Loss functions for the neural network based estimation of the local spectrum from sample band variances

Working paper

M Tsyrulnikov and A Sotskiy
HydroMetCenter of Russia

December 27, 2022

Goal

Training a neural network (NN) to recover the local spatial spectrum f_ℓ ($\ell = 0, 1, \dots, \ell_{\max}$) from the vector of local ensemble band variances $\widehat{v}_{(j)}$ ($j = 1, 2, \dots, J$) requires a *loss function*. The loss function $\mathcal{L}(f^{(1)}, f^{(2)})$ measures a difference between the two spectra, $f^{(1)}$ and $f^{(2)}$. During the NN training process, $\mathcal{L}(.,.)$ measures the deviation of the NN-generated spectrum $f^{\rm NN}$ from the "true" spectrum $f^{\rm true}$. Which loss function can be used? Which are more meaninful than others?

We suppress the dependence of the local spectrum f on the spatial point x for clarity of presentation (because the estimation is performed independently for different spatial points).

Below we list four candidate loss functions.

1 Naive ℓ^2 loss function

Just take

$$\boxed{\mathcal{L}_{\rm naive\,L2}(f^{(1)}, f^{(2)}) = \sum_{\ell=0}^{\ell_{\max}} (f_\ell^{(1)} - f_\ell^{(2)})^2.} \tag{1}$$

2 A bit more advanced ℓ^2 loss function

As spectra normally change drastically from low to high wavenumbers, it looks reasonable to apply the log transform before computing the ℓ^2 squared:

$$\boxed{\mathcal{L}_{\log\rm L2}(f^{(1)}, f^{(2)}) = \sum_{\ell=0}^{\ell_{\max}} (\log f_\ell^{(1)} - \log f_\ell^{(2)})^2.} \tag{2}$$

The log transform here makes the spectra to change only moderately within the wavenumber range.

3 A more educated ℓ^2 loss function

Consider a stationary process $\xi(x)$ on the sphere. Its spectral decomposition reads

$$\xi(x) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} \widetilde{\xi}_{\ell m} \, Y_{\ell m}(x), \tag{3}$$

where all $\widetilde{\xi}_{\ell m}$ are mutually uncorrelated random complex numbers such that $\mathbb{E}\,\widetilde{\xi}_{\ell m} = 0$ and $\mathbb{E}\,|\widetilde{\xi}_{\ell m}|^2 = f_\ell$. This allows us to rewrite Eq.(3) ass follows:

$$\xi(x) = \sum_{\ell=0}^{\ell_{\max}} \sqrt{f_\ell} \sum_{m=-\ell}^{\ell} \widetilde{\alpha}_{\ell m} \, Y_{\ell m}(x), \tag{4}$$

where all $\widetilde{\alpha}_{\ell m}$ are mutually uncorrelated random complex numbers such that $\mathbb{E}\,\widetilde{\alpha}_{\ell m} = 0$ and $\mathbb{E}\,|\widetilde{\alpha}_{\ell m}|^2 = 1$.

Now, generate two random processes that share the driving-noise random variables $\widetilde{\alpha}_{\ell m}$ but have different spectra, $f_\ell^{(1)}$ and $f_\ell^{(2)}$:

$$\xi^{(1)}(x) = \sum_{\ell=0}^{\ell_{\max}} \sqrt{f_\ell^{(1)}} \sum_{m=-\ell}^{\ell} \widetilde{\alpha}_{\ell m} \, Y_{\ell m}(x), \tag{5}$$

and

$$\xi^{(2)}(x) = \sum_{\ell=0}^{\ell_{\max}} \sqrt{f_\ell^{(2)}} \sum_{m=-\ell}^{\ell} \widetilde{\alpha}_{\ell m} \, Y_{\ell m}(x), \tag{6}$$

Finally, compute the *variance* of the difference between the two processes:

$$\mathsf{Var}\,(\xi^{(2)}(x) - \xi^{(1)}(x)) = \sum_{\ell=0}^{\ell_{\max}} \frac{2\ell+1}{4\pi} (\sqrt{f_\ell^{(2)}} - \sqrt{f_\ell^{(2)}})^2. \tag{7}$$

This variance can be used as a more justified variant of the ℓ^2 loss function

$$\boxed{\mathcal{L}_{\mathrm{var\,L2}}(f^{(1)}, f^{(2)}) = \sum_{\ell=0}^{\ell_{\max}} \frac{2\ell+1}{4\pi} (\sqrt{f_\ell^{(2)}} - \sqrt{f_\ell^{(2)}})^2.} \tag{8}$$

4 Custom loss function

Each of the above three loss functions can be used as a baseline to compare with a *problem-specific loss function*. The latter penalizes the *analysis error variance* for a hypothetical simplified analysis. This loss function looks most justified because minimizing the analysis error variance is precisely what we seek in data assimilation. See the article, Appendix F ("Spectral loss function").