Evaluation of the Doctoral Thesis entitled 'Covariance estimation for filtering in high dimensions' by Mgr. Marie Turcicova.

The thesis focuses on estimating covariance matrices from small ensembles. It consists of an introductory chapter on covariance operators and an overview chapter on existing methods for covariance estimation in high dimensions. This is followed by 3 chapters on maximum likelihood estimators, score matching estimators, and M-estimators containing many new results. Chapter 6 provides an introduction to Stochastic Ensemble Kalman filters (SEnKF), and chapter 7 uses the results from chapters 3,4 and 5 to develop new filtering algorithms that are tested in numerical experiments and their performance compared to the SEnKF and its diagonal variant.

The thesis is well-written and the logical development of ideas is well laid out. It was a joy to read. Below the different chapters are evaluated, and some discussion points raised.

Chapter 1 defines the covariance operator and its connection with covariance matrices in finite-dimensional spaces. It further introduces the idea of transforming a covariance matrix to a (near) diagonal form to reduce the estimation burden in the transformed space. **Chapter 2** contains a very useful and clear overview of methods that generate full rank covariance matrices with minimal spurious correlations, including the tapering method via a Shur product which is most used in the geosciences. Then parametric methods are discussed, specifically regularization in the spectral and inverse.

For Maximum Likelihood Estimators (MLE), **Chapter 3** contains a very elegant proof that a minimum parameter set gives a lower asymptotic posterior error covariance than any larger parameter set. Asymptotic analytical solutions are provided for several settings where the to-be-estimated covariance matrix is diagonal, after a Fourier transformation. Simple numerical experiments are provided that illustrate the accuracy of the theory, even for ensemble sizes that are much smaller than the size of the system. Furthermore, experiments with a non-diagonal covariance matrix also showed the superiority of using all prior knowledge on the covariance structure.

Unfortunately, in practice we will seldom know details of the structure of the covariance matrix, and the size of the minimal parameter set is unknown. I would love to discuss this with the candidate as it would be interesting to see what happens if the parameter set is chosen too low, or, more importantly, how one would determine the minimal parameter set, given a finite ensemble size.

Score matching estimators (SME) are introduced in **chapter 4**. Their introduction to covariance estimation in data assimilation is brilliant as it does provide closed-form expressions for the estimators where MLE does not. After a much-appreciated thorough introduction, a small Lemma (7) is provided and proven which states that if an N-sample operator converges for N to infinity to an operator that has an inverse, than the inverse of the N-sample operator exists with probability 1 for N to infinity. This result will be quite useful for later developments.

This chapter also proves continuity of the Score matching Estimator with respect to random perturbations from the exponential family, a crucial result for application to data assimilation. A main result is the closed-form expression for the SME for the parameters in a linear model of the precision matrix. Furthermore, for Gaussian Markov random fields this SME is proven to be consistent. These results hold when the matrix formed by elements $tr(SA_kA_j)$, where S is the sample covariance and A_i are the design matrices for the linear model for the precision matrix, is invertible. It is very useful that a simple to apply condition for this invertibility is given: the A_i have to be linearly independent.

A question came up while studying this chapter. MLE can be considered the mode of the posterior pdf on theta using a flat prior. Can the SME also be connected to Bayes Theorem, and if so, how? A bit more discussion on the fact that while the MLE and SME result in similar precision matrices, the corresponding covariance matrices can be quite different. For instance, in Fig 4.8 on

page 61 MLE does not have negative values, while SME does have negative off diagonal elements. It would great to discuss this further with the candidate.

Chapter 5 discusses the asymptotic variance of M-estimators, of which both MLE and SME are examples. This chapter nicely generalizes the nested results from chapter 3.

The standard Stochastic Ensemble Kalman Filter is introduced in **Chapter 6**. It uses the formulation that needs the full ensemble covariance matrix and does not discuss more recent algorithms in which this matrix is never formed by transforming the problem to ensemble space. The matrix to be inverted is of the size of the ensemble in these formulations. However, interestingly, removal of spurious correlations is still essential for accurate data assimilation. It would be interesting to discuss how these ideas connect to the developments in this thesis.

Chapter 7 develops three new ensemble Kalman filter algorithms based on the SME and applies them to two toy examples. The assumption is made that the model error covariance and the observation error covariance are diagonal, but I must confess I do not see why. That said, the methods are elegant, simple but efficient. One of them is based on Gaussian resampling, and a much-wanted proof of the consistency of this filter for linear model and observation operator is provided. The performance of two methods is tested and is remarkable (the third is only of use for small dimensional models with fixed design matrices for the precision matrix). The so-called Score Matching Ensemble Filter (SMEF) even systematically outperforms the standard Stochastic Ensemble Kalman Filter, although a consistency proof could not be provided because Gaussianity is lost in this filter.

The numerical experiments could have been described in slightly more detail (How does the ensemble spread, crucial for useful weather forecasting, perform compared to the RMSE? How are the truth and the observations generated? Note that Lorenz 1996 is for mid-latitudes, not the equator.), but provide a valuable demonstration of the filters in action. It is found that the SMEF performs better than SMF-GR, and the explanation given is that the latter explicitly assumes Gaussian processes, while the system under study is non-Gaussian. However, no proof of the non-Gaussianity is given and my guess is that the Gaussian assumption is not far off. This would point to a deeper reason, perhaps related to the performance discussed in chapter 4, which would be great to discuss further.

To conclude, this thesis is a very valuable addition to existing knowledge, both on the fundamental mathematical side and for operational applications. It triggers further scientific discussion, which is exactly the way it should be. I expect that the material in this thesis can have serious impact in numerical weather prediction and similar fields because of 1) the thorough mathematical analysis of the consistency of the schemes, so that we understand their limitations and applicability, and 2) covariance estimation from small samples is a hot topic in operational weather prediction and related fields and will be so for some time in the future, and 3) the developed schemes are practical and can be implemented with minor modifications in real operational systems.

Peter Jan van Leeuwen
Professor in Data Assimilation
Department of Atmospheric Sciences,
Colorado State University, USA,
and
Department of Meteorology,
University of Reading, UK