The University of Melbourne
ARC Centre of Excellence for Climate Extremes
Bureau of Meteorology

The Tensor Hybrid forecast covariance model: unifying localization and hybridization

ISDA 2024

Francesco Sardelli

Joint work with **Craig Bishop**

Motivation

The background error covariance matrix is crucial

Met Office upgrades: July 2011 to November 2017

Date	PS	Major Change	KPI impact	
Jul 2011	PS27	DA: Hybrid 4DVar implementation	+2.44	(1)
Jan 2012	PS28	Minor global changes	−0.40	
Mar 2012	PS29	Minor global changes	−0.00	
Sep 2012	PS30	Technical change	−0.00	
Jan 2013	PS31	DA: Improved hybrid 4DVar	+1.18	(3)
Apr 2013	PS32	SA and Dust Assimilation	+0.08	
Feb 2014	PS33	Rose Implementation	−0.10	
Jul 2014	PS34	ENDGAME dynamical core	−0.50	
Feb 2015	PS35	SA package + Dust	+0.86	(5)
Aug 2015	PS36	Transition from IBM to Cray	+0.68	
Mar 2016	PS37	DA/SA: VarBC/NewSats/CVT	+2.40	(2)
Nov 2016	PS38	SA package	+0.99	(4)
Jul 2017	PS39	17 $\longrightarrow$ 10 km UM	+0.53	
Nov 2017	PS40	DA: CVT/waveband localisation	+0.75	

Hybrid covariance matrix $\mathbf{P}^h$

$$\mathbf{P}^h = \beta_e \, \mathbf{P}^e \odot \mathbf{C} + \beta_c \, \mathbf{P}^c$$

β_e and β_c are positive scalar constant coefficients.

$\mathbf{C}$ is the localization matrix.

$\mathbf{P}^c$ is the climatological covariance matrix. It is defined through an average over time. It is almost ***constant***.

$\mathbf{P}^e \odot \mathbf{C}$ is the localized ensemble covariance matrix. It is *ad hoc* technique to correct the rank deficiency and sample noise in $\mathbf{P}^e$.

$\mathbf{P}^e$ is the ensemble covariance matrix generated by an ensemble data assimilation scheme. $\mathbf{P}^e$ is ***time-dependent***.

Introduced in Hamill and Snyder, 2000[a].
Ad hoc approach: little theoretical justification.

[a] Thomas M. Hamill and Chris Snyder (2000). "A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme". In: *Monthly Weather Review* 128.8, pp. 2905–2919.

Promise to the audience

By the end of this talk, I will introduce a new Hybrid:

The **T**-matrix Hybrid

Significantly better than the current Hybrid
in two synthetic experiments

Promise to the audience

By the end of this talk, I will introduce a new Hybrid:

The **T**-matrix Hybrid

Significantly better than the current Hybrid
in two synthetic experiments

To get to that, we will . . .

Our path

- Deepen our conceptual understanding of

the background error covariance matrix

also known as

true forecast error covariance matrix $\mathbf{P}^f$

Our path

- Deepen our conceptual understanding of

the background error covariance matrix

also known as

true forecast error covariance matrix $\mathbf{P}^f$

- Construct a theoretical framework for Hybrid forecast covariance models

The true forecast error covariance matrix $\mathbf{P}^f$

A way to rigorously describe the forecast uncertainty

The true forecast error covariance matrix $\mathbf{P}^f$

A way to rigorously describe the forecast uncertainty

Definition through a thought experiment

The true forecast error covariance matrix $\mathbf{P}^f$

A way to rigorously describe the forecast uncertainty

Definition through a thought experiment

Infinite copies of the Earth with:

- the same atmospheric true state $\mathbf{x}^t$;
- identical observational networks;
- identical numerical weather prediction system.

What varies?

- the results of the observations from one replicate system to the other;
- the resulting forecast $\mathbf{x}^f$.

Forecast error: $\epsilon^f \doteq \mathbf{x}^f - \mathbf{x}^t$

$$\mathbf{P}^f \doteq \mathbb{E}\left[\epsilon^f (\epsilon^f)^\top\right],$$

averaging over all replicate Earth's.

The climatological distribution of $\mathbf{P}^f$

$\mathbf{P}^f$ depends on time

The climatological distribution of $\mathbf{P}^f$

$\mathbf{P}^f$ depends on time

There is a time series of $\mathbf{P}^f$'s

The climatological distribution of $\mathbf{P}^f$

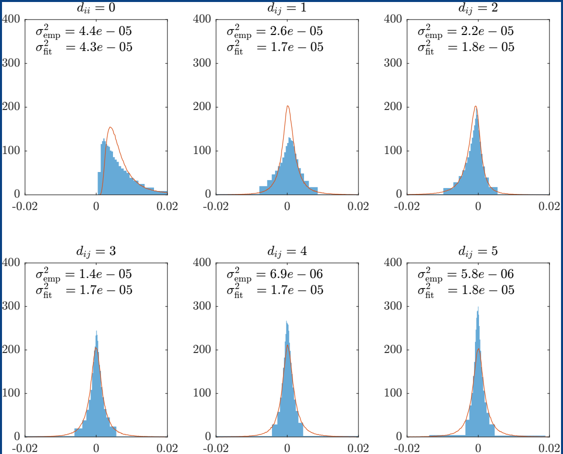
$\mathbf{P}^f$ depends on time

There is a time series of $\mathbf{P}^f$'s

The time series of $\mathbf{P}^f$ defines its climatological distribution

The thought experiment on a computer

- Model: Lorenz 96
- Number of replicate Earth's: 3200



Generalizing Bishop and Satterfield, 2013[1] to full matrices.

[1] Craig H. Bishop and Elizabeth A. Satterfield (2013). "Hidden Error Variance Theory. Part I: Exposition and Analytic Model". In: *Monthly Weather Review* 141.5, pp. 1454–1468

The thought experiment on a computer

- Model: Lorenz 96
- Number of replicate Earth's: 3200

The first ever revelation of a climatological distribution of $\mathbf{P}^f$

Generalizing Bishop and Satterfield, 2013[1] to full matrices.

[1] Craig H. Bishop and Elizabeth A. Satterfield (2013). "Hidden Error Variance Theory. Part I: Exposition and Analytic Model". In: *Monthly Weather Review* 141.5, pp. 1454–1468

Question

Is it possible to know the true forecast error covariance matrix exactly?

Question

Is it possible to know the true forecast error covariance matrix exactly?

No, we would need an infinite number of copies of the planet Earth.

Question

Is it possible to know the true forecast error covariance matrix exactly?

No, we would need an infinite number of copies of the planet Earth.

How can we estimate $\mathbf{P}^f$?

Is there a better approximation to $\mathbf{P}^f$ than $\mathbf{P}^h$?

Bayesian data assimilation

Data assimilators know how to estimate things

Bayesian data assimilation

Data assimilators know how to estimate things

Prior distribution $\mathbf{x}$

Forecast $\mathbf{x}^f = \mathbb{E}[\mathbf{x}]$

Likelihood function $\mathbf{y}|\mathbf{x}$

Observations $\mathbf{y}$

Posterior distribution

Analysis $\mathbf{x}^a = \mathbb{E}[\mathbf{x}|\mathbf{y}]$

An analogy

In the simple case when $\mathbf{H}$ is the identity $\mathbf{I}$,

the observations $\mathbf{y}$ are a *statistically noisy* version of $\mathbf{x}^t$

An analogy

In the simple case when $\mathbf{H}$ is the identity $\mathbf{I}$,

the observations $\mathbf{y}$ are a *statistically noisy* version of $\mathbf{x}^t$

Analogously,

$\mathbf{P}^e$ is a *statistically noisy* version of $\mathbf{P}^f$

A prior estimate for $\mathbf{P}^f$

When you know nothing but the time series of a quantity

you use the time average as a rough estimate

A prior estimate for $\mathbf{P}^f$

When you know nothing but the time series of a quantity

you use the time average as a rough estimate

Thus,

we can use $\mathbf{P}^c$ as a rough estimate of $\mathbf{P}^f$

Hyperfilter

also referred to as secondary filter

In past work on estimating forecast covariances, e.g. Ménétrier and Auligné, 2015[2] and Tsyrulnikov and Rakitko, 2017[3], there is a missing ingredient.

Prior: climatological distribution for $\mathbf{P}^f$

$$\mathbf{P}^c = \mathbb{E}[\,\mathbf{P}^f\,]$$

[2] Benjamin Ménétrier and Thomas Auligné (2015). "Optimized Localization and Hybridization to Filter Ensemble-Based Covariances". In: *Monthly Weather Review* 143.10, pp. 3931 –3947

[3] Michael Tsyrulnikov and Alexander Rakitko (2017). "A Hierarchical Bayes Ensemble Kalman Filter". In: *Physica D: Nonlinear Phenomena* 338, pp. 1–16

Hyperfilter

also referred to as secondary filter

In past work on estimating forecast covariances, e.g. Ménétrier and Auligné, 2015[2] and Tsyrulnikov and Rakitko, 2017[3], there is a missing ingredient.

Prior: climatological distribution for $\mathbf{P}^f$

$\mathbf{P}^c = \mathbb{E}[\,\mathbf{P}^f\,]$

Likelihood function $\mathbf{P}^e|\mathbf{P}^f$

$\mathbf{P}^e$ is known

Posterior distribution $\mathbf{P}^f|\mathbf{P}^e$

Estimate of $\mathbf{P}^f$ given by $\mathbb{E}[\,\mathbf{P}^f|\mathbf{P}^e\,]$

---

[2] Benjamin Ménétrier and Thomas Auligné (2015). “Optimized Localization and Hybridization to Filter Ensemble-Based Covariances”. In: *Monthly Weather Review* 143.10, pp. 3931 –3947

[3] Michael Tsyrulnikov and Alexander Rakitko (2017). “A Hierarchical Bayes Ensemble Kalman Filter”. In: *Physica D: Nonlinear Phenomena* 338, pp. 1–16

Gaussian data assimilation

Specifying the distributions

Prior distribution $\mathbf{x}$

Gaussian distribution: $\mathbf{x} \sim \mathcal{N}(\mathbf{x}^f, \mathbf{P}^f)$

Likelihood function $\mathbf{y}|\mathbf{x}$

Gaussian distribution: $\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{x}^f, \mathbf{R})$

Posterior distribution

Gaussian distribution: $\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mathbf{x}^f, \mathbf{P}^a)$

Physical meaning of the Hybrid

Prior distribution $\mathbf{P}^f$

Inverse Wishart: $\mathbf{P}^f \sim \mathscr{I}\mathscr{W}\left(\nu_{\mathbf{P}^f}, \boldsymbol{\Psi}_{\mathbf{P}^f}\right)$

Likelihood function $\mathbf{P}^e|\mathbf{P}^f$

Wishart:
$\mathbf{P}^e|\mathbf{P}^f \sim \mathscr{W}\left(\nu_{\mathbf{P}^e|\mathbf{P}^f}, \boldsymbol{\Sigma}_{\mathbf{P}^e|\mathbf{P}^f}\right)$

Posterior distribution $\mathbf{P}^f|\mathbf{P}^e$

Inverse Wishart: $\mathbf{P}^f|\mathbf{P}^e \sim \mathscr{I}\mathscr{W}\left(\nu_{\mathbf{P}^f|\mathbf{P}^e}, \boldsymbol{\Psi}_{\mathbf{P}^f|\mathbf{P}^e}\right)$

Estimate of $\mathbf{P}^f$: $\mathbb{E}\left[\mathbf{P}^f|\mathbf{P}^e\right] = \beta_e \mathbf{P}^e + \beta_c \mathbf{P}^c = \mathbf{P}^h$

This is the Hybrid (without localization)!

Kalman filter data assimilation

Linear approximation

Prior distribution $\mathbf{x}$

Forecast $\mathbf{x}^f = \mathbb{E}[\mathbf{x}]$

Likelihood function $\mathbf{y}|\mathbf{x}$

Observations $\mathbf{y}$

Posterior distribution

Analysis $\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{x}^f - \mathbf{y}) \approx \mathbb{E}[\mathbf{x}|\mathbf{y}]$

When $\mathbf{H} = \mathbf{I}, \quad \mathbf{K} = \mathbf{P}^f(\mathbf{P}^f + \mathbf{R})^{-1}$

Vectorizing matrices

Listing the matrix elements of $\mathbf{P}^f$ in a vector $\mathrm{vec}(\mathbf{P}^f)$

(only the upper triangular part needed for symmetric matrices)

$$\mathbf{P}^f = \begin{bmatrix} P^f_{11} & P^f_{12} & \dots & P^f_{1\,n} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & P^f_{n-1\,n} \\ P^f_{n\,1} & \dots & \dots & P^f_{n\,n} \end{bmatrix} \quad \mapsto \quad \mathrm{vec}(\mathbf{P}^f) = \begin{bmatrix} P^f_{11} \\ \vdots \\ P^f_{nn} \\ P^f_{12} \\ \vdots \\ P^f_{n-1\,n} \end{bmatrix}$$

Same for:

$$\mathbf{P}^e \mapsto \mathrm{vec}(\mathbf{P}^e) \quad \text{and} \quad \mathbf{P}^c \mapsto \mathrm{vec}(\mathbf{P}^c)$$

Tensor Hybrid

Linear approximation in the Hyperfilter

Applying the Kalman filter to estimate $\mathbf{P}^f$

Prior: climatological distribution $\mathrm{vec}(\mathbf{P}^f)$

Climatological mean:

$$\mathrm{vec}(\mathbf{P}^c) = \mathbb{E}[\mathrm{vec}(\mathbf{P}^f)]$$

Likelihood function $\mathrm{vec}(\mathbf{P}^e) | \mathrm{vec}(\mathbf{P}^f)$

$\mathrm{vec}(\mathbf{P}^e)$ is known

Posterior distribution $\mathrm{vec}(\mathbf{P}^f) | \mathrm{vec}(\mathbf{P}^e)$

$$\mathrm{vec}(\mathbf{P}^h) = \mathrm{vec}(\mathbf{P}^c) + \mathcal{K}(\mathrm{vec}(\mathbf{P}^e) - \mathrm{vec}(\mathbf{P}^c))$$

$\mathbf{P}^h$ is the Tensor Hybrid

$$\mathcal{K} = \mathbf{\Pi}(\mathbf{\Pi} + \mathcal{R})^{-1}$$

Keep it affordable

$$\mathcal{K} = \mathbf{\Pi}(\mathbf{\Pi} + \mathcal{R})^{-1} \text{ has } \mathcal{O}(n^4) \text{ elements.}$$

How can we modify the Tensor Hybrid to make it more affordable?

First:

- Kalman filter estimation for the diagonal $\mathrm{diag}(\mathbf{P}^f)$ only

Second:

- Kalman filter estimation for each off-diagonal element P^f_{ij}

The resulting Hybrid is the **T**-*matrix Hybrid*

$\mathbf{T}$-matrix Hybrid

First part: diagonal

Applying the Kalman filter to estimate $\mathrm{diag}(\mathbf{P}^f)$

Prior: climatological distribution $\mathrm{diag}(\mathbf{P}^f)$

Climatological mean:

$$\mathrm{diag}(\mathbf{P}^c) = \mathbb{E}[\mathrm{diag}(\mathbf{P}^f)]$$

Likelihood function $\mathrm{diag}(\mathbf{P}^e) | \mathrm{diag}(\mathbf{P}^f)$

$\mathrm{diag}(\mathbf{P}^e)$ is known

Posterior distribution $\mathrm{vec}(\mathbf{P}^f) | \mathrm{vec}(\mathbf{P}^e)$

$$\mathrm{diag}(\mathbf{P}^h) = \mathrm{diag}(\mathbf{P}^c) + \mathbf{T}(\mathrm{diag}(\mathbf{P}^e) - \mathrm{diag}(\mathbf{P}^c))$$

$\mathbf{P}^h$ is the Tensor Hybrid

$$\mathbf{T} = \mathbf{\Pi}_{\mathrm{diag}}(\mathbf{\Pi}_{\mathrm{diag}} + \mathcal{R}_{\mathrm{diag}})^{-1}$$

T-matrix Hybrid

Second part: off-diagonal elements P^f_{ij}

Applying the Kalman filter to estimate P^f_{ij} individually.

Prior: climatological distribution P^f_{ij}	Likelihood function $P^e_{ij}\|P^f_{ij}$
Climatological mean: $P^c_{ij} = \mathbb{E}[\,P^f_{ij}\,]$	P^e_{ij} is known

Posterior distribution $P^f_{ij}|P^e_{ij}$

$$P^h_{ij} = P^c_{ij} + \Gamma_{ij}(P^e_{ij} - P^c_{ij})$$

Scalar quantities: $\Gamma_{ij} = p_{ij}(p_{ij} + r_{ij})^{-1}$

Climatological means

The climatological mean of $\mathbf{P}^f$ is $\mathbf{P}^c$

But

The climatological mean of the current $\mathbf{P}^h$ is not $\mathbf{P}^c$

The climatological mean of
the **T**-matrix Hybrid is $\mathbf{P}^c$

New variables

The previous formulas for the $\mathbf{T}$-matrix Hybrid simplify if we represent the state $\mathbf{x}$ with new variables $\tilde{\mathbf{x}}$:

The *standardized variables*: $\tilde{\mathbf{x}} = (\mathbf{P}^c)^{-\frac{1}{2}}\,\mathbf{x}$

$$\text{with} \quad (\mathbf{P}^c)^{-\frac{1}{2}}\,(\mathbf{P}^c)^{\frac{1}{2}} = \mathbf{I}$$

In standardized variables $\tilde{\mathbf{x}}$, the climatological covariance matrix $\tilde{\mathbf{P}}^c$ is just the identity:

$$\boxed{\tilde{\mathbf{P}}^c = \mathbf{I}}$$

New variables

$\mathbf{P}^f$ is positive semi-definite

Any estimate of it should be positive semi-definite

The *standardized variables* help with the positive semi-definiteness of the **T** matrix Hybrid

Affordably computing $(\mathbf{P}^c)^{-\frac{1}{2}}$

Implementing the standardized variable gain Hybrid requires computing $(\mathbf{P}^c)^{-\frac{1}{2}}$.

Inverses of very big matrices are in general not computationally affordable.

But, we have a

new swift algorithm to compute $(\mathbf{P}^c)^{-\frac{1}{2}}$

I tested this algorithm in 3D synthetic experiment

using spherical harmonics, showing that its output

matches a brute force algorithm.

Determining **T** and Γ_{ij}

- The parametric method

- The HyperInnovation method

- The HyperVariational method

In particular, the HyperVariational method is a **general automated method** to tune the parameters of any Hybrid.

I tested this method in two simple synthetic experiments.

Experiments to test the $\mathbf{T}$-matrix Hybrid

Localized inverse matrix gamma climatology of $\mathbf{P}^f$'s

- Generate $\mathbf{P}^f$ through $\mathbf{P}^f = \mathbf{C} \odot \mathbf{P}^l$ where $\mathbf{P}^l \sim \text{IMG}(\nu_l, \mathbf{P}^c)$

The **T**-matrix Hybrid significantly outperforms the current Hybrid.

Synthetic $\mathbf{P}^f$'s with Gaussian covariance functions

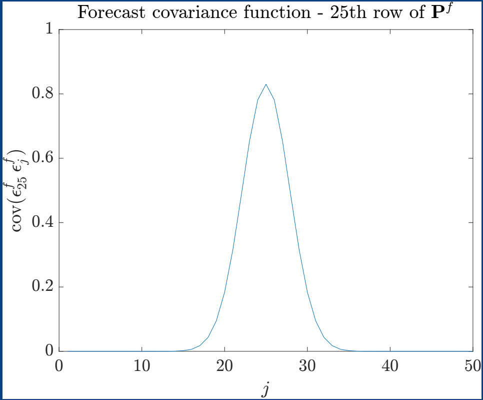
- Generate a length scale l and an amplitude a as random draws from a gamma distribution;
- Use l and a to define Gaussian covariance function, and, with it, a $\mathbf{P}^f$;

The **T**-matrix Hybrid significantly outperforms the current Hybrid.

Synthetic $\mathbf{P}^f$'s with Gaussian covariance functions

In this experiment:

$$f_{25}(j) = v^f \ \exp\left(-\frac{d(25,j)^2}{l^2_{\mathbf{P^f}}}\right)$$



Synthetic $\mathbf{P}^f$'s with Gaussian covariance functions

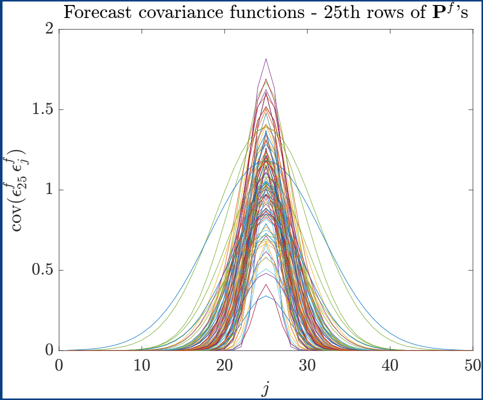
How did I generate a climatology of $\mathbf{P}^f$'s? For each $\mathbf{P}^f$:

1. Random v^f and random $l_{\mathbf{P}^f}$;
2. Construct all forecast covariance functions $f_i(j)$;
3. You have $\mathbf{P}^f$.

Synthetic $\mathbf{P}^f$'s with Gaussian covariance functions

How did I generate a climatology of $\mathbf{P}^f$'s? For each $\mathbf{P}^f$:

1. Random v^f and random $l_{\mathbf{P}^f}$;
2. Construct all forecast covariance functions $f_i(j)$;
3. You have $\mathbf{P}^f$.

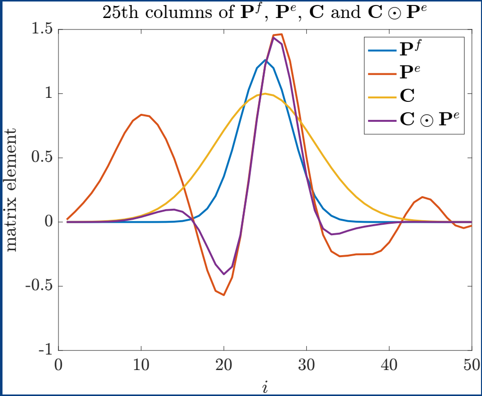


Synthetic $\mathbf{P}^f$'s with Gaussian covariance functions

For each $\mathbf{P}^f$, I generated a corresponding synthetic ensemble with its $\mathbf{P}^e$.

To construct the current Hybrid:

$$\mathbf{P}^h = \beta_e \mathbf{P}^e \odot \mathbf{C} + \beta_c \mathbf{P}^c$$



Synthetic $\mathbf{P}^f$'s with Gaussian covariance functions

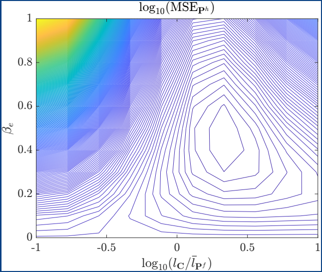
Tuning the current Hybrid $\mathbf{P}^h$

$$\mathbf{P}^h = \beta_e \mathbf{P}^e \odot \mathbf{C} + \beta_c \mathbf{P}^c$$

A consistency equation for the time average $\mathbb{E}[P^h_{ii}]$ yields $\beta_c = 1 - \beta_e$.

Let us consider a localization matrix $\mathbf{C}$ which only depends on a length scale $l_{\mathbf{C}}$.

We only need to tune β_e and $l_{\mathbf{C}}$ to get an optimal $\mathbf{P}^h$



Synthetic $\mathbf{P}^f$'s with Gaussian covariance functions

Results

Analysis (spatio-temporal) mean square errors:

$$\mathrm{MSE}_{\mathbf{P}^c} = 0.4637$$
$$\mathrm{MSE}_{\mathbf{P}^h} = 0.4545$$
$$\mathrm{MSE}_{\mathbf{P}^h_\mathbf{T}} = 0.4475$$

$$\frac{\mathrm{MSE}_{\mathbf{P}^h} - \mathrm{MSE}_{\mathbf{P}^h_\mathbf{T}}}{\mathrm{MSE}_{\mathbf{P}^c} - \mathrm{MSE}_{\mathbf{P}^h}} = 0.75$$

The standardized variables **T**-matrix Hybrid is significantly better than $\mathbf{P}^h$.

The big picture

Some of the most important results

The first ever revelation of a climatological distribution of $\mathbf{P}^f$.

New HyperBaysian models producing a new Hybrid:

1. Whose time mean equals the climatological covariance matrix.
2. Which performs significantly better than the current Hybrid in two synthetic experiments.

A general automated tuning method for the Hybrid parameters.

A new swift algorithm to compute $(\mathbf{P}^c)^{-\frac{1}{2}}$.

This provides a pathway to a geographically dependant static covariance matrix in Hybrid models.