Sampling error in the estimation of observation error covariance matrices using observation-minus-background and observation-minus-analysis statistics

Guannan Hu[1] and Sarah L. Dance[1,2]

[1]*University of Reading, United Kingdom*

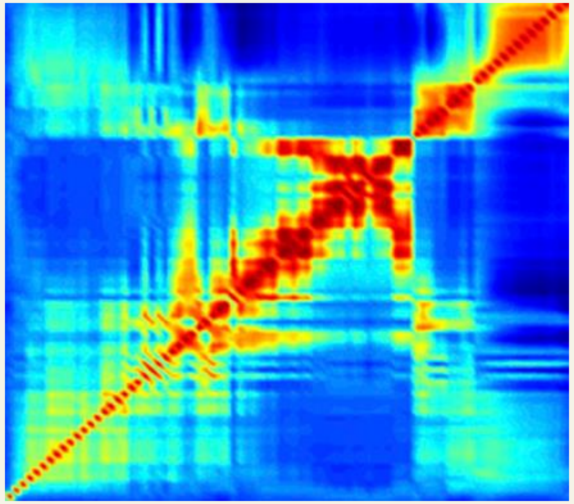
[2]*National Centre for Earth Observation (NCEO), United Kingdom*

Motivation

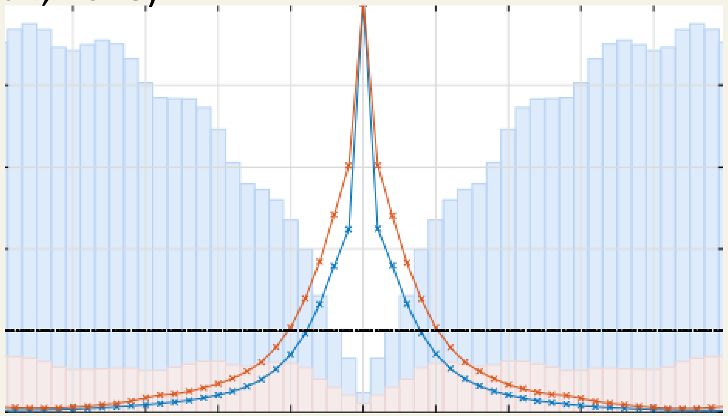
- **Observation error covariance** matrices affect the accuracy of analyses and forecasts.

- An indirect sampling approach is widely used to estimate correlated observation error statistics **(Desroziers et al., 2005)**.

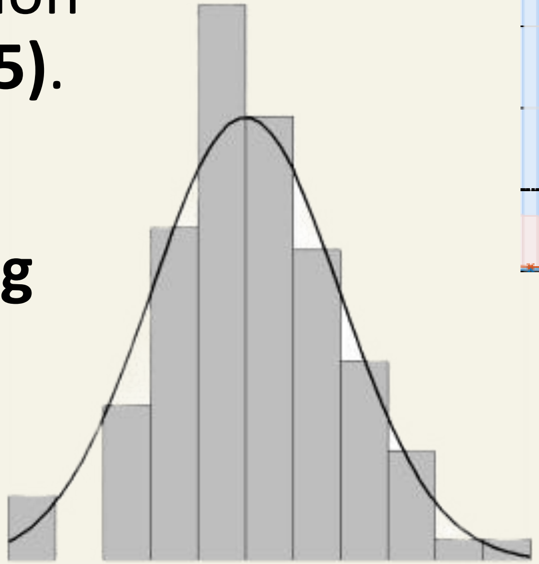
- Our goal is to investigate the **sampling error** of this method.



Interchannel correlation
(Gauthier et al., 2018)



Spatial correlation
(Waller et al.,2016, Remote sensing)



Desroziers et al. diagnostics

- Observation-minus-background (O-B) statistics

$$
\begin{aligned}
\mathbf{d}^{o-b} &= \mathbf{y} - H(\mathbf{x}^b) \\
&= (\mathbf{y} - H(\mathbf{x}^t)) - (H(\mathbf{x}^b) - H(\mathbf{x}^t)) \\
&\approx \boldsymbol{\epsilon}^o - \mathbf{H}(\mathbf{x}^b - \mathbf{x}^t) \\
&\approx \boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^b
\end{aligned}
$$

$\mathbf{y} \in \mathbb{R}^m$: observation vector

$\mathbf{x}^{\mathrm{b}} \in \mathbb{R}^n$: background model state vector

H: nonlinear observation operator $\mathbb{R}^n \to \mathbb{R}^m$

$\mathbf{x}^{\mathrm{t}} \in \mathbb{R}^n$: true model state vector

$\mathbf{H} \in \mathbb{R}^{m \times n}$: linearised observation operator

$\boldsymbol{\epsilon}^o$: observation error

$\boldsymbol{\epsilon}^b$: background error

Desroziers et al. diagnostics

- Statistical expectation

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{d}^{o-b}\left(\mathbf{d}^{o-b}\right)^{\mathrm{T}}\right] &= \mathbb{E}\left[\left(\boldsymbol{\epsilon}^{o} - \mathbf{H}\boldsymbol{\epsilon}^{b}\right)\left(\boldsymbol{\epsilon}^{o} - \mathbf{H}\boldsymbol{\epsilon}^{b}\right)^{\mathrm{T}}\right] \\
&= \mathbb{E}[\boldsymbol{\epsilon}^{o}(\boldsymbol{\epsilon}^{o})^{\mathrm{T}}] - \mathbb{E}\left[\boldsymbol{\epsilon}^{o}\left(\boldsymbol{\epsilon}^{b}\right)^{\mathrm{T}}\right]\mathbf{H}^{\mathrm{T}} + \mathbf{H}\mathbb{E}[\boldsymbol{\epsilon}^{b}(\boldsymbol{\epsilon}^{o})^{\mathrm{T}}] + \mathbf{H}\mathbb{E}\left[\boldsymbol{\epsilon}^{b}\left(\boldsymbol{\epsilon}^{b}\right)^{\mathrm{T}}\right]\mathbf{H}^{\mathrm{T}} \\
&= \mathbb{E}[\boldsymbol{\epsilon}^{o}(\boldsymbol{\epsilon}^{o})^{\mathrm{T}}] + \mathbf{H}\mathbb{E}\left[\boldsymbol{\epsilon}^{b}\left(\boldsymbol{\epsilon}^{b}\right)^{\mathrm{T}}\right]\mathbf{H}^{\mathrm{T}} \\
&= \mathbf{R} + \mathbf{HBH}^{\mathrm{T}} \\
&= \mathbf{D}
\end{aligned}
$$

$\mathbf{D} \in \mathbb{R}^{m\times m}$: Innovation covariance matrix
$\mathbf{R} \in \mathbb{R}^{m\times m}$: observation error covariance matrix
$\mathbf{B} \in \mathbb{R}^{n\times n}$: background error covariance matrix

Desroziers et al. diagnostics

- Observation-minus-analysis (O-A) residuals

$$
\begin{aligned}
\mathbf{d}^{o-a} &= \mathbf{y} - H(\mathbf{x}^a) \\
&= \mathbf{y} - H\left(\mathbf{x}^b + \delta\mathbf{x}\right) \\
&= \mathbf{y} - H\left(\mathbf{x}^b\right) - \mathbf{H}\delta\mathbf{x} - \mathcal{O}(\|\delta\mathbf{x}\|^2) \\
&\approx \mathbf{d}^{o-b} - \mathbf{H}\delta\mathbf{x} \\
&= (\mathbf{I} - \mathbf{HK})\mathbf{d}^{o-b} \\
&= \mathbf{R}\left(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \mathbf{d}^{\mathrm{o-b}}
\end{aligned}
$$

$\delta\mathbf{x} = \mathbf{K}\mathbf{d}^{o-b}$

$\mathbf{x}^{\mathrm{a}} \in \mathbb{R}^n$: analysis model state vector

$\mathbf{K} \in \mathbb{R}^{n\times m}$: Kalman gain matrix

Desroziers et al. diagnostics

- We obtain in the previous slides

$$\mathbb{E}\left[\mathbf{d}^{o-b}\left(\mathbf{d}^{o-b}\right)^{\mathrm{T}}\right] = \mathbf{R} + \mathbf{HBH}^{\mathrm{T}}$$

$$\mathbf{d}^{o-a} = \mathbf{R}\left(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \mathbf{d}^{\mathrm{o-b}}$$

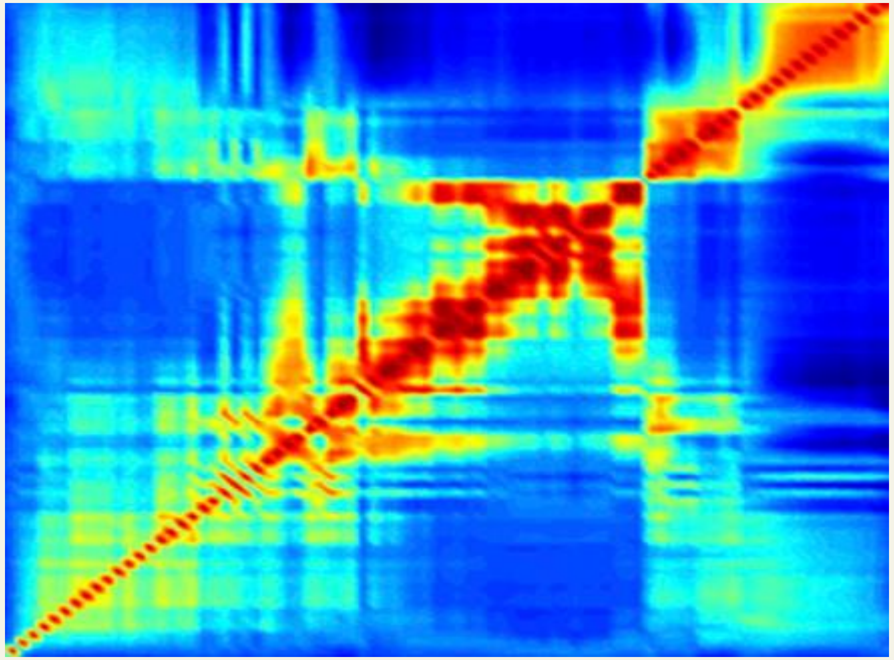
- The expectation of the outer product of $\mathbf{d}^{o-a}$ and $\mathbf{d}^{o-b}$

$$\mathbb{E}\left[\mathbf{d}^{o-a}\left(\mathbf{d}^{o-b}\right)^{\mathrm{T}}\right] = \mathbf{R}\left(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \mathbb{E}\left[\mathbf{d}^{o-b}\left(\mathbf{d}^{o-b}\right)^{\mathrm{T}}\right] = \mathbf{R}$$

NOTE: we assume that the error covariance matrices used describe the truth completely accurately, otherwise (Waller et al., 2016; Janjic et al., 2018)

$$\mathbb{E}\left[\mathbf{d}^{o-a}\left(\mathbf{d}^{o-b}\right)^{\mathrm{T}}\right] \neq \mathbf{R}$$

Desroziers et al. diagnostics



Interchannel correlation
(Gauthier et al., 2018)

Some remarks:

- A commonly used technique to estimate correlated $\mathbf{R}$ (interchannel and spatial correlations).
- The matrices estimated are noisy and have to be reconditioned for operational use.

Sample covariance matrix

Innovation covariance matrix (direct sampling)

$$\mathbf{D} = \mathbb{E}\left[\mathbf{d}^{o-b}\left(\mathbf{d}^{o-b}\right)^{\mathrm{T}}\right]$$

$$\widehat{\mathbf{D}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{d}_i^{o-b}\left(\mathbf{d}_i^{o-b}\right)^{\mathrm{T}} - \overline{\mathbf{d}^{o-b}}\left(\overline{\mathbf{d}^{o-b}}\right)^{\mathrm{T}}$$

N: sample size

$\mathbf{d}_i^{o-b}$: i-th O-B residual

$\overline{\mathbf{d}^{o-b}}$: sample mean

Sample covariance matrix

Observation error covariance matrix (indirect sampling)

$$\mathbf{R} = \mathbb{E}\left[\mathbf{d}^{o-a}\left(\mathbf{d}^{o-b}\right)^{\mathrm{T}}\right]$$

$$\widehat{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{d}_i^{o-a}\left(\mathbf{d}_i^{o-b}\right)^{\mathrm{T}} - \overline{\mathbf{d}^{o-a}}\left(\overline{\mathbf{d}^{o-b}}\right)^{\mathrm{T}}$$

Previous work on direct sampling error (Ledoit and Wolf, 2004)

- Expected quadratic loss of sample covariance matrix $\widehat{\mathbf{D}}$

$$\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{D}}-\mathbf{D}\right\|_F^2\right]=\alpha(\mu^2+\theta)-\beta,$$

$\|\cdot\|_F$: Frobenius norm (elementwise difference)

- **The 1st Factor**: ratio of the number of observations and sample size

$$\alpha = m/N$$

m: the number of observations
N: sample size

Previous work on direct sampling error (Ledoit and Wolf, 2004)

- Expected quadratic loss of sample covariance matrix $\widehat{\mathbf{D}}$

$$\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{D}}-\mathbf{D}\right\|_F^2\right]=\alpha(\mu^2+\theta)-\beta$$

- **The 2nd Factor**: square of the average size of the diagonal elements

$$\mu^2=[trace(\mathbf{D})/m]^2$$

The *trace* of a square matrix is defined to be the sum of its diagonal elements.

Previous work on direct sampling error (Ledoit and Wolf, 2004)

- Eigenvalue decomposition

$$\mathbf{D} = \mathbf{U}\,\mathrm{diag}\big(\lambda_1(\mathbf{D}),\ \lambda_2(\mathbf{D}), \cdots, \lambda_m(\mathbf{D})\big)\,\mathbf{U}^{\mathrm{T}}$$

$\mathbf{U} \in \mathbb{R}^{m\times m}$: matrix whose columns are the eigenvectors of $\mathbf{D}$
$\lambda_i(\mathbf{D})$: the i-th eigenvalue of $\mathbf{D}$

- Uncorrelated observation errors

$$\mathbf{\Gamma} = \mathbf{U}^{\mathrm{T}}[\boldsymbol{\epsilon}_1^o, \ldots, \boldsymbol{\epsilon}_N^o]$$

$\mathbf{\Gamma} \in \mathbb{R}^{m\times N}$: matrix of N observations on a system of m uncorrelated random variables that spans the same space as $[\boldsymbol{\epsilon}_1^o, \ldots, \boldsymbol{\epsilon}_N^o]$.

Previous work on direct sampling error (Ledoit and Wolf, 2004)

- Expected quadratic loss of $\widehat{\mathbf{D}}$

$$\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{D}}-\mathbf{D}\right\|_F^2\right]=\alpha(\mu^2+\theta)-\beta$$

- **The 3rd Factor**: variation in the mean size of the squared observation error between samples

$$\theta = Var\left[\frac{1}{m}\sum_{i=1}^{m}\gamma_{i1}^2\right],$$

γ_{i1}: the i-th element of the first column of $\mathbf{\Gamma}$

- θ is bounded as N goes to infinity.

Previous work on direct sampling error (Ledoit and Wolf, 2004)

- Expected quadratic loss of $\widehat{\mathbf{D}}$

$$\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{D}}-\mathbf{D}\right\|_F^2\right]=\alpha(\mu^2+\theta)-\beta$$

- **The 4th Factor**: the sum of the squares of the eigenvalues divided by m and N

$$\beta=\frac{1}{mN}\sum_{i=1}^{m}\lambda_i^2\left(\mathbf{D}\right)$$

- β converges to zero as N goes to infinity.

Our contribution to indirect sampling error

Let

$$\mathbf{W} = \mathbf{R}\left(\mathbf{HBH}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

Then we have

$$\mathbf{R} = \mathbf{WD} \quad (\mathbf{D} = \mathbf{R} + \mathbf{HBH}^{\mathrm{T}})$$

$$\widehat{\mathbf{R}} = \mathbf{W}\widehat{\mathbf{D}}$$

and we can write

$$\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{R}} - \mathbf{R}\right\|_F^2\right] = \frac{1}{m}\mathrm{E}\left[\left\|\mathbf{W}(\widehat{\mathbf{D}} - \mathbf{D})\right\|_F^2\right],$$

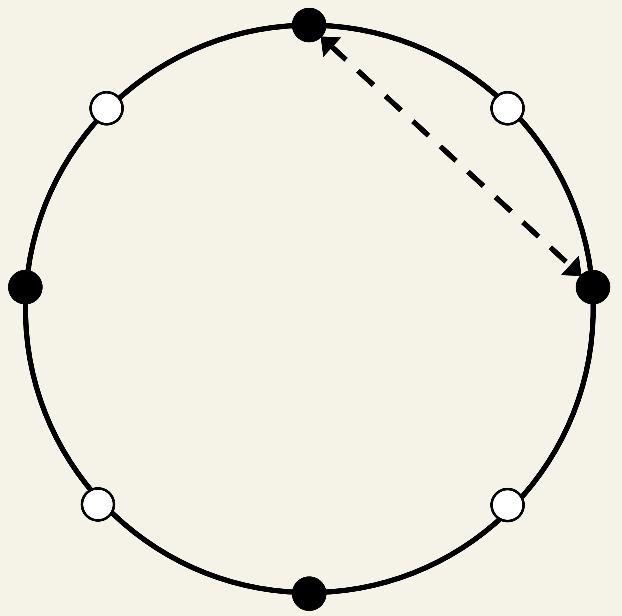
which satisfies the inequality

$$\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{R}} - \mathbf{R}\right\|_F^2\right] \leq s_1^2(\mathbf{W})\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{D}} - \mathbf{D}\right\|_F^2\right] = s_1^2(\mathbf{W})[\alpha(\mu^2 + \theta) - \beta]$$

$s_1^2(\mathbf{W})$: the square of the largest singular value of $\mathbf{W}$ (not necessarily symmetric)

Experimental design

- Equally spaced model grid points on a latitude circle on the Earth
- Observations at alternate grid points
- Chordal distance
- Error covariance modelling



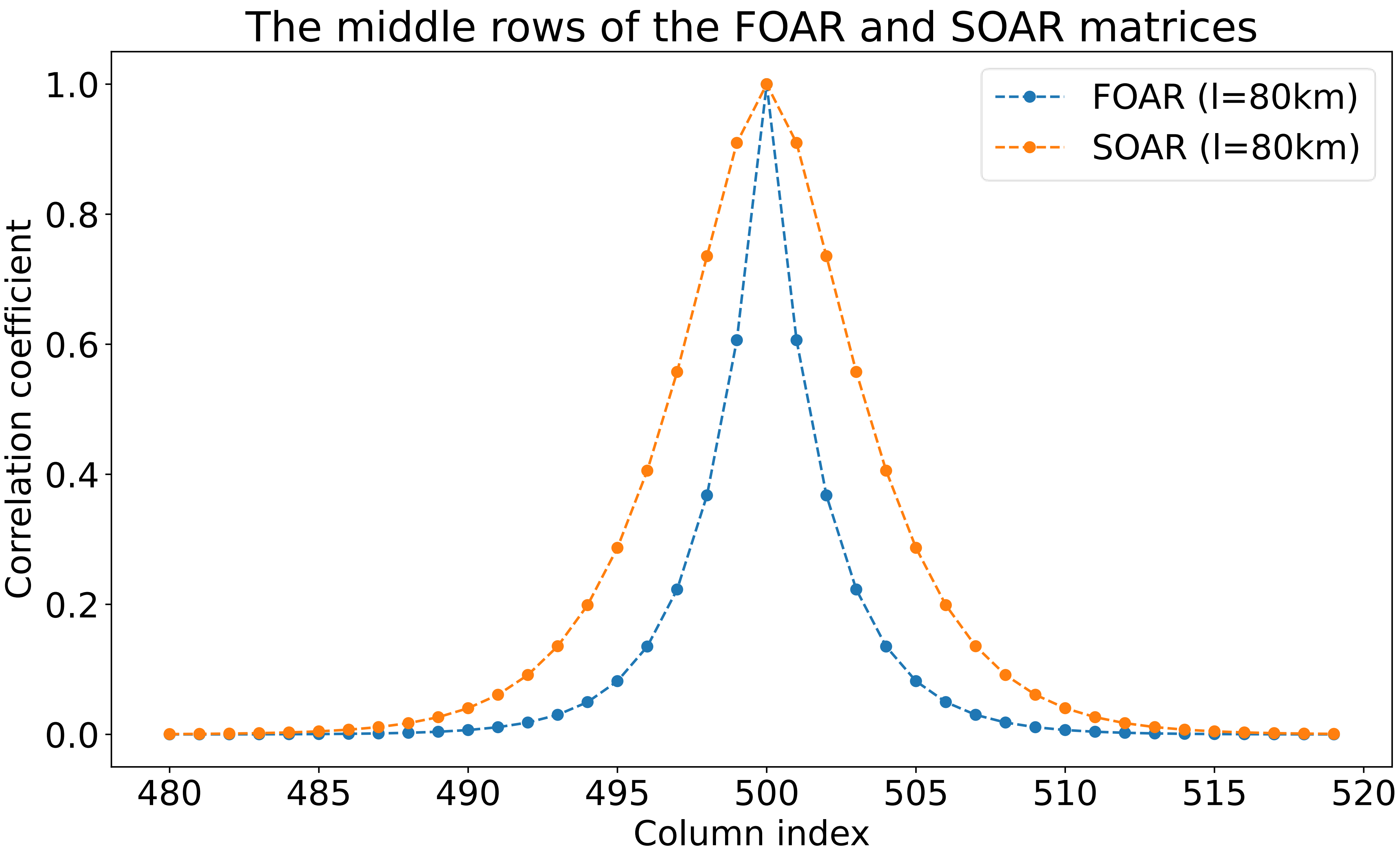
$$\mathbf{B} = \sigma_b^2 \mathbf{C}_{SOAR}(l_b)$$

$$\mathbf{R} = \sigma_o^2 \mathbf{C}_{FOAR}(l_o)$$

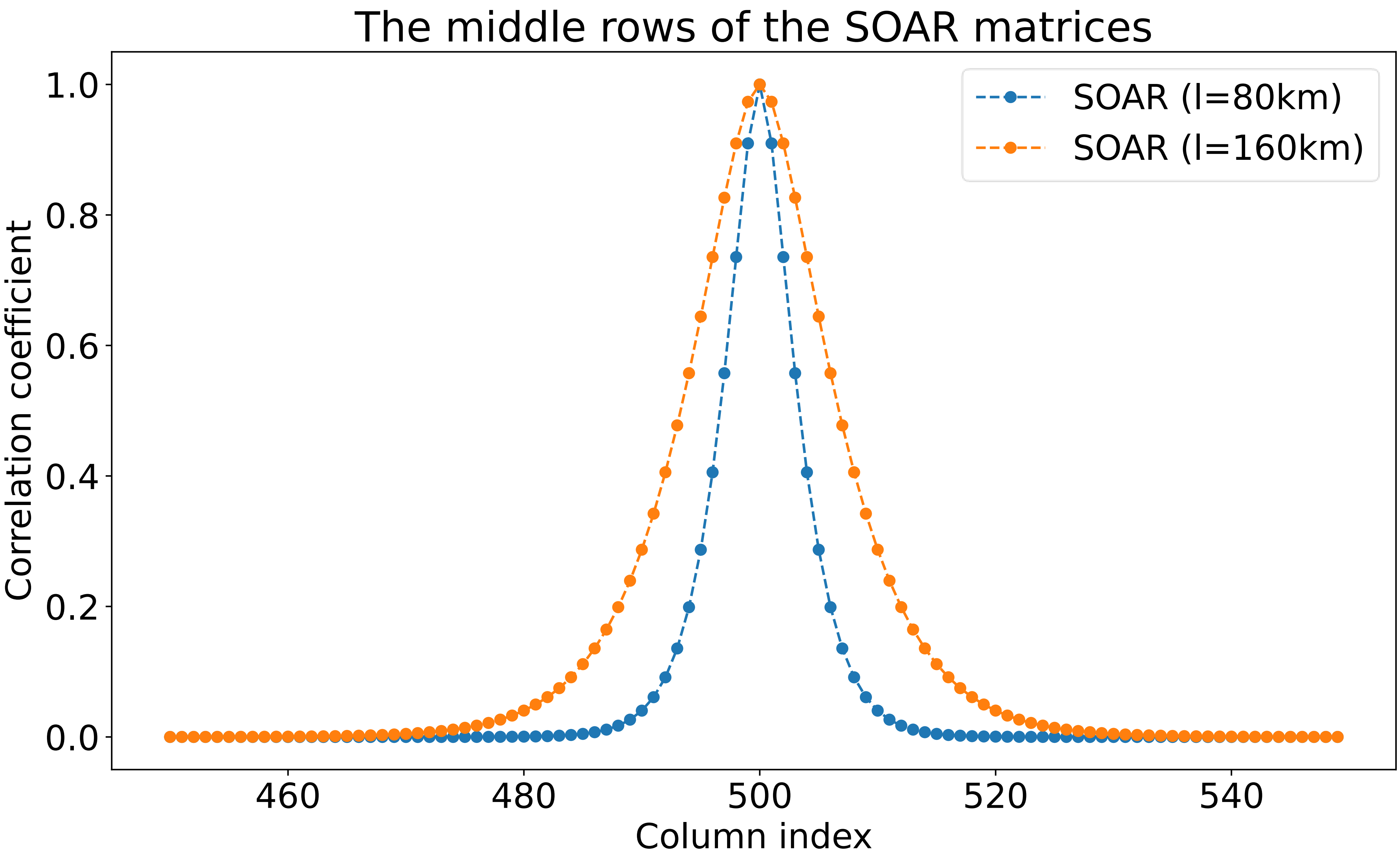
$\mathbf{C}_{FOAR}$ and $\mathbf{C}_{SOAR}$: first-order and second-order auto-regression correlation functions

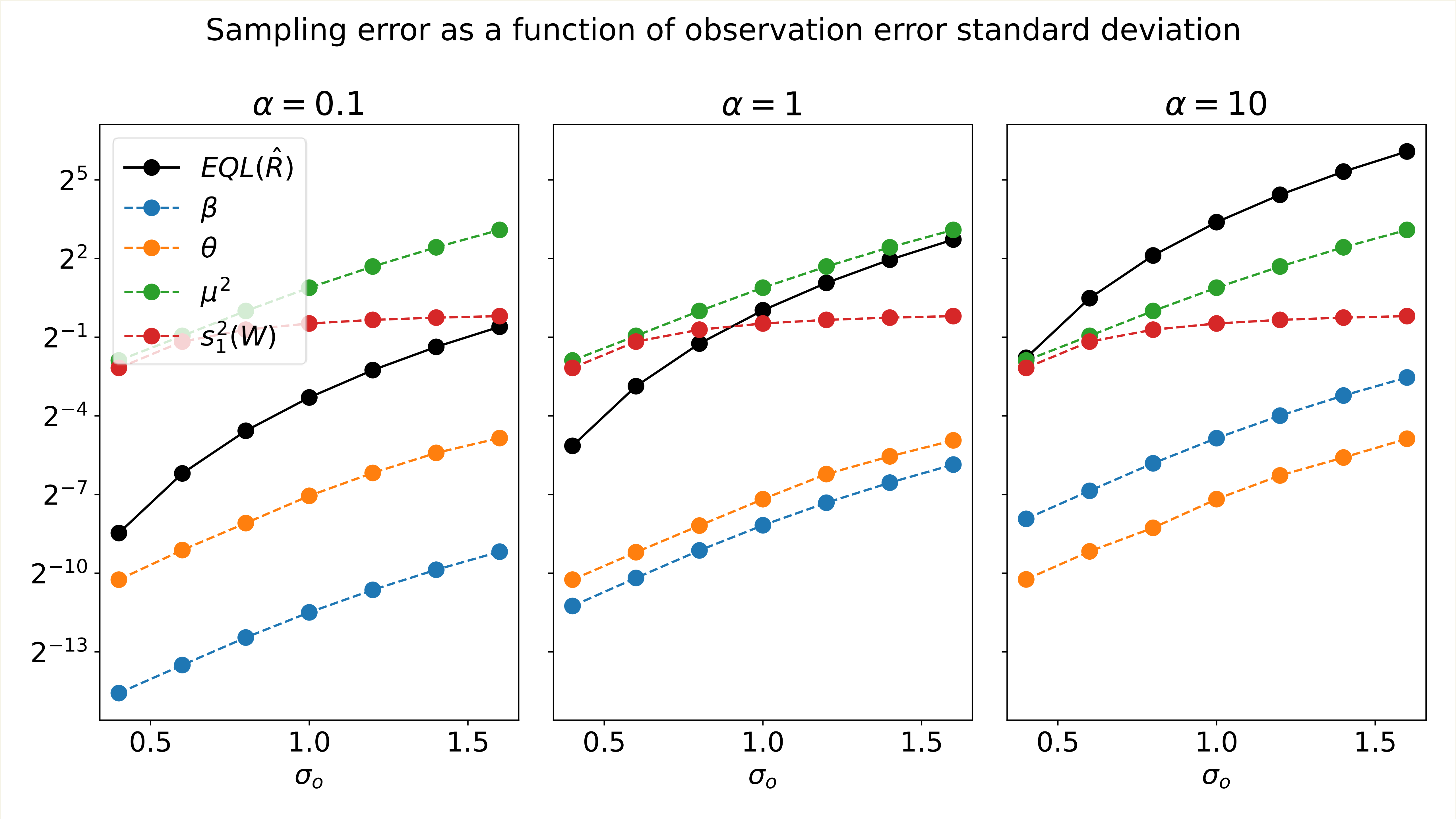
σ_b and σ_o: background and observation error standard deviations

l_b and l_o: background and observation error correlation lengthscales

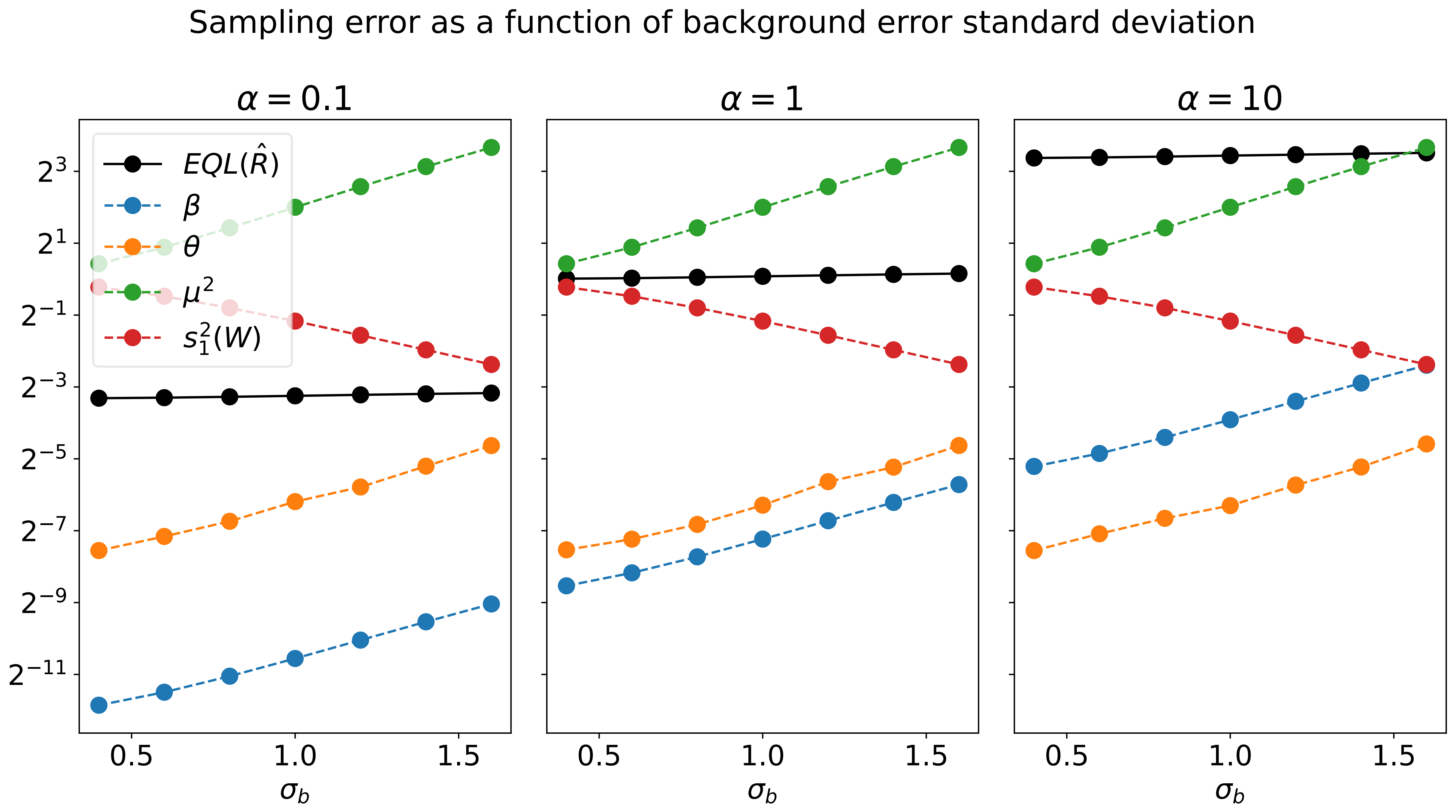


The middle rows of the FOAR and SOAR matrices
FOAR (l=80km)
SOAR (l=80km)
Correlation coefficient
Column index
0.0
0.2
0.4
0.6
0.8
1.0
480
485
490
495
500
505
510
515
520

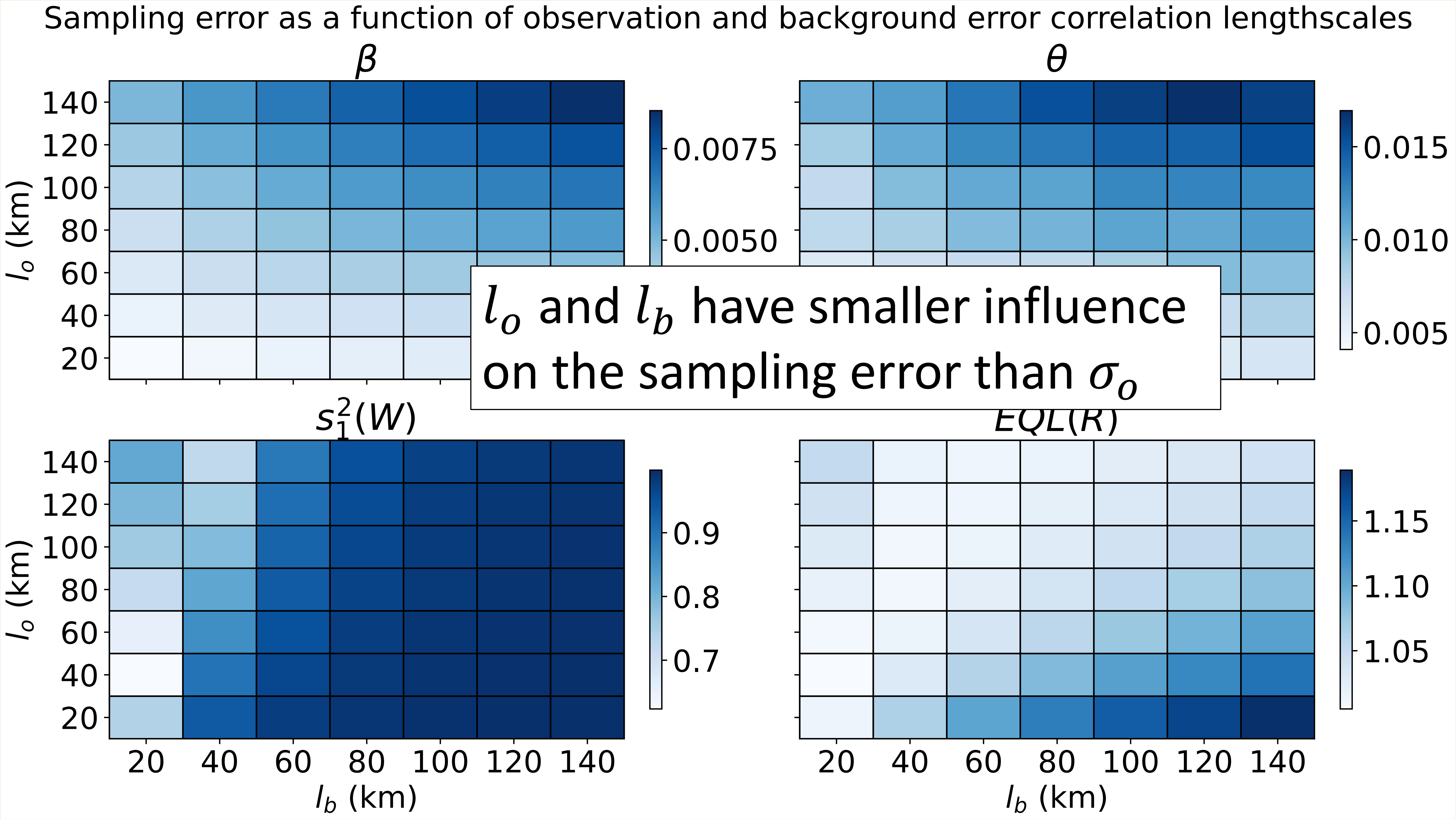

The middle rows of the SOAR matrices
SOAR (l=80km)
SOAR (l=160km)
Correlation coefficient
0.0
0.2
0.4
0.6
0.8
1.0
460
480
500
520
540
Column index



Sampling error as a function of observation error standard deviation
$\alpha = 0.1$
$\alpha = 1$
$\alpha = 10$
$EQL(\hat{R})$
β
θ
μ^2
$s_1^2(W)$
2^5
2^2
2^{-1}
2^{-4}
2^{-7}
2^{-10}
2^{-13}
0.5
1.0
1.5
σ_o



Sampling error as a function of background error standard deviation
$\alpha = 0.1$
$\alpha = 1$
$\alpha = 10$
$EQL(\hat{R})$
β
θ
μ^2
$s_1^2(W)$
2^3
2^1
2^{-1}
2^{-3}
2^{-5}
2^{-7}
2^{-9}
2^{-11}
0.5
1.0
1.5
σ_b


Sampling error as a function of observation and background error correlation lengthscales
β
θ
$s_1^2(W)$
$EQL(R)$
l_o (km)
l_b (km)
20
40
60
80
100
120
140
0.0050
0.0075
0.005
0.010
0.015
0.7
0.8
0.9
1.05
1.10
1.15
l_o and l_b have smaller influence on the sampling error than σ_o

Cross-sectional dispersion of sample eigenvalues

- The spread of the sample eigenvalues around the mean of the true eigenvalues

$$\delta = \mathbb{E}\left[\sum_{i=1}^{m} (\lambda_i(\widehat{\mathbf{R}}) - \mu_R)^2\right], \mu_R = \frac{1}{m}\sum_{i}^{m} \lambda_i(\mathbf{R}).$$

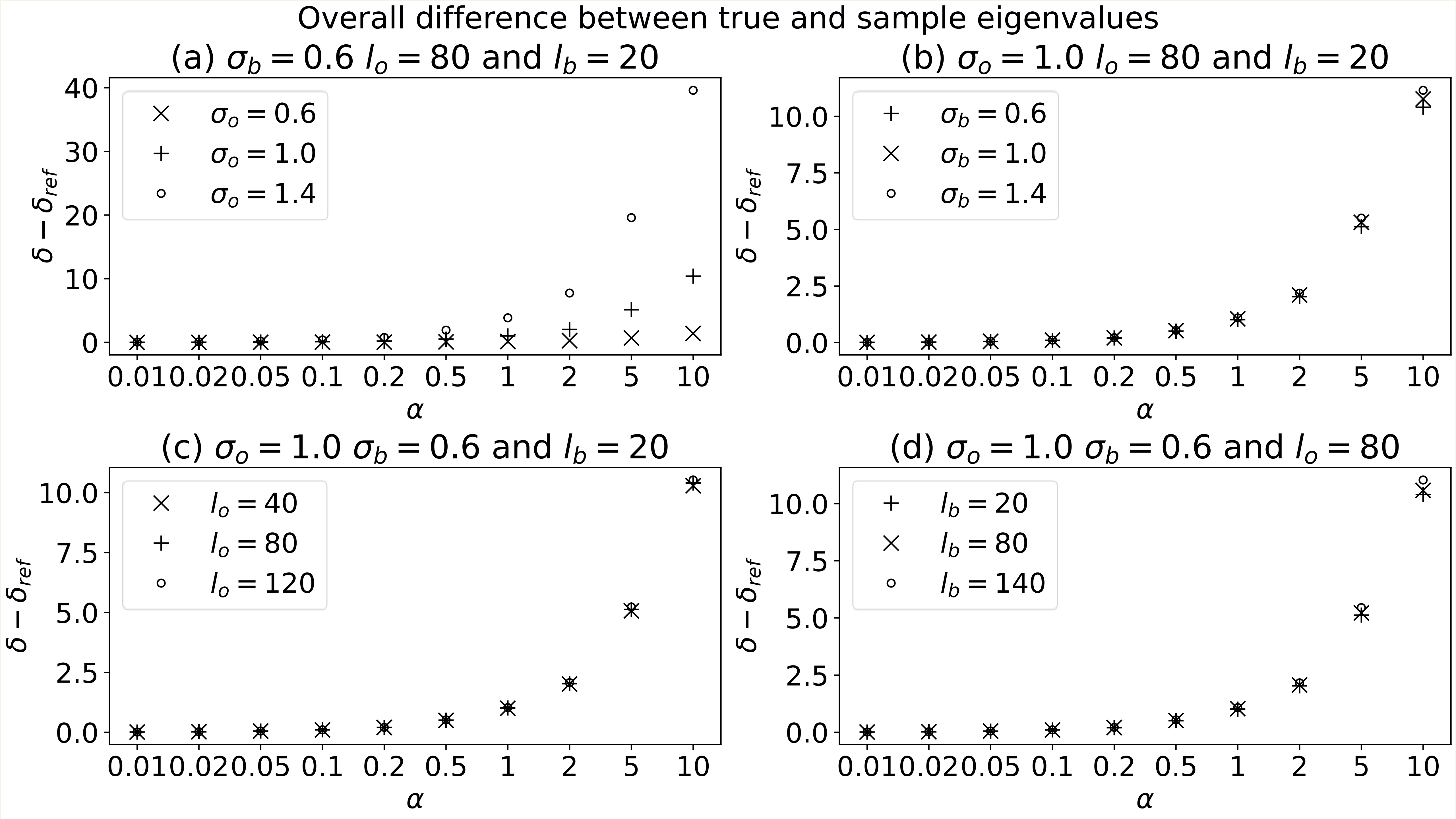
- The variance of the true eigenvalues

$$\delta_{ref} = \sum_{i=1}^{m} [\lambda_i(\mathbf{R}) - \mu_R)]^2$$

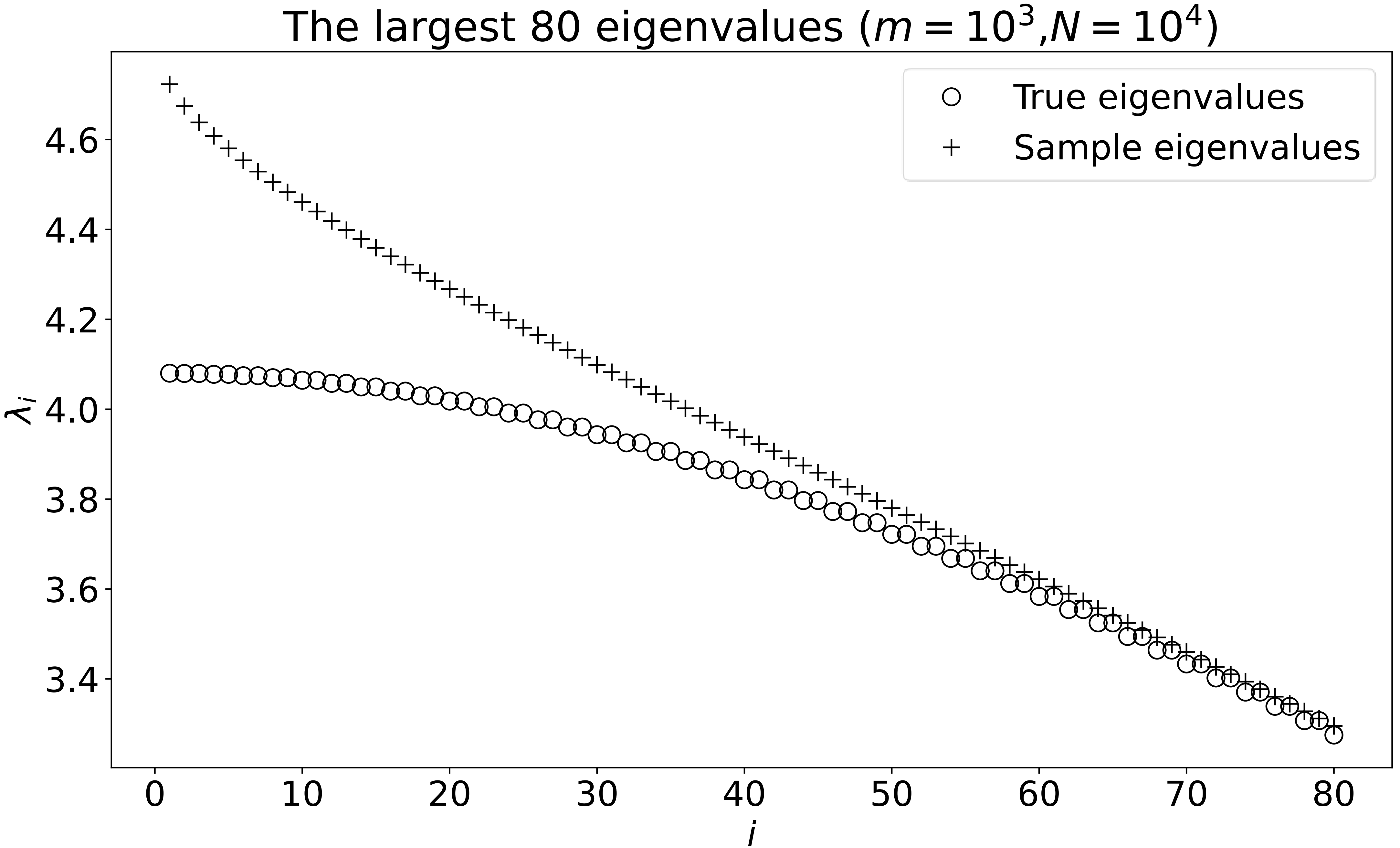
- Ledoit and Wolf (2004) showed that

$$\frac{1}{m}\mathrm{E}\left[\left\|\widehat{\mathbf{R}} - \mathbf{R}\right\|_F^2\right] = \frac{1}{\mathrm{m}}\left(\delta - \delta_{ref}\right),$$

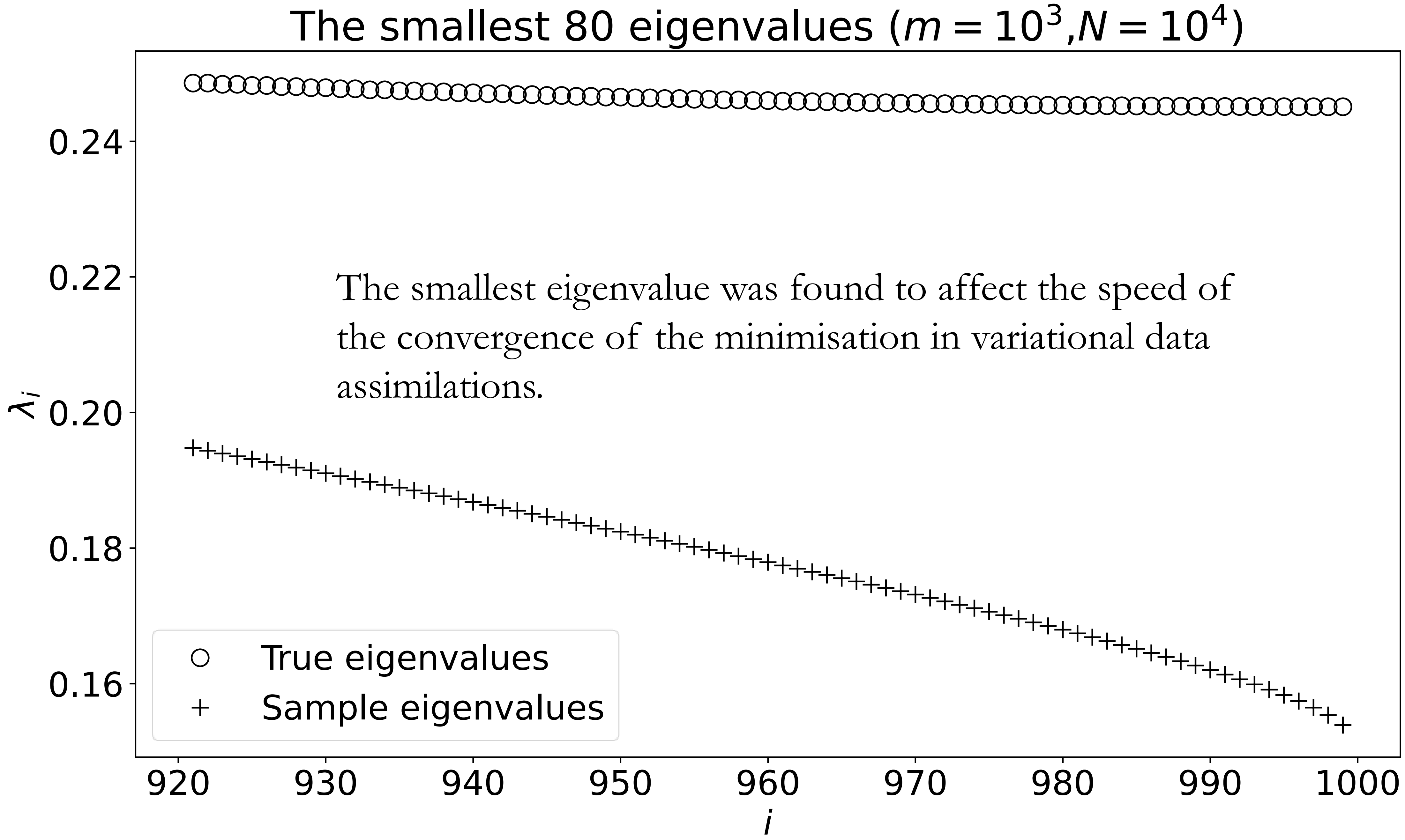
which indicates that δ converges to δ_{ref} as sampling error decreases.



Overall difference between true and sample eigenvalues
(a) $\sigma_b = 0.6$ $l_o = 80$ and $l_b = 20$
$\sigma_o = 0.6$
$\sigma_o = 1.0$
$\sigma_o = 1.4$
(b) $\sigma_o = 1.0$ $l_o = 80$ and $l_b = 20$
$\sigma_b = 0.6$
$\sigma_b = 1.0$
$\sigma_b = 1.4$
(c) $\sigma_o = 1.0$ $\sigma_b = 0.6$ and $l_b = 20$
$l_o = 40$
$l_o = 80$
$l_o = 120$
(d) $\sigma_o = 1.0$ $\sigma_b = 0.6$ and $l_o = 80$
$l_b = 20$
$l_b = 80$
$l_b = 140$
$\delta - \delta_{ref}$
α
0.01 0.02 0.05 0.1 0.2 0.5 1 2 5 10



The largest 80 eigenvalues ($m = 10^3$,$N = 10^4$)
True eigenvalues
Sample eigenvalues
4.6
4.4
4.2
4.0
3.8
3.6
3.4
λ_i
0
10
20
30
40
50
60
70
80
i



The smallest eigenvalue was found to affect the speed of the convergence of the minimisation in variational data assimilations.

Summary

- The sampling error is mainly affected by observation error standard deviation (compared to other error characteristics).
- Numerical results showed that the largest sample eigenvalues are greater than the true values and the smallest sample eigenvalues are smaller than the true values.
- Our results can provide guidance in deciding on appropriate sample sizes and choosing parameters for matrix reconditioning techniques.

Reference

- G. Desroziers, L. Berre, B. Chapnik, and P. Poli (2005). Diagnosis of observation, background and analysis-error statistics in observation space. QJRMS.
- O. Ledoit and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis.
- J. A. Waller, S. L. Dance, and N. K. Nichols (2016). Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. QJRMS.
- Janjić, T, Bormann, N, Bocquet, M, Carton, JA, Cohn, SE, Dance, SL, Losa, SN, Nichols, NK, Potthast, R, Waller, JA, Weston, P (2018). On the representation error in data assimilation, QJRMS.