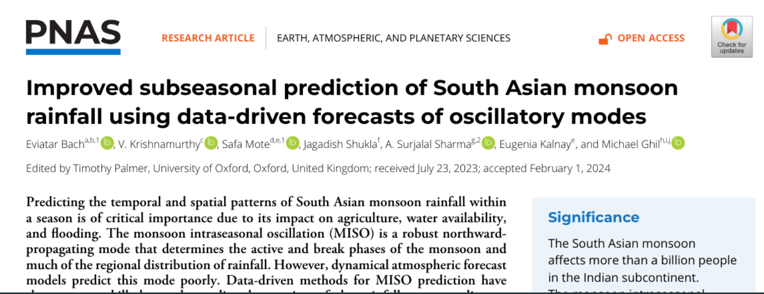
A demonstration that data-driven forecasts can significantly improve physical model forecasts when combined.

Future work: Application to other important modes of climate variability, in particular the Madden–Julian Oscillation and El Niño.

EnOC is a way of combining physical model forecasts with data-driven forecasts. Can be generalized using the Multi-Model Ensemble Kalman Filter (Bach and Ghil 2023).



E. Bach, V. Krishnamurthy, S. Mote, J. Shukla, A. S. Sharma, E. Kalnay, and M. Ghil (2024). **“Improved Subseasonal Prediction of South Asian Monsoon Rainfall Using Data-Driven Forecasts of Oscillatory Modes”.** *Proceedings of the National Academy of Sciences*


PNAS RESEARCH ARTICLE | EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES OPEN ACCESS

Check for updates

Improved subseasonal prediction of South Asian monsoon rainfall using data-driven forecasts of oscillatory modes

Eviatar Bach[a,b,1], V. Krishnamurthy[c], Safa Mote[d,e,1], Jagadish Shukla[f], A. Surjalal Sharma[g,2], Eugenia Kalnay[e], and Michael Ghil[h,i,j]

Edited by Timothy Palmer, University of Oxford, Oxford, United Kingdom; received July 23, 2023; accepted February 1, 2024

Predicting the temporal and spatial patterns of South Asian monsoon rainfall within a season is of critical importance due to its impact on agriculture, water availability, and flooding. The monsoon intraseasonal oscillation (MISO) is a robust northward-propagating mode that determines the active and break phases of the monsoon and much of the regional distribution of rainfall. However, dynamical atmospheric forecast models predict this mode poorly. Data-driven methods for MISO prediction have

Significance

The South Asian monsoon affects more than a billion people in the Indian subcontinent.

My email: eviatarbach@protonmail.com

My website: eviatarbach.com

References i

- Alexander, R., Z. Zhao, E. Székely, and D. Giannakis (2017). "Kernel Analog Forecasting of Tropical Intraseasonal Oscillations". *Journal of the Atmospheric Sciences.*
- Bach, E. and M. Ghil (2023). "A Multi-Model Ensemble Kalman Filter for Data Assimilation and Forecasting". *Journal of Advances in Modeling Earth Systems.*
- Bach, E., V. Krishnamurthy, S. Mote, J. Shukla, A. S. Sharma, E. Kalnay, and M. Ghil (2024). "Improved Subseasonal Prediction of South Asian Monsoon Rainfall Using Data-Driven Forecasts of Oscillatory Modes". *Proceedings of the National Academy of Sciences.*
- Bach, E., S. Mote, V. Krishnamurthy, A. S. Sharma, M. Ghil, and E. Kalnay (2021). "Ensemble Oscillation Correction (EnOC): Leveraging Oscillatory Modes to Improve Forecasts of Chaotic Systems". *Journal of Climate.*

References ii

- Charney, J. G. and J. Shukla (1981). **“Predictability of Monsoons”.** *Monsoon Dynamics*. Ed. by J. Lighthill and R. P. Pearce. Cambridge, U.K.: Cambridge University Press.
- Chen, N., A. J. Majda, C. T. Sabeerali, and R. S. Ajayamohan (2018). **“Predicting Monsoon Intraseasonal Precipitation Using a Low-Order Nonlinear Stochastic Model”.** *Journal of Climate*.
- Ghil, M. et al. (2002). **“Advanced Spectral Methods for Climatic Time Series”.** *Reviews of Geophysics*.
- Goswami, B. N. (2012). **“South Asian Monsoon”.** *Intraseasonal Variability in the Atmosphere–Ocean Climate System*. Ed. by W. K. M. Lau and D. E. Waliser. 2nd edition. Springer-Verlag Berlin Heidelberg.
- Jiang, X., T. Li, and B. Wang (2004). **“Structures and Mechanisms of the Northward Propagating Boreal Summer Intraseasonal Oscillation”.** *Journal of Climate*.

References iii

Jie, W., F. Vitart, T. Wu, and X. Liu (2017). **"Simulations of the Asian Summer Monsoon in the Sub-Seasonal to Seasonal Prediction Project (S2S) Database".** *Quarterly Journal of the Royal Meteorological Society.*

Krishnamurthy, V. and A. S. Sharma (2017). **"Predictability at Intraseasonal Time Scale".** *Geophysical Research Letters.*

Vigaud, N., A. W. Robertson, M. K. Tippett, and N. Acharya (2017). **"Subseasonal Predictability of Boreal Summer Monsoon Rainfall from Ensemble Forecasts".** *Frontiers in Environmental Science.*

Predictability of MISO

Various studies have demonstrated predictability of MISO using data-driven methods (Krishnamurthy and Sharma 2017; Alexander et al. 2017).

Predictability of MISO

Various studies have demonstrated predictability of MISO using data-driven methods (Krishnamurthy and Sharma 2017; Alexander et al. 2017).

The data-driven methods generally predict MISO better than models.

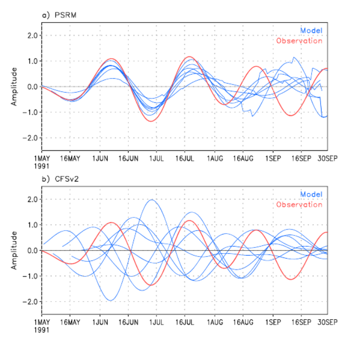


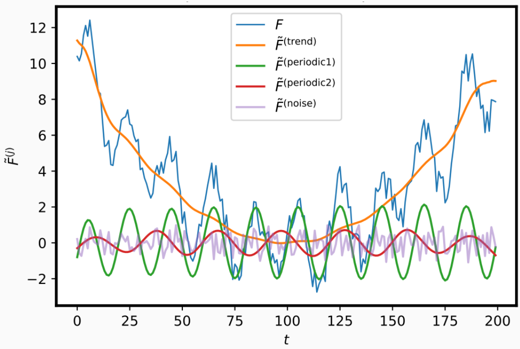
Figure 1: From Krishnamurthy and Sharma 2017

Singular spectrum analysis

SSA is a form of principal component analysis applied to delay-embedded time-series data. Multivariate version called multi-channel SSA (M-SSA)

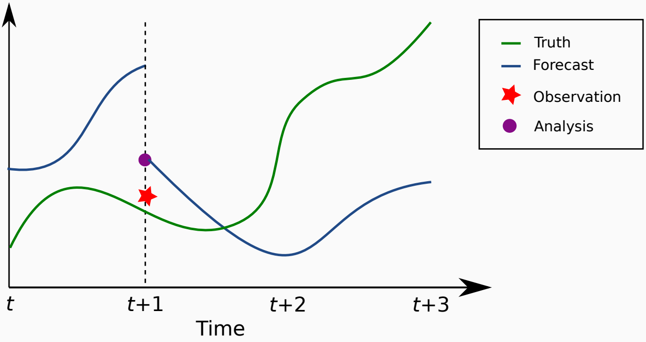
Singular spectrum analysis

SSA is a form of principal component analysis applied to delay-embedded time-series data. Multivariate version called multi-channel SSA (M-SSA)

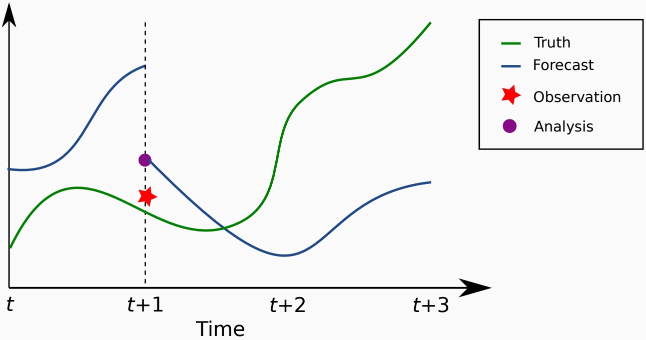


By Jordan D'Arcy

When the physical model is *perfect*, its forecast from the last analysis is the best we can do.



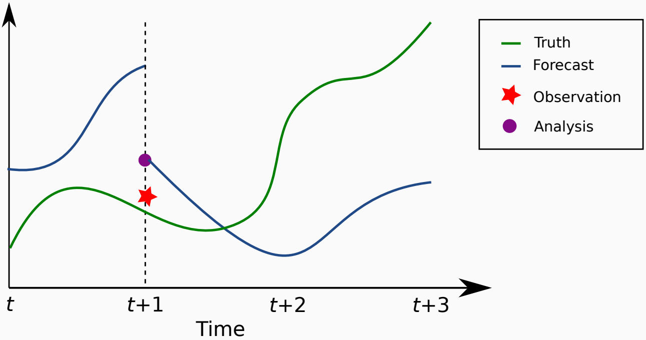
When the physical model is *perfect*, its forecast from the last analysis is the best we can do.



Past observational record provides no extra information (**ho_bayesian_1964**):

$$p(\mathbf{x}(t) \,|\, \mathbf{y}(t-1), \ldots, \mathbf{y}(0)) = p(\mathbf{x}(t) \,|\, \mathbf{x}^a(t-1)) \tag{2}$$

When the physical model is *perfect*, its forecast from the last analysis is the best we can do.

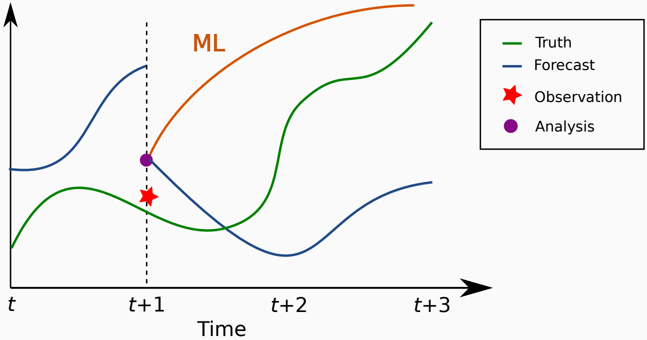


Past observational record provides no extra information (**ho_bayesian_1964**):

$$p(\mathbf{x}(t) \mid \mathbf{y}(t-1), \ldots, \mathbf{y}(0)) = p(\mathbf{x}(t) \mid \mathbf{x}^{\mathrm{a}}(t-1)) \tag{2}$$

If model is *imperfect*, ML can bring in helpful information from the past about the system's dynamics.

When the physical model is *perfect*, its forecast from the last analysis is the best we can do.

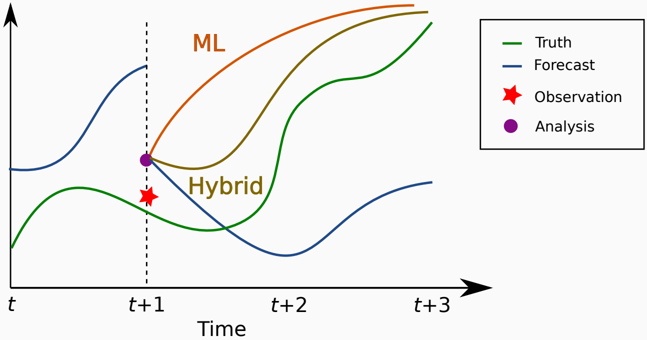


Past observational record provides no extra information (ho_bayesian_1964):

$$p(\mathbf{x}(t) \,|\, \mathbf{y}(t-1), \ldots, \mathbf{y}(0)) = p(\mathbf{x}(t) \,|\, \mathbf{x}^{\mathrm{a}}(t-1)) \tag{2}$$

If model is *imperfect*, ML can bring in helpful information from the past about the system's dynamics.

When the physical model is *perfect*, its forecast from the last analysis is the best we can do.



Past observational record provides no extra information (ho_bayesian_1964):

$$p(\mathbf{x}(t) \mid \mathbf{y}(t-1), \ldots, \mathbf{y}(0)) = p(\mathbf{x}(t) \mid \mathbf{x}^{\mathrm{a}}(t-1)) \qquad (2)$$

If model is *imperfect*, ML can bring in helpful information from the past about the system's dynamics.

Ensemble Oscillation Correction (EnOC)

- Generally, in ensemble forecasting, we weight members equally.

Ensemble Oscillation Correction (EnOC)

- Generally, in ensemble forecasting, we weight members equally.
- However, what if we have some reason to believe some ensemble members are better than others?

Ensemble Oscillation Correction (EnOC)

- Generally, in ensemble forecasting, we weight members equally.
- However, what if we have some reason to believe some ensemble members are better than others?
- We can forecast oscillations accurately using data-driven methods.

Ensemble Oscillation Correction (EnOC)

- Generally, in ensemble forecasting, we weight members equally.
- However, what if we have some reason to believe some ensemble members are better than others?
- We can forecast oscillations accurately using data-driven methods.
- Idea: for ensemble mean, weight members by members' distance from a data-driven oscillation forecast.

Ensemble Oscillation Correction (EnOC)

- Generally, in ensemble forecasting, we weight members equally.
- However, what if we have some reason to believe some ensemble members are better than others?
- We can forecast oscillations accurately using data-driven methods.
- Idea: for ensemble mean, weight members by members' distance from a data-driven oscillation forecast.
- Simple choice works well: pick out best m' ensemble members, exclude the rest. m' is picked based on historical record.

Ensemble Oscillation Correction (EnOC)

- Generally, in ensemble forecasting, we weight members equally.
- However, what if we have some reason to believe some ensemble members are better than others?
- We can forecast oscillations accurately using data-driven methods.
- Idea: for ensemble mean, weight members by members' distance from a data-driven oscillation forecast.
- Simple choice works well: pick out best m' ensemble members, exclude the rest. m' is picked based on historical record.
- More sophisticated weighting: EnOC with data assimilation (EnOC-DA)

Big picture

We can draw upon three sources of information to make a forecast of a system:

1. A dynamical model based on theory

Big picture

We can draw upon three sources of information to make a forecast of a system:

1. A dynamical model based on theory
2. A historical record of observations of the system

Big picture

We can draw upon three sources of information to make a forecast of a system:

1. A dynamical model based on theory
2. A historical record of observations of the system
3. Observations of the system's current state

Big picture

We can draw upon three sources of information to make a forecast of a system:

1. A dynamical model based on theory
2. A historical record of observations of the system
3. Observations of the system's current state

(1) and (3) is traditional DA.

Big picture

We can draw upon three sources of information to make a forecast of a system:

1. A dynamical model based on theory
2. A historical record of observations of the system
3. Observations of the system's current state

(1) and (3) is traditional DA.

(2) and (3) is DA applied to an ML model.

Big picture

We can draw upon three sources of information to make a forecast of a system:

1. A dynamical model based on theory
2. A historical record of observations of the system
3. Observations of the system's current state

(1) and (3) is traditional DA.

(2) and (3) is DA applied to an ML model.

(1) and (2) is, e.g., physics-informed neural networks.

Big picture

We can draw upon three sources of information to make a forecast of a system:

1. A dynamical model based on theory
2. A historical record of observations of the system
3. Observations of the system's current state

(1) and (3) is traditional DA.

(2) and (3) is DA applied to an ML model.

(1) and (2) is, e.g., physics-informed neural networks.

The MM-EnKF is a framework and methodology to combine all three.

Multi-model data assimilation

- The multi-model Kalman filter assimilation step is

$$\mathbf{x}^{\mathrm{a}} = \mathbf{P}^{\mathrm{a}} \left(\sum_{m=1}^{M} \mathbf{G}_m^T \left(\mathbf{P}_m^{\mathrm{f}} \right)^{-1} \mathbf{x}_m^{\mathrm{f}} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y} \right), \tag{3}$$

  where

$$\mathbf{P}^{\mathrm{a}} = \left(\sum_{m=1}^{M} \mathbf{G}_m^T \left(\mathbf{P}_m^{\mathrm{f}} \right)^{-1} \mathbf{G}_m + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1}. \tag{4}$$

Multi-model data assimilation

- The multi-model Kalman filter assimilation step is

$$\mathbf{x}^{\mathrm{a}} = \mathbf{P}^{\mathrm{a}} \left(\sum_{m=1}^{M} \mathbf{G}_m^T \left(\mathbf{P}_m^{\mathrm{f}} \right)^{-1} \mathbf{x}_m^{\mathrm{f}} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y} \right), \tag{3}$$

  where

$$\mathbf{P}^{\mathrm{a}} = \left(\sum_{m=1}^{M} \mathbf{G}_m^T \left(\mathbf{P}_m^{\mathrm{f}} \right)^{-1} \mathbf{G}_m + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1}. \tag{4}$$

- Now, the weights for each model m are inversely proportional to $\mathbf{P}_m^{\mathrm{f}}$. If we set $M = 1$, we recover the regular Kalman filter equations.

Monsoon intraseasonal oscillations

- One theory for the northward propagation involves a vertical shear mechanism (Jiang, Li, and Wang 2004):

Monsoon intraseasonal oscillations

- One theory for the northward propagation involves a vertical shear mechanism (Jiang, Li, and Wang 2004):
    - Convection + easterly vertical shear induces a barotropic vorticity anomaly north of convection.

Monsoon intraseasonal oscillations

- One theory for the northward propagation involves a vertical shear mechanism (Jiang, Li, and Wang 2004):
  - Convection + easterly vertical shear induces a barotropic vorticity anomaly north of convection.
  - This vorticity anomaly leads to low-level moisture convergence, and a northward shift of convection.

Monsoon intraseasonal oscillations

- One theory for the northward propagation involves a vertical shear mechanism (Jiang, Li, and Wang 2004):
  - Convection + easterly vertical shear induces a barotropic vorticity anomaly north of convection.
  - This vorticity anomaly leads to low-level moisture convergence, and a northward shift of convection.

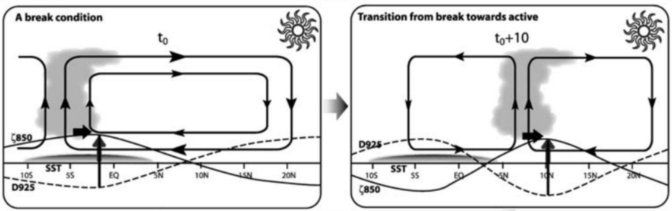


Figure 2: From Goswami 2012