Non-Sequential Ensemble Kalman Filtering using Distributed Arrays

Cédric Travelletti*
Institute of Mathematical Statistics and Actuarial Science
University of Bern
Bern, Switzerland
cedric.travelletti@stat.unibe.ch

David Ginsbourger
Institute of Mathematical Statistics and Actuarial Science
University of Bern
Bern, Switzerland
david.ginsbourger@stat.unibe.ch

Jörg Franke
Oeschger Centre for Climate Change Research
University of Bern
Bern, Switzerland
franke@giub.unibe.ch

Stefan Brönnimann
Oeschger Centre for Climate Change Research
University of Bern
Bern, Switzerland
stefan.broennimann@giub.unibe.ch

Veronika Valler
Oeschger Centre for Climate Change Research
University of Bern
Bern, Switzerland
veronika.valler@giub.unibe.ch

Abstract

We present new methods for distributed estimation of large covariance matrices.

1 Introduction

Kalman filtering is a data assimilation method that aims at ... It has been used successfully in a variety of domains such as ... and ... In spite of these achievements, the original Kalman filter suffers from several limitations that prevent its application at larger scales. This has given rise to a variety of modified Kalman filters, such as ..., ... and ... . Of all the difficulties associated with the vanilla Kalman filter, most of them can be traced back to the background covariance matrix. Indeed, this matrix is quadratic in the number of prediction points, which makes it hard to handle and store at

*C. T. acknowledges funding from the Swiss National Science Fundation under project nr. 178858.

Preprint. Under review.

larger model sizes. One approach proposed to alleviate these difficulties is the ensemble Kalman filter (EnKF), first introduced by Evensen (1994) and subsequently developped in (Evensen, 2003; Evensen et al., 2009). Ensemble Kalman filtering gets rid of the background covariance matrix by instead updating sequentially a sample (ensemble) of the prior background distribution and using the empirical sample covariance as an approximation of the true covariance. Thus, instead of storing a full covariance matrix, EnKF only has to store the (updated) ensemble members at each step.

For most applications, the EnKF provides an efficient, tractable approximation to the vanilla Kalman filter, but it has an Achilles heel: for any application that requires meddling with the covariance, the difficulties of the vanilla Kalman filter come back in full force. An example of such an application is covariance regularization. Indeed, at tractable (small) ensemble sizes, the empirical covariance suffers from undersampling, which significantly degrades the performance of the filter (long-range correllations, variance underestimation, artifacts, ...). This means that, in practice, the empirical covariance has to be regularized before in can be used for filtering (Anderson and Anderson, 1999; Hamill et al., 2001), hence requiring access to the full matrix, bringing us back to the chores of the original filter.

In practice, these problems are usually solved by considering some sequential version of the filter (Houtekamer and Mitchell, 2001), where observations are assimilated in batches, so that, at each iteration, only the parts of the covariance associated to the data points being currently assimilated need to be computed. It has long been known to the community that, when used in conjunction with covariance regularization techniques, sequential processing breaks down the theoretical guarantees of the EnKF and can yield to unwanted consequences. These include, among others, dependency of the results on observations ordering. Heuristical approaches for "optimal" observation ordering during assimilation have been proposed (Whitaker et al., 2008), and while the problem is known, practical applications of the EnKF resort to sequential assimilation as a last resort, owing to the computational impossibility of assimilating all data synchronously, as would be required by the original filter formulation.

The first, and to our knowledge only, systematic study of order dependency in sequential ensemble Kalman filtering was undertaken by (Nerger, 2015). This study demonstrates that the use of covariance localization (Hamill et al., 2001) leads to observation ordering dependency when used in the square root ensemble Kalman filter of Tippett et al. (2003). The study concludes that when observational erros are similar in magnitude to model errors, the ordering effect is of moderate significance. While such an assumption can be met in some setups, it does not hold in general for ensemble-based climate reconstructions (Bhend et al., 2012) where model uncertainties can be large.

In this work, we introduce a new implementation of the ensemble Kalman filter that leverages the most recent progress in distributed computing. Through the use of distributed arrays (Dask Development Team, 2016; Rocklin, 2015), the full model error covariance matrix can be built in (distributed) memory, which allows our implementation to assimilate all observations in a single batch (all-at-once), thus getting rid of ordering dependencies. This allows to use localization in EnKF while still retaining the theoretical properties of the original filter. The computational scalability of this new implementation allows it to conduct all-at-once,localized, EnKF assimilation on large-scale problems such as global climate reconstruction (Bhend et al., 2012; Franke et al., 2017). We believe that by leveraging the most recent developments in distributed computing, this new approach opens the door to a principled use of covariance localization in large-scale applications of ensemble (square-root) Kalman filtering.

To demonstrate our new implementation, we compare the performance of sequential assimilation and all-at-once assimilation on synthetic problems and on real-world paleoclimatic reconstruction applications such as reconstruction of the Tambora eruption event and climate re-analysis of the whole 20th century. In passing, we introduce the use of scoring rules (Gneiting and Raftery, 2007) for comparison of different climate forecasts. While scoring rules are broadly known in the statistics community; climatology and the natural sciences that leverage data assimilation techniques usually resort to metrics that consider the assimilation result as a pointwise prediction. In contrast, scoring rules consider the full probabilitsic nature of the assimilation, allowing for a finer assesment of the forecast accuracy (note that here the word forecast is also used for prediction of past climate states).

This paper is structured as follows: in section 2 we review the basics of the ensemble Kalman filter and its variants. In section 3 we introduce the distributed ensemble Kalman filter (dEnKF) and demonstrate its capabilites. Finally, in section 4, we compare the performances of all-at-once

assimilation with sequential assimilation in synthetic experiments as well as on real-world large-scale climate reconstruction problems.

Our study shows that all-at-once assimilation performs significantly better on all considered problems, while still being computationally tractable. This opens new venues for the study of the effects of covariance localization in climate reconstruction (Valler et al., 2019) and provides the first large-scale extension of the investigations on sequential vs synchronous assimilation performed in (Nerger, 2015) that were limited to small-scale setups, due to lack of computational resources.

2 The Kalman Filter and its Variants

Ensemble Kalman filtering (EnKF) is a data assimilation technique that aims at providing an estimation of the state of a dynamical system by combining observed data with a prior model of the unknown system state. Before going into the details of the EnKF, we review the classic (discrete) Kalman filter, adopting a Bayesian point of view for our exposition. The interested reader is refered to (Snyder, 2014) and (Welch and Bishop, 2000) for more details. The exposition presented here is inspired from the one in (Katzfuss et al., 2016).

In the following, let $\boldsymbol{\psi}_t^* \in \mathbb{R}^m$ denote the unknown state vector of some physical system at time t and assume that the system is governed by the following dynamics

$$\boldsymbol{\psi}_{t+1}^* = \mathcal{F}_t \boldsymbol{\psi}_t^* + \boldsymbol{\delta}_t,\ \boldsymbol{\delta}_t \sim \mathcal{N}(0, \boldsymbol{\Delta}_t), \tag{1}$$

where, for all $t \in \mathbb{N}$, the dynamics of the system is determined by a linear operator $\mathcal{F}_t : \mathbb{R}^m \to \mathbb{R}^m$ and $\boldsymbol{\delta}$ is a random vector modelling our uncertainty in the dynamics of the system.

At each time step t, we are given observations that stem from a linear operator $\boldsymbol{G}_t$ applied to the current unknown state of the system:

$$\boldsymbol{y}_t = \boldsymbol{G}_t \boldsymbol{\psi}_t^* + \boldsymbol{\epsilon}_t,\ \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{E}_t), \tag{2}$$

where $\boldsymbol{G}_t : \mathbb{R}^m \to \mathbb{R}^{n_t}$ is a linear measurement operator and $\boldsymbol{\epsilon}_t$ is a random observation noise vector. Kalman filtering aims at sequentially estimating (filtering) the state of the system, based on the observations available up to the current time step and on some prior knowledge about the initial state at time 0. In a Bayesian formulation, Kalman filtering assumes that the initial state $\boldsymbol{\psi}_0$ is a realization of random vector $\boldsymbol{\Psi}_0$ with Gaussian prior $\boldsymbol{\Psi}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and then approximates the unknown state at time t using the conditional (filtering) distribution of $\boldsymbol{\Psi}_0$ conditionally on the dynamics and the observations up to t. This conditional distribution is Gaussian with mean and covariance that are given by:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_t^{\boldsymbol{f}} + \boldsymbol{K}_t \left(\boldsymbol{y}_t - \boldsymbol{G}_t \boldsymbol{\mu}_t^{\boldsymbol{f}} \right), \tag{3}$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^{\boldsymbol{f}} - \boldsymbol{K}_t \boldsymbol{G}_t \boldsymbol{\Sigma}_t^{\boldsymbol{f}}, \tag{4}$$

where the matrix $\boldsymbol{K}_t$ is called the *Kalman gain* and is given by

$$\boldsymbol{K}_t = \boldsymbol{\Sigma}_t^{\boldsymbol{f}} \boldsymbol{G}_t^T \left(\boldsymbol{G}_t \boldsymbol{\Sigma}_t^{\boldsymbol{f}} \boldsymbol{G}_t^T + \boldsymbol{E}_t \right)^{-1}, \tag{5}$$

assuming the matrix in parenthesis to be invertible, and the *forecast mean* $\boldsymbol{\mu}_t^{\boldsymbol{f}}$ and *forecast covariance* $\boldsymbol{\Sigma}_t^{\boldsymbol{f}}$ are given by:

$$\begin{aligned} \boldsymbol{\mu}_t^{\boldsymbol{f}} &= \mathcal{F}_t \boldsymbol{\mu}_{t-1}, \\ \boldsymbol{\Sigma}_t^{\boldsymbol{f}} &= \mathcal{F}_t \boldsymbol{\Sigma}_{t-1} \mathcal{F}_t^T + \boldsymbol{\Delta}_t. \end{aligned}$$

The key quantity to compute at each iteration is the *Kalman gain*, which is a $m \times n_t$ matrix. In practice, the use of the above Kalman filter algorithm requires computing the $m \times m$ conditional covariance matrix $\boldsymbol{\Sigma}_t$ at each step, which can be prohibitive or even impossible when the state space is large, as in the case, for example, in climate science applications. The Ensemble Kalman Filter (EnKF) was introduced to overcome these difficulties.

2.1 Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) can be seen as a Monte-Carlo approximation of the classical Kalman filter that aims at bypassing the difficulties creaed by large conditional covariance matrices. Its key idea is to approximate the conditional distribution at each step by a sample (ensemble) drawn from the distribution. Then, instead of having to update mean vectors and covariance matrices, one only has to update the ensemble members, thereby reducing the computional burden and memory footprint.

Considering the same setting as in the previous section, ensemble Kalman filtering begins with an *ensemble*, $\boldsymbol{\psi}_0^{(1)}, \dots, \boldsymbol{\psi}_0^{(p)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ which is an i.i.d. sample from the prior state distribution. At each step $t \in \mathbb{N}$, the EnKF updates the current ensemble members $\boldsymbol{\psi}_{t-1}^{(1)}, \dots, \boldsymbol{\psi}_{t-1}^{(p)}$ to a new ensemble $\boldsymbol{\psi}_t^{(1)}, \dots, \boldsymbol{\psi}_t^{(p)}$ so that the new ensemble constitutes an i.i.d. sample of the conditional distribution at time t. The traditional EnKF produces an ensemble that aproximates the conditional distribution by updating the mean and deviations using *perturbed observations*. Indeed, as noted by (Burgers et al., 1998), one needs to inject randomness at each update step so that the filter does not underestimate the variance. Hence, at each step t, given data $\boldsymbol{y}_t$, the EnKF proceeds as follows

1. Sample evolution errors $\boldsymbol{\delta}_t^{(i)} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \boldsymbol{\Delta}_t)$.
2. Forecast members
$$\boldsymbol{\psi}_t^{f(i)} = \boldsymbol{G}_t \boldsymbol{\psi}_{t-1}^{(i)} + \boldsymbol{\delta}_t^{(i)}.$$
3. Sample *perturbed observations* $\boldsymbol{y}_t^{(i)} \overset{\text{i.i.d.}}{\sim} \boldsymbol{y}_t + \boldsymbol{\epsilon}_t$
4. Update the ensemble members:
$$\boldsymbol{\psi}_t^{(i)} = \boldsymbol{\psi}_t^{f(i)} + \hat{\boldsymbol{K}} \left(\boldsymbol{y}_t^{(i)} - \boldsymbol{G} \boldsymbol{\psi}_t^{f(i)} \right). \tag{6}$$

Here, the *ensemble Kalman gain* matrix $\hat{\boldsymbol{K}}$ is the matrix obtained by replacing the true forecast covariance matrix $\boldsymbol{\Sigma}_t^{\boldsymbol{f}}$ in the equation for the Kalman gain Equation (5) by some estimator $\hat{\boldsymbol{\Sigma}}_t^{\boldsymbol{f}}$ of the (unknown) forecast covariance matrix. Indeed, in an ensemble setting, this covariance matrix is not known and has to be estimated from the ensemble itself.

One sees that, at each step, the EnKF maintains an ensemble of p state vectors that approximates the conditional distribution, and that the update equations are analogous to the classical Kalman update equations, with the state forecast covariance replaced by some ensemble estimate thereof and the observations replaced by perturbed ones. The ensemble update equations eq. (6) were chosen so that, in the limit of infinite ensemble size, the empirical ensemble covariance converges to the true covariance. Although theoretically justified, the use of perturbed observations at every step is a source of sampling errors that reduces filter performances at small ensemble sizes (see (Whitaker and Hamill, 2002) for a detailed study). Therefore, deterministic approaches that avoid the use of perturbed observations have been developped. Most of these approaches fall under the category of *square root filters* (Tippett et al., 2003), which we present next.

2.2 Ensemble Square Root Filters

Ensemble square root filters (EnSRF) are a class of ensemble Kalman filters that use deterministic updates to avoid the use of perturbed observations (Whitaker and Hamill, 2002; Tippett et al., 2003). As for the EnKF, our exposition of EnSRF will mostly follow the one in (Katzfuss et al., 2016) and we refer interested readers to this work for more details.

Denoting by $\bar{\boldsymbol{\psi}}_t$ the mean of the ensemble members at time t and by $\boldsymbol{\psi}_t^{(i)'} := \boldsymbol{\psi}_t^{(i)} - \bar{\boldsymbol{\psi}}_t$ the deviations of the ensemble members from the mean, the EnSRF replaces the update equations of the EnKF by an update of the mean and an update of the deviations:

$$\bar{\boldsymbol{\psi}}_t = \bar{\boldsymbol{\psi}}_t^f + \hat{\boldsymbol{K}} \left(\boldsymbol{y}_t - \boldsymbol{G} \bar{\boldsymbol{\psi}}_t^f \right), \tag{7}$$

$$\boldsymbol{\psi}_t^{(i)} = \boldsymbol{\psi}_t^{f(i)} - \tilde{\boldsymbol{K}}_t \boldsymbol{G} \boldsymbol{\psi}_t^{f(i)}, \tag{8}$$

where the *square root Kalman gain* $\tilde{\boldsymbol{K}}_t$ is a matrix defined by:

$$\tilde{\boldsymbol{K}}_t = \hat{\boldsymbol{\Sigma}}_t^{\boldsymbol{f}} \boldsymbol{G}_t^T \left(\sqrt{\boldsymbol{G}_t \hat{\boldsymbol{\Sigma}}_t^{\boldsymbol{f}} \boldsymbol{G}_t^T + \boldsymbol{E}_t} \right)^{-1} \left(\sqrt{\boldsymbol{G}_t \hat{\boldsymbol{\Sigma}}_t^{\boldsymbol{f}} \boldsymbol{G}_t^T + \boldsymbol{E}_t} + \sqrt{\boldsymbol{E}_t} \right)^{-1} \tag{9}$$

In the above, we assume the matrices in parenthesis to be invertible, and the square roots are allowed to be any matrix square root of the original matrix. The EnSRF has been found to overperform the EnKF with perturbed observations for small ensemble sizes Whitaker and Hamill (2002).

2.3 Sequential Filtering and the Problem of Localization

When using ensemble Kalman filters, great care has to be taken in the choice of the ensemble forecast covariance estimator $\hat{\boldsymbol{\Sigma}}_t^{\boldsymbol{f}}$ that is used for the definition of both the *ensemble Kalman gain* in the EnKF Equation (6) and of the *square root Kalman gain* in the EnSRF Equation (9). Indeed, in most applications, the ensemble size is drastically smaller than the state vector, so that estimators based on the empirical covariance of the ensemble will suffer from undersampling errors.

For spatial data assimilation problems, these undersampling errors often show as spurious correlations between distant locations. *Localization* is a procedure to mitigate those and produce a better estimate of the ensemble covariance than the empirical one (Houtekamer and Mitchell, 2001; Hamill et al., 2001). The idea behind localization is to taper the long-range correlations in the empirical covariance by using a distance-depent smoothing. Formally, *localization* estimates the ensemble forecast covariance matrix by

$$\hat{\boldsymbol{\Sigma}}_t^{\boldsymbol{f}} = \mathrm{Cov}\left(\left(\boldsymbol{\psi}_t^{f(i)} \right)_{i=1,\dots,p} \right) \circ \boldsymbol{\rho},$$

where the first term denotes the empirical covariance of the ensemble forecast, $\boldsymbol{\rho}$ is a symmetric positive-definite matrix and $\circ$ denotes elementwise product. In applications where the state vector represents the values of some spatially varying field $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, $(\boldsymbol{\psi}_t)_j = f(x_j, t)$, localization matrices are usually defined by applying a symmetric positive definite function ρ to the spatial locations $\boldsymbol{\rho}_{jk} = \rho(x_j, x_k)$. We note that there also exists non-distance-based localization approaches, such as the ones developed by (Furrer and Bengtsson, 2007).

Sequential Filtering:

3 Distributed Non-Sequential Ensemble Kalman Filtering

Inverses and square roots can be (approximately) computed in distributed memory by using the randomized SVD algorithm from (Halko et al., 2011), which is implemented in `DASK` as `dask.array.linalg.compressed_svd`. We stress that the idea of using randomized SVD in ensemble Kalman filter has already been developed elsewhere, most notably in (Bopardikar, 2017; Farchi and Bocquet, 2019). While our work shares this central idea with the ones just cited, our originality is threefold: i) we provide a **distributed** implementation in a user-friendly Python package, ii) we perform a **detailed study of the improvements** of all-at-once assimilation over the sequential one and iii) we demonstrate our algorithm one **large-scale** problems. On the other hand, the previously cited works do not provide a distributed implementation and do not compare themselves to sequential assimilation.

In the following, let λ_i and $u_i, i = 1, \dots, k$ be the k-largest (approximate) singular values and corresponding left singular vectors of the estimated background covariance $\hat{\boldsymbol{\Sigma}}$ delivered by the rank k randomized truncated SVD algorithm of (Halko et al., 2011). Then, one can compute approximate inverses and square roots of $\hat{\boldsymbol{\Sigma}}$ via:

$$\hat{\boldsymbol{\Sigma}}^{-1} \approx [u_1, \dots, u_k] \begin{pmatrix} 1/\lambda_1 & & \\ & \ddots & \\ & & 1/\lambda_k \end{pmatrix} [u_1, \dots, u_k]^T$$

$$\sqrt{\hat{\boldsymbol{\Sigma}}} \approx [u_1, \dots, u_k] \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_k} \end{pmatrix} [u_1, \dots, u_k]^T.$$

In a lazy setting, this means that one only has to store $2k(m+1)$ values, compared to the $2m^2$ ones needed for storing the full inverse and square root. We can then use the following implementation for a single update step of the EnSRF:

Algorithm 1 Distributed EnSRF update

Require: Ensemble $\boldsymbol{\psi}_t^{(1)}, \ldots, \boldsymbol{\psi}_t^{(p)}$, observation operator $\boldsymbol{G}_t$ and observed data $\boldsymbol{y}_t$
SVD cutoff rank k.
Ensure: Updated ensemble $\boldsymbol{\psi}_t^{(1)}, \ldots, \boldsymbol{\psi}_t^{(p)}$.
Build localized estimated covariance $\hat{\boldsymbol{\Sigma}}_t$ in distributed memory.

$(\lambda_i, u_i)_{i=1,\ldots,k} \leftarrow \textsc{DistributedSVD}(\hat{\boldsymbol{\Sigma}}_t, \text{rank} = k)$
$\hat{\boldsymbol{\Sigma}}_t^{-1} \leftarrow \textsc{ApproximateInverse}((\lambda_i, u_i)_{i=1,\ldots,k})$
$\sqrt{\hat{\boldsymbol{\Sigma}}_t} \leftarrow \textsc{ApproximateSqrt}((\lambda_i, u_i)_{i=1,\ldots,k})$

$(\boldsymbol{\psi}_t^{(i)})_{i=1,\ldots,p} \leftarrow \textsc{KalmanUpdate}((\boldsymbol{\psi}_t^{(i)})_{i=1,\ldots,p}, \hat{\boldsymbol{\Sigma}}_t^{-1}, \sqrt{\hat{\boldsymbol{\Sigma}}_t})$ ▷ Update using eqs. (7) and (8)

4 Experimental Comparison of Sequential and Non-Sequential Filtering

Even if non-sequential ensemble Kalman filtering is theoretically guaranteed to overperform its sequential counterpart when the assumptions of the Kalman filter are met; practical applications of the EnKF almost always lie outside of the validity range of these assumptions. We thus here provide numerical comparisons between all-at-once (aao) and sequential (seq) assimilation schemes for the EnKF on practical assimilation problems. In order to compare performances, we run the two schems in a variety of different scenarios where ground-truth for the state to be estimated is available. We assess the reconstruction quality of the approaches using three different metrics:

- Root mean-square error (RMSE) when predicting with the analysis mean,
- Reduction of error (RE) skill score (Cook et al., 1994),
- Energy score (ES) multivariate scoring rule.

While the first two are commonly used metrics in the Kalman filter community, scoring rules have yet to see broad application in those fields. Compared to the other metrics, the energy score takes the full analysis ensemble into account and also allows for theoretically sound ranking of different models. We provide a more detailed account of scoring rules and their advantages in appendix A.

4.1 Synthetic Test Case

We first compare the all-at-once (aao) and sequential (seq) implementation of the EnKF on a synthetic test case. The task is to reconstruct a two-dimensional scalar field on the unit square, discretized on a $m \times m$ square grid. The ground truth is generated by sampling a Matérn $3/2$ Gaussian process with unit variance and correllation length $\lambda = 0.1$. A starting ensemble (background) is also generated by sampling p realizations from the same process, putting ourselves in a "well-specified" setup. All-at-once and sequential implementations are then compared by assimilating n data points at randomly chosen locations.

To compare the two assimilation schemes, a set of 20 different ground truths is generated by sampling from the GP model. For each such realization, a starting ensemble is also sampled from the same GP model. Then, a set of $n = 300$ randomly chosen data points is assimilated using each scheme and the assimilation results are compared using the aforementioned metrics. The ground truth are discretized on a 80×80 square grid and the goal is to reconstruct the discretized ground truth vector. Figure 1 and fig. 2 depict the starting setup for one iteration of the experiments. The observational noise standard deviation is set to $\sigma_\epsilon = 0.01$, which corresponds to 1% of the standard deviation of the GP model used to generate the ground truths and the ensembles. The effect of the observational noise level on the assimilation is further studied in followup experiments.

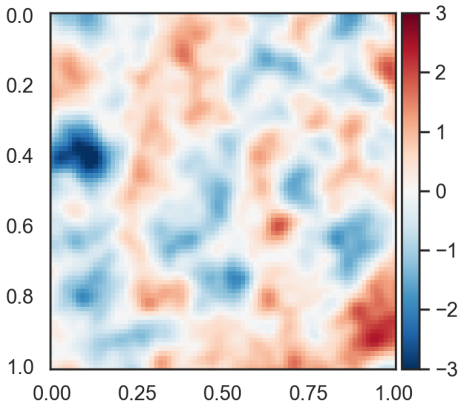
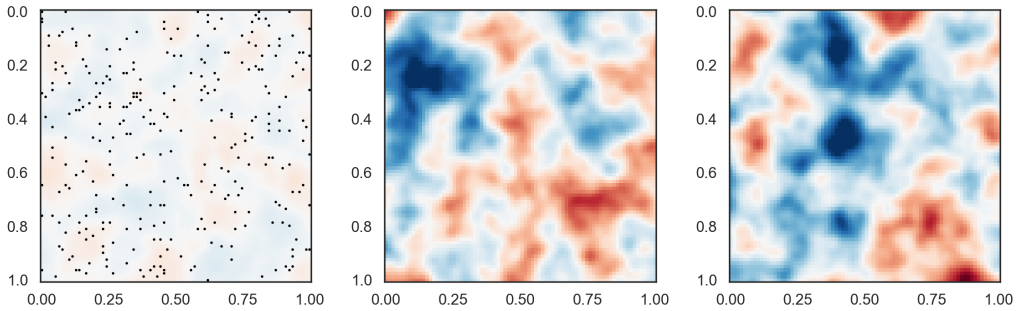


Figure 1: Example of ground truth sampled from the GP model.

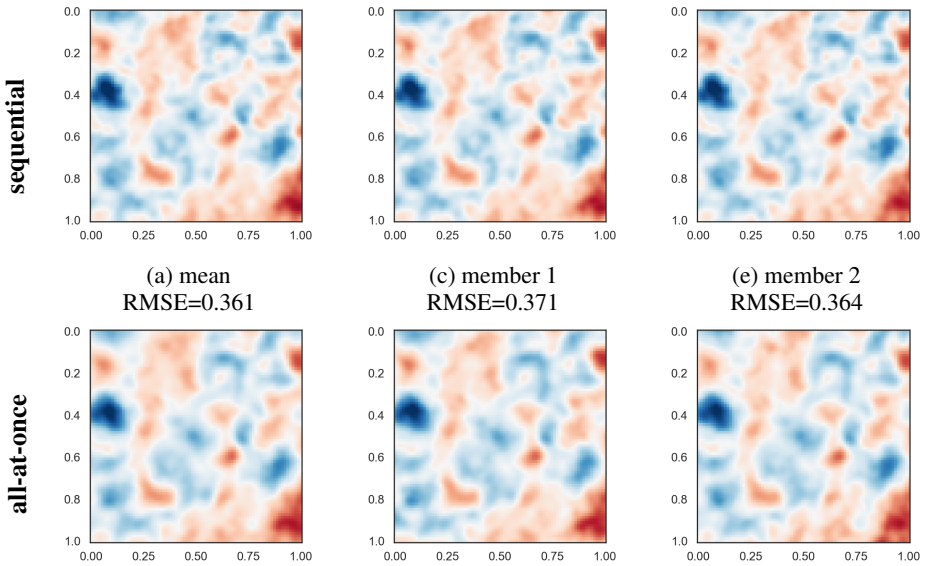


(a) RMSE: 0.808 (b) RMSE: 1.283 (c) RMSE: 1.315

Figure 2: Prior mean (a) and selected prior ensemble members (b) and (c). Observation locations depicted in black. RMSE for prediction of the ground truth in fig. 1 also depicted for each member.

The covariance localization used in the experiment was composition with a spd function (see section 2.3). The localization function used was a Matérn $3/2$ kernel with correllation length set to 0.2 (twice the one of the GP used to generate the ground truths).

Results for assimilation of the data corresponding to the ground truth from Figure 1 are depicted in Figure 3. No obvious qualitative differences between the two assimilation schemes are to be found. Quantitative differences can be computed by considering, the root-mean-square error (RMSE) when reconstructing the ground truth using the data assimilation results. The RMSE when predicting the ground truth with the updated mean and two selected ensemble members is also depicted in Figure 3 for each assimilation scheme. One sees that all-at-once assimilation tends to perform better than the sequential approach.



(a) mean
RMSE=0.361

(c) member 1
RMSE=0.371

(e) member 2
RMSE=0.364

(b) mean
RMSE=0.322

(d) member 1
RMSE=0.346

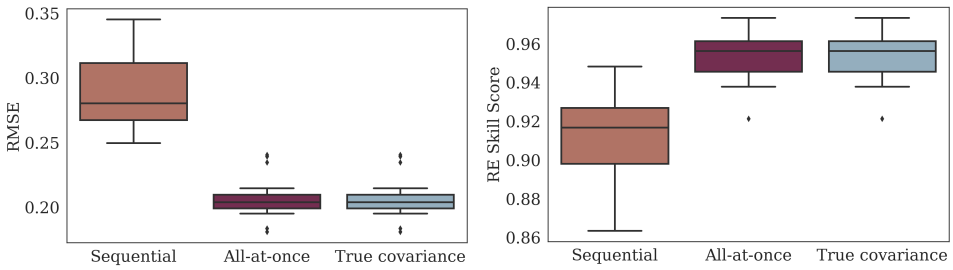
(f) member 2
RMSE=0.361

Figure 3: Updated means and selected updated ensemble members, after assimilation of the data generated from the ground truth in Figure 1.

In order to get a quantitative assesment of the difference in accuracy between sequential and all-at-once assimilation, we repeat the assimilation experiments 20 times by sampling different ground truth from the GP model described before. Apart from the reconstruction RMSE, one can also compute the difference in the *reduction of error* skill score for the different schemes. The *reduction of error* (RE) skill score (Murphy and Epstein, 1989) is a widely used metric for ranking model performances in atmospheric science. For a given set of forecast vectors $\boldsymbol{\psi}_t^f$, the RE score aims at assessing how well the forecast improves the reconstruction a given reference $\boldsymbol{\psi}_t^r$ compared to some original background forecast $\boldsymbol{\psi}_t^b$, over some time period $t \in T$. The RE score is given by (time index ommited for brevity):

$$\mathrm{RE}\left(\boldsymbol{\psi}^f, \boldsymbol{\psi}^r, \boldsymbol{\psi}^b\right) := 1 - \frac{\sum_{t\in T} ||\boldsymbol{\psi}_t^f - \boldsymbol{\psi}_t^r||^2}{\sum_{t\in T} ||\boldsymbol{\psi}_t^b - \boldsymbol{\psi}_t^r||^2}$$

Note that while the original definition of the RE skill score does not include a time dimension, it is customary in climate science to sum over the considered time steps. Compared to the RMSE, this metric weighs the errors by the best improvement that can be achieved, i.e. errors where the background is far from the reference are penalized less heavily. The RE score ranges from 1 to $-\infty$, with a value of 1 corresponding to perfect reconstruction. Figure 4 shows the distribution of the RMSE and RE score for sequential and all-at-once assimilation. Since these results depend on the choice of localization used, we also include the results for all-at-once assimilation using the true covariance matrix that was used to generate the ground truth. This gives an idea of the best possible reconstruction that can be achieved under the current conditions.



(a) RMSE

(b) RE skill score

Figure 4: Comparison of the distributions (20 repetitions) of the RMSE and RE score.

We see that all-at-once outperforms sequential assimilation for all metrics by an order of 2 to 5% (see (Wheatcroft, 2019) on interpreting skill scores).

While the RE skill score is widely used in climate science, it has been pointed out in the litterature (Wheatcroft, 2019) that skill scores can be misleading and biased compared to raw scores. Also, we would like to stress that, in the context of ensemble Kalman filters, the RE score does not leverage the full content of the forecast since it only considers pointwise prediction (ensemble mean) when we have a full ensemble at hand. In order to incorporate the probabilitsic information contained in the spread of the forecast ensemble in the ranking of the assimilation schemes, one should use *proper scoring rules* (Gneiting and Raftery, 2007). We here only focus on the *energy score* (Gneiting et al., 2008), which is a proper scoring rule for multivariate forecasts. For a given ensemble of forecast vectors $\boldsymbol{\psi}_t^f, i = 1, \ldots, p$ and a reference vector $\boldsymbol{\psi}_t^r$ to be reconstructed, the energy score is given by (Jordan et al., 2019):

$$\mathrm{ES}\left(\boldsymbol{\psi}_t^{f(i)}, \boldsymbol{\psi}_t^r\right) := \frac{1}{p}\sum_{i=1}^{p} ||\boldsymbol{\psi}_t^{f(i)} - \boldsymbol{\psi}_t^r|| - \frac{1}{2p^2}\sum_{i=1}^{p}\sum_{j=1}^{p} ||\boldsymbol{\psi}_t^{f(i)} - \boldsymbol{\psi}_t^{f(j)}||$$

Scoring rules are used for model comparison and ranking, lower scores corresponding to better models, according to the scoring at hand. While the energy score has some shortcomings (see (Bjerregård et al., 2021) for a review and presentation of alternative scoring rules), such multivariate scoring rules have yet to find broad use in climate science and we firmly believe that the multivariate and probabilitsic nature of ES already offers an improvement over the univariate point-prediction scorings used in the community. Figure 5 shows that all-at-once also outperforms sequential assimilation in terms of energy score.

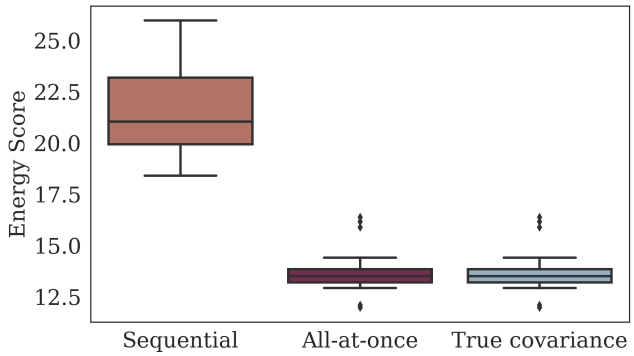
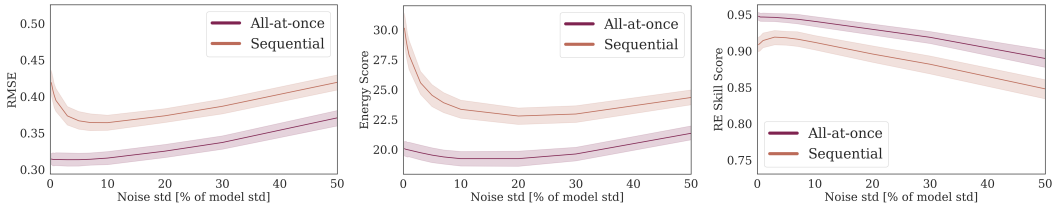


Figure 5: Comparison of the enery score for the different assimilation methods.

The amount by which all-at-once and sequential assimilation differ for the EnSRF depends on the magnitude of the observation erros compared to the spread of the starting ensemble. Indeed, as noted by (Nerger, 2015) *"When the observation errors [are] of a similar magnitude as the initial errors of the state estimate, both filter methods [show] a similar behavior. When the observation errors [are] decreased, the EnSRF [shows] a stronger tendency to diverge and larger minimum RMS errors*

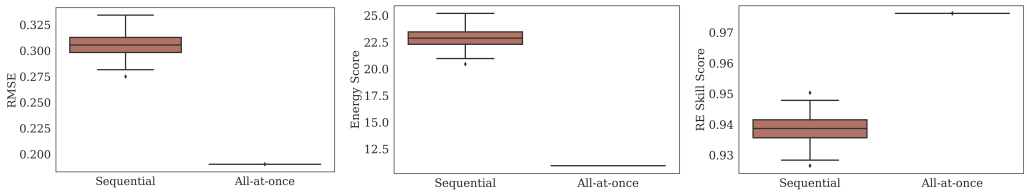
[...] than variant of the EnSRF that assimilates all observations at once." Figure 6 demonstrates this effect by increasing the observational error standard deviation σ_ϵ. Degradation of performance for the sequential EnSRF at small observational errors is clearly visible.



(a) RMSE (b) Energy score (c) RE skill score

Figure 6: Evolution of the different accuracy metrics as a function of the observational noise standard deviation.

On top of the above effect, incorrect updates in the localized, sequential version of the EnSRF introduce dependencies on the order of the observations, as was also noted by (Nerger, 2015). While this effect is well known in the data assimilation community, it is usually assumed to be of little practical relevance. Apart from the small-scale experiments (40 observations) perfomed by (Nerger, 2015), evidence for this assumption is lacking. Our aim here and in the next section is to provide a study of this effect in larger, more realistic situations. Figure 7



(a) RMSE (b) Energy score (c) RE skill score

Figure 7: Comparison of the distributions of the acuracy metrics for different observation orderings (sequential assimilation).

4.2 Paleoclimate Reconstruction Test Case

In order to demonstrate the superiority of all-at-once EnSRF assimilation over the sequential one on real-world problems, we now turn to a paleoclimate reconstruction task and perform the same comparisons as in the last section. This application reproduces a popular Kalman-based method for climate reconstruction Bhend et al. (2012); Franke et al. (2017). The idea is to blend historical temperature recordings with computer simulated climate models to get a global reconstruction of past climate states.

The goal of our paleoclimatic reconstruction task is to reconstruct monthly average temperatures for the period 1960-1980 using an EnSRF. As a reference (groud truth) for the climate state over this period, we use the *Climatic Research Unit gridded Time Series* (CRU TS) dataset (Harris et al., 2020). CRU TS is a high-resolution ($0.5°$ by $0.5°$) dataset of monthly reconstructed climate variables (mean temperature, precipitation rate, ...) produced by interpolating station data. For our practical purpose, this dataset can be considered as embodying the state of the art in climate reconstruction. Our starting ensemble is a set of 30 model simulations from the general circulation model *ECHAM5.4* (Roeckner et al., 2003) and is the same as used for several state-of-the-art climate reconstructions (Bhend et al., 2012; Franke et al., 2017). The simulations incorporate known external forcings such as solar irradiance, volcanic activity and greenhous gas concentration. To bring the simulations back towards reality, we assimilate station data of monthly average temperature. The dataset used for the station data is the *international surface temperature initiative (ISTI) global land surface databank* stage 3 station dataset (Rennie et al., 2014). Figure 8 shows an overview of the stations available in the dataset for January 1960. As for the synthetic test case, we compare the RE skill score, enery score and RMSE when reconstructing using the forecast mean for the different assimilation schemes.

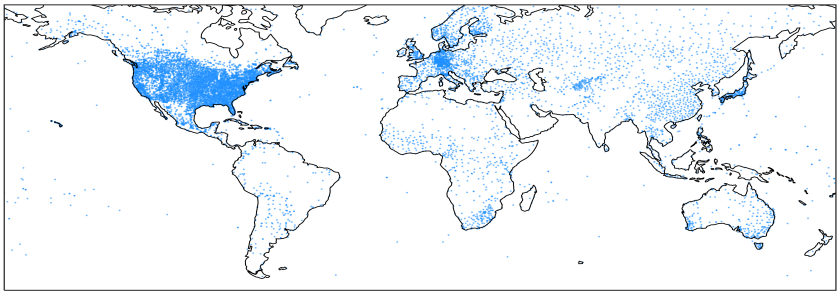
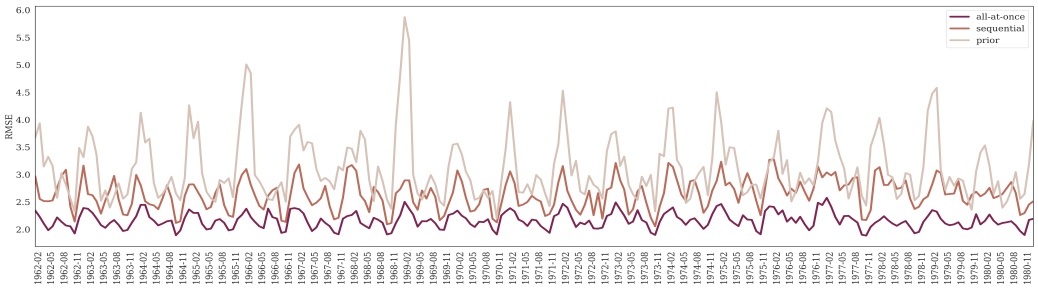
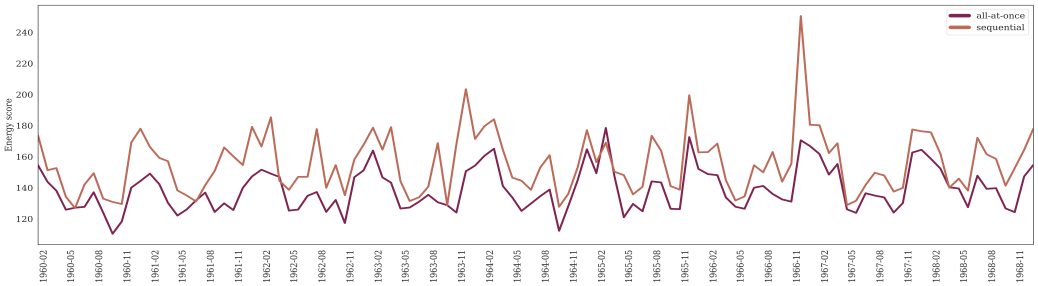


Figure 8: Stations in the ISTI dataset for January 1960.

Figure 9 shows the monthly RMSE and ES and demonstrates how all-at-once assimilation by a magnitude that is often of half a degree in the RMSE. Energy score also indicates that aao should be preffered over seq.

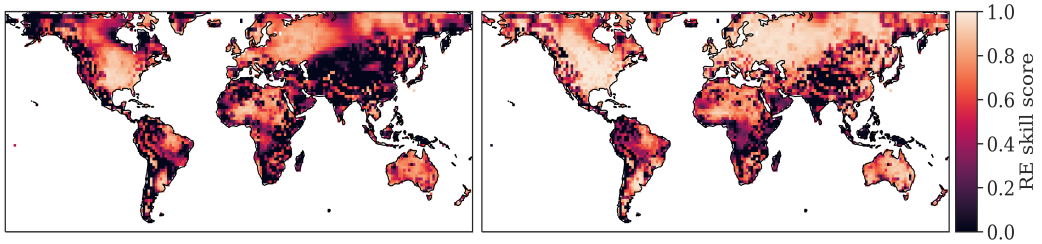


(a) RMSE



(b) Energy score

Figure 9: RMSE of the reconstruction and energy score for different assimilation schemes.



(a) sequential (b) all-at-once

Figure 10: RE skill scores for reconstruction of the reference dataset over the 1960-1980 period.

5 Conclusion

References

Anderson, J. L. and Anderson, S. L. (1999). A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741 – 2758.

Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S. (2012). An ensemble-based approach to climate reconstructions. *Climate of the Past*, 8(3):963–976.

Bjerregård, M. B., Møller, J. K., and Madsen, H. (2021). An introduction to multivariate probabilistic forecast evaluation. *Energy and AI*, 4:100058.

Bopardikar, S. D. (2017). Randomized matrix factorization for kalman filtering. In *2017 American Control Conference (ACC)*, pages 5795–5800.

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126(6):1719 – 1724.

Cook, E. R., Briffa, K. R., and Jones, P. D. (1994). Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *International Journal of Climatology*, 14(4):379–402.

Dask Development Team (2016). *Dask: Library for dynamic task scheduling*.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367.

Evensen, G. et al. (2009). *Data assimilation: the ensemble Kalman filter*, volume 2. Springer.

Farchi, A. and Bocquet, M. (2019). On the efficiency of covariance localisation of the ensemble kalman filter using augmented ensembles. *Frontiers in Applied Mathematics and Statistics*, 5.

Franke, J., Brönnimann, S., Bhend, J., and Brugnara, Y. (2017). A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations. *Scientific Data*, 4(1):170076.

Furrer, R. and Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227–255.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., and Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST*, 17(2):211–235.

Halko, N., Martinsson, P.-G., and Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, 129(11):2776 – 2790.

Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1):109.

Houtekamer, P. L. and Mitchell, H. L. (2001). A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123 – 137.

Jordan, A., Krüger, F., and Lerch, S. (2019). Evaluating probabilistic forecasts with scoringrules. *Journal of Statistical Software*, 90(12):1–37.

Katzfuss, M., Stroud, J. R., and Wikle, C. K. (2016). Understanding the ensemble kalman filter. *The American Statistician*, 70(4):350–357.

Murphy, A. H. and Epstein, E. S. (1989). Skill scores and correlation coefficients in model verification. *Monthly Weather Review*, 117(3):572 – 582.

Nerger, L. (2015). On serial observation processing in localized ensemble kalman filters. *Monthly Weather Review*, 143(5):1554 – 1567.

Rennie, J. J., Lawrimore, J. H., Gleason, B. E., Thorne, P. W., Morice, C. P., Menne, M. J., Williams, C. N., de Almeida, W. G., Christy, J., Flannery, M., Ishihara, M., Kamiguchi, K., Klein-Tank, A. M. G., Mhanda, A., Lister, D. H., Razuvaev, V., Renom, M., Rusticucci, M., Tandy, J., Worley, S. J., Venema, V., Angel, W., Brunet, M., Dattore, B., Diamond, H., Lazzara, M. A., Le Blancq, F., Luterbacher, J., Mächel, H., Revadekar, J., Vose, R. S., and Yin, X. (2014). The international surface temperature initiative global land surface databank: monthly temperature data release description and methods. *Geoscience Data Journal*, 1(2):75–102.

Rocklin, M. (2015). Dask: Parallel computation with blocked algorithms and task scheduling. In Huff, K. and Bergstra, J., editors, *Proceedings of the 14th Python in Science Conference*, pages 130 – 136.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A. (2003). The atmospheric general circulation model echam 5. part i: model description. *Report / Max Planck Institute for Meteorology*, 349.

Snyder, C. (2014). Introduction to the Kalman filter. In *Advanced Data Assimilation for Geosciences: Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012*. Oxford University Press.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S. (2003). Ensemble square root filters. *Monthly weather review*, 131(7):1485–1490.

Valler, V., Franke, J., and Brönnimann, S. (2019). Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation. *Climate of the Past*, 15(4):1427–1441.

Welch, G. and Bishop, G. (2000). An introduction to the kalman filter. Technical Report TR95041, Department of Computer Science, University of North Carolina, Chapel Hill.

Wheatcroft, E. (2019). Interpreting the skill score form of forecast performance metrics. *International Journal of Forecasting*, 35(2):573–579.

Whitaker, J. S. and Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913 – 1924.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., and Toth, Z. (2008). Ensemble data assimilation with the ncep global forecast system. *Monthly Weather Review*, 136(2):463–482.

A Scoring Rules for Evalutation of Probabilistic Forecasts

SCRIPTS: /home/cedric/PHD/Dev/DIESEL/reporting/paleoclimate/plot_scores_synthetic.ipynb