Constraining the Multiscale Structure of Geophysical Fields in Machine Learning: The Case of Precipitation

Clément Guilloteau, Phong V.V. Le, and Efi Foufoula-Georgiou

***Abstract*— The use of deep-learning algorithms for estimating the value of geophysical variables from remotely sensed information is rapidly expanding. The typical objective function minimized in such algorithms is the mean square error (MSE), which is known to lead to smooth estimates with compressed dynamical range as compared to the true distribution of the variable of interest. Here, we introduce and evaluate alternative objective functions, focusing on the retrieval of precipitation rates from satellite passive microwave radiometric measurements using a deep convolutional neural network. For this testbed application, the results show that explicitly imposing the preservation of the statistical distribution and spatial wavelet power spectrum of the target variable allows to accurately reproduce extreme values and sharp gradients across multiple scales.**

***Index Terms*— Machine learning, objective function, precipitation, remote sensing, satellite.**

I. Introduction

In THE last two decades, the fields of image recognition and computer vision have been revolutionized by the emergence of deep-learning algorithms, with deep convolutional neural networks in particular, being popular and efficient at numerous tasks [1]. Inevitably, the field of geophysical remote sensing (i.e., estimating the value or state of geophysical variables from remote radiometric measurements) is also moving toward this class of algorithms. In many cases, these algorithms have already demonstrated superior performance when compared to older algorithms [2], [3]. An appealing aspect of the deep-learning framework is its flexibility, as the selection of the input data and of the hyperparameters of a neural network does not a priori restrict, which prediction task can be performed; the objective of the prediction is indeed solely determined by the objective function in deep learning. When the task is to predict the value of a physical variable (target) from a set of observed radiometric parameters (predictors), i.e., a regression task, supervised learning is necessary. Supervised learning relies on training data, for which, in addition to the predictors, the target values are also known a priori. In the supervised regression learning, the objective function optimized during the training measures the "distance" between the prediction and the a priori known "truth."

Geophysical variables of various types, such as atmospheric or surface temperature, wind vectors, soil or atmospheric moisture, vegetation density, ocean salinity, and precipitation rates, are not randomly distributed across space and time but exhibit a coherent spatial and temporal organization across a wide range of scales. In this respect, geophysical fields are similar to grayscale or colorscale images or to video flux: they contain "features," localized in space and time, which can be characterized by their size, orientation, shape, and texture. However, the preservation of these features and of their properties is not typically what remote-sensing retrieval algorithms, deep-learning algorithms in particular, are trained and optimized for. Indeed, in most cases, algorithms are optimized to minimize a "pixelwise" distance metric (i.e., a metric that only considers differences between the prediction and the target in coincident pixels); the most commonly used objective function in deep-learning retrieval algorithms for geophysical remote sensing is the mean squared error (MSE) or alternatively the mean absolute error (MAE).

While an infinitely small MSE (or MAE) obviously guarantees the preservation of all features in a geophysical field, such an objective is often unattainable in geophysical remote sensing, in fact, it is not uncommon for the MSE to be in the same order of magnitude as the variance of the target variable. For example, in the case of retrieval of precipitation rates from passive satellite radiometric measurements at scales finer than 12 h and 100 km, comparisons with high-accuracy ground-based measurements indicate that the state-of-the art products can rarely resolve more than 75% of the ground-observed variance and often around or less than 50% [4], [5]. When the residual errors cannot be reduced to a negligible quantity, MSE-optimal estimation algorithms tend to favor smooth solutions and compress the dynamical range of the retrieved variable [6].

In this letter, we propose novel alternatives to the MSE objective function used "by default" in the training of most of the deep-learning algorithms. The application we present as demonstration is the estimation of surface precipitation rates from satellite passive microwave measurements, a data-rich but highly underdetermined retrieval problem [7], [8] of high interest to the geoscience community [9]. Our focus is to preserve the statistical distribution of precipitation rates (especially the extreme values in the tail of the distribution) and the multiscale variance of the precipitation fields, which

Manuscript received 30 January 2023; revised 29 April 2023; accepted 26 May 2023. Date of publication 8 June 2023; date of current version 22 June 2023. This work was supported in part by the NASA through the Global Precipitation Measurement Program under Grant 80NSSC19K0684 and Grant 80NSSC22K0597 and in part by the National Science Foundation (NSF) through the TRIPODS-X Program under Grant DMS-1839336. *(Corresponding author: Clément Guilloteau.)*

Clément Guilloteau is with the Civil and Environmental Engineering Department, University of California at Irvine, Irvine, CA 94143 USA (e-mail: cguillot@uci.edu).

Phong V.V. Le is with the Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA (e-mail: lepv@ornl.gov).

Efi Foufoula-Georgiou is with the Department of Civil and Environmental Engineering and the Department of Earth System Science, University of California at Irvine, Irvine CA 94143 USA (e-mail: efi@uci.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/LGRS.2023.3284278, provided by the authors.

Digital Object Identifier 10.1109/LGRS.2023.3284278

relates directly to the covariance function and to the variogram. While a few published studies have already demonstrated the utility of constraining the statistical distribution of precipitation rates [10], we show that doing so at only one scale is insufficient and that a multiscale constraint (applied here through the spatial wavelet power spectrum) allows a more comprehensive preservation of the statistical properties of the field at multiple scales, including its spatial covariance.

II. Data and Neural Network Architecture

A mixed-scale dense convolutional neural network (MSDNet) [11] is used here to perform the retrieval of instantaneous surface precipitation rates over oceans. The inputs of the MSDNet are scenes of brightness temperatures, measured by the passive Global Precipitation Measurement (GPM) Microwave Imager (GMI) onboard the GPM Core Observatory satellite. Each input scene consists of the superposition of 13 radiometric images corresponding to the 13 GMI channels and covers a 245 × 245 km area. Each radiometric image is composed of 931 fields of view (overlapping pixels), corresponding to 49 scan positions and 19 consecutive GMI scans. GMI is a conical-scanning imager, and its channels, covering the microwave spectrum between 10 and 183 GHz, comprise five horizontal-polarization single-band channels (at 11, 19, 37, 89, and 166 GHz), six vertical-polarization single-band channels (at 11, 19, 23, 37, 89, and 166 GHz), and two vertical-polarization double-sideband channels (at 183 ± 3 and 183 ± 7 GHz) [12]. The brightness temperatures measured by GMI at the top of the atmosphere are essentially the outcome of the surface emission, emission and absorption by water vapor and liquid rain drops, and scattering by ice particles in the atmosphere [13].

The objective is, from the GMI observations only, to predict the 5-km-resolution surface precipitation fields as is currently retrieved by the NASA GPM combined radar-radiometer algorithm (CORRA [14]). CORRA relies not only on the passive GMI observations but also on the measurement from the active dual-frequency precipitation radar (DPR), also onboard the GPM Core observatory satellite. While the radar-radiometer combined measurements are known to allow much more accurate estimates of precipitation rates than those made only from passive radiometric measurements, the DPR is currently the only scanning precipitation radar orbiting the Earth and has a swath width of only 245 km and average revisit time of more than 3 days. Comparatively, GMI has an 885-km swath width, and, in combination with several other similar passive instruments orbiting the Earth such AMSR-2 and the SSMI/S series, it can achieve complete coverage of the globe within a few hours average revisit time [15], hence the interest in providing accurate precipitation measurements from passive microwave imagers only. We note that GMI's scanning pattern provides a 5-km sampling distance in the along-scan direction and a 13-km sampling distance in the cross-scan direction. Therefore, in the cross-scan direction, the 5-km target resolution of the MSDNet prediction is higher than the instrumental resolution of the inputs, which amplifies the underconstrained nature of the prediction task. Supervised training of the MSDNet is performed with a training database consisting of 7000 scenes of GMI-observed brightness temperatures over global oceans, collocated with CORRA surface precipitation rates. The MSDNet used in this study is composed of 70 densely connected convolution layers (see supplemental online material for more details about the MSDNet architecture and the data). The retrieval performance is tested on a dataset of 1000 unique oceanic precipitation scenes completely independent of the training dataset.

III. Objective Functions

The objective function in deep learning algorithms (also referred to as loss or cost function) measures the similarity between the prediction and the target during the training. It can be as simple as the MSE or MAE or be constructed as a combination of several distance metrics (it may also contain one or several regularization terms). Here, we present a few distance metrics that can be used in combination with or as alternative to the MSE to build objective functions that efficiently preserve the multiscale statistical distribution of precipitation fields and avoid the smoothing effect inherent to MSE-optimal estimates.

Let us consider a target tensor $\mathbf{T}$ and its prediction $\mathbf{P}$. The first distance we consider is the MSE, empirically computed as the mean squared distance between $\mathbf{P}$ and $\mathbf{T}$, which is the square of the Euclidian ($L2$) distance divided by the tensor length

$$D_1(\mathbf{P}, \mathbf{T}) = \frac{1}{n}\sum_{i=1}^{n}(p_i - t_i)^2 \tag{1}$$

where n is the number of elements in the N-dimensional tensors $\mathbf{P}$ and $\mathbf{T}$, p_i is the ith element of the tensor $\mathbf{P}$, and t_i is the ith element of the tensor $\mathbf{T}$. The MSE can be decomposed as follows:

$$D_1(\mathbf{P}, \mathbf{T}) = (\mu_P - \mu_T)^2 + \sigma_T^2 + \sigma_P^2 - 2\mathrm{Cov}(\mathbf{P}, \mathbf{T}) \tag{2}$$

where μ_T, μ_P, σ_T^2, and σ_P^2 are, respectively, the mean and variance of $\mathbf{P}$ and $\mathbf{T}$, and $\mathrm{Cov}(\mathbf{P}, \mathbf{T})$ denotes the covariance between $\mathbf{P}$ and $\mathbf{T}$. For any properly optimized minimal-MSE estimator, the estimation should be unbiased (at least over the training data); therefore, the first term $(\mu_P - \mu_T)^2$ should be close to null at the end of the training. For a "perfect" retrieval, $\sigma_T^2 = \sigma_P^2 = \mathrm{Cov}(\mathbf{P}, \mathbf{T})$, and therefore, $D_1 \approx 0$. However, when residual errors are not negligible, the optimal-MSE solution is achieved with $\mathrm{Cov}(\mathbf{P}, \mathbf{T}) = \sigma_P^2$ and $\sigma_P^2 < \sigma_T^2$, hence the "compressing" effect of MSE-optimal estimators, with the variance of the prediction being lower than the variance of the target.

To avoid this compressing effect, we propose an alternative distance metric between $\mathbf{P}$ and $\mathbf{T}$, defined as follows:

$$D_2(\mathbf{P}, \mathbf{T}) = (\mu_P - \mu_T)^2 + \left|\sigma_T^2 - \sigma_P^2\right| + \sigma_T^2(1 - \mathrm{CC}(\mathbf{P}, \mathbf{T})) \tag{3}$$

where $\mathrm{CC}(\mathbf{P}, \mathbf{T})$ denotes the linear Pearson correlation coefficient between $\mathbf{P}$ and $\mathbf{T}$. This distance is minimized when $\mu_P = \mu_T$ and $\sigma_T^2 = \sigma_P^2$ and when $\mathrm{CC}(\mathbf{P}, \mathbf{T})$ is as close as possible to the unit value. We note that the Pearson correlation coefficient is agnostic to any linear distortion of $\mathbf{P}$. We also note that the terms that compose our proposed D_2 distance metric (mean, variance, and correlation coefficient) are somewhat similar to the terms that define the Kling–Gupta efficiency [16] coefficient, except that in our case, the quantities are not rendered dimensionless and that the optimal value is 0 instead of 1.

If the preservation of the whole statistical distribution of the target variable (beyond its mean and variance as described above) is strongly desired, a specific constraint to achieve this might be added. The consistency of two empirical probability density functions (PDF) or cumulative distribution functions

(CDF) can always be measured through a distance metric. In the case of precipitation, we are particularly interested in preserving the tail of the distribution because of the hydrological and climatological importance of precipitation extremes. The precipitation rates corresponding to the 99% or 99.9% percentiles of the distribution even if relatively infrequent (occurring over a few hours per year on average) are quasi-systematically associated with impactful weather systems (mostly hurricanes and other deep convective storms). For this reason, to represent the statistical distribution of precipitation in the present study, we chose to rely on a modified distribution function, which we designate as the cumulative contribution function (CCF) and which puts more emphasis on extremes as compared to the CDF. Let us consider a variable X and its PDF f_X. The CDF of X is defined as follows:

$$F_X(v) = \int_{-\infty}^{v} f_x(u)du. \tag{4}$$

The CDF is related to the probability of exceedance (PoE) function P_X with $P_X(v) = 1-F_X(v)$. When X is a nonnegative variable, as is the case for precipitation rates, we define the CCF of X as follows:

$$\Phi_X(v) = \frac{1}{E[X]} \int_0^v u \times f_x(u)du \tag{5}$$

where $E[X]$ is the expected value of X defined as $E[X] = \int_0^\infty u \times f_x(u)du$. With such a definition, the value $\Phi_X(v)$ represents the relative contribution of precipitation rates between 0 and v to the mean value of X. The CCF, while it has not been rigorously formalized and has been given several different names, has been used in the literature to represent and analyze statistical distributions of precipitation rates [17]. We note again that the CCF only makes sense if the variable of interest takes only nonnegative values and is a quantity that can be "accumulated" over time. In this letter, we measure distance between the CCF of the target $\mathbf{T}$ and prediction $\mathbf{P}$ tensors through the standard continuous $L1$ distance and define

$$D_3(\mathbf{P}, \mathbf{T}) = \int_0^\infty |\Phi_P(v) - \Phi_T(v)|dv. \tag{6}$$

While preserving the statistical distribution of the target variable at the scale at which the retrieval is performed (the "pixel" scale) is often desired, it does not guarantee at all the preservation of the statistical distribution at any other scale. A constraint relying on a statistical distance such as the D_3 distance may be applied several times at different scales, however, selecting the scales at which to apply it may be cumbersome and, eventually, a large number of constraints in a deep-learning algorithms may render the convergence of the training run more difficult. An alternative to constraining the full statistical distribution at multiple scales is to constrain the multiscale variance only. The multiscale variance of a field is determined by its spatial autocorrelation function or alternatively by its Fourier or wavelet power spectral density (PSD). As for the distribution functions, distance metrics can be used to evaluate the consistency between two autocorrelation functions or two Fourier or wavelet PSD functions. In this letter, we choose to rely on the discrete wavelet PSD obtained from a 2-D Haar wavelet decomposition. Discrete wavelet PSDs are generally easier to handle than the continuous Fourier PSD or autocorrelation function, particularly in multidimensional spaces. Moreover, they are fast to compute, can be robustly estimated from relatively small sample sizes, and allow easy handling of edge effects.

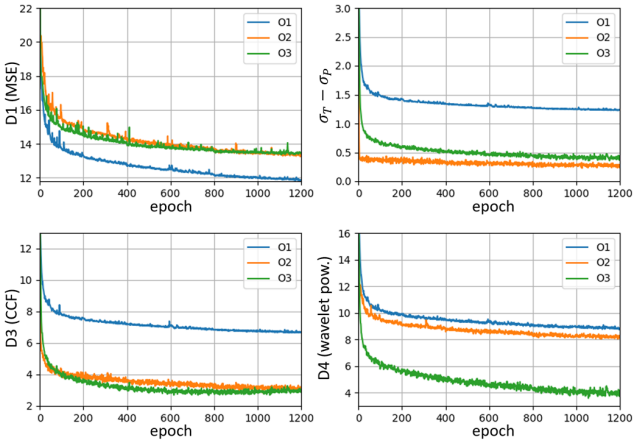

Fig. 1. Evolution of the performance metrics (distances D_1, D_3, and D_4 between the prediction and the target, and difference between the standard deviation of the prediction and of the target) during the training of the MSDNet for different objective functions O_1, O_2, and O_3. At each epoch, the performance metrics are computed over the 7000 scenes of the training dataset.

Let $S_X(\lambda, \omega)$ be the Haar wavelet power (i.e., the mean of the squared wavelet coefficient) of the N-dimensional variable X at the scale λ and in the direction ω. We measure the distance between the discrete wavelet power spectra of the tensors $\mathbf{P}$ and $\mathbf{T}$ through the standard discrete $L1$ distance

$$D_4(\mathbf{P}, \mathbf{T}) = \sum_{i=1}^{J} \sum_{k=1}^{M} |S_P(\lambda_i, \omega_k) - S_T(\lambda_i, \omega_k)| \tag{7}$$

where λ_i and ω_k are, respectively, the discrete scales and directions of an N-dimensional discrete wavelet transform of depth J. In N dimensions, the number of directions of a discrete wavelet transform is $M = 2^N - 1$. In this letter, we apply the wavelet transform on 2-D precipitation fields ($N = 2$ and $M = 3$) and with $J = 5$. We use the distances D_1, D_2, D_3, and D_4 to construct three different objective functions O_1, O_2, and O_3 to be optimized during the training of the MSDNet

$$O_1 = D_1(\mathbf{P}, \mathbf{T}) \tag{8}$$

$$O_2 = D_2(\mathbf{P}, \mathbf{T}) + \gamma_3 D_3(\mathbf{P}, \mathbf{T}) \tag{9}$$

$$O_3 = D_2(\mathbf{P}, \mathbf{T}) + \gamma_3 D_3(\mathbf{P}, \mathbf{T}) + \gamma_4 D_4(\mathbf{P}, \mathbf{T}) \tag{10}$$

where γ_3 and γ_4 are the weights used to control the relative influence of the different terms in the cost functions. Here, we take $\gamma_3 = 3$ and $\gamma_4 = 1$. The same MSDNet is trained three times with the three different objective functions, while the training data and hyperparameters are kept the same for the three training runs.

IV. Results

The metric distances D_1, D_2, D_3, and D_4 between the prediction and the target along with other performance metrics and statistics are computed over the 7000 scenes of the training dataset at each epoch of the training of the MSDNets, not only to allow the optimization of the objective function but also to monitor the efficiency of the training and its convergence toward a stable solution. The evolutions during the three training runs of the distances D_1, D_3, and D_4, along with standard deviation of the predicted precipitation rates minus that of the target precipitation rates, are shown

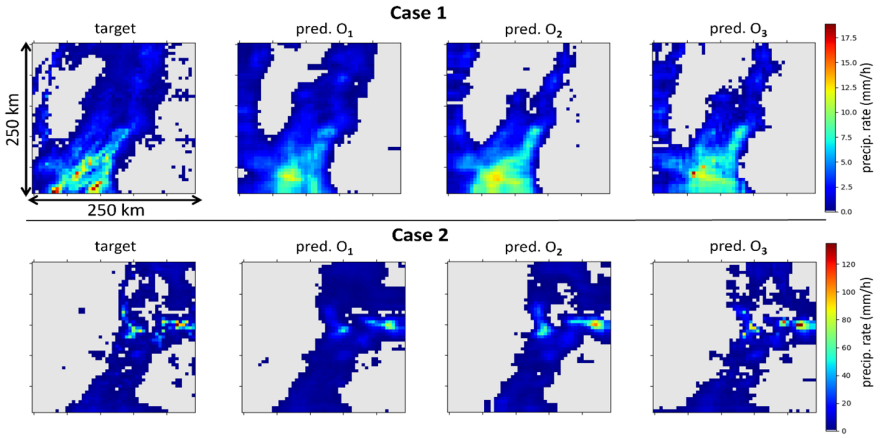

Fig. 2. Two cases of precipitation scenes of the testing dataset retrieved by the CORRA algorithm (target) and predicted from GMI brightness temperature only by the MSDNets trained with different objective functions O_1, O_2, and O_3. While the MSDNet trained with O_1 (MSE minimization) produces spatially smooth estimations and compresses the high precipitation rates, the MSDNet trained with O_3 (with constraints on the CCF and on the wavelet PSD) preserves the dynamical range of precipitation rates and the texture of the target fields.

TABLE I

PERFORMANCE OF THE MSDNETS TRAINED WITH THE THREE DIFFERENT OBJECTIVE FUNCTIONS O_1, O_2, AND O_3.

	Pred. O1	Pred. O2	Pred. O3
CC	0.77	0.77	0.76
D1 (MSE)	12.8 mm^2h^2	14.3 mm^2h^2	14.8 mm^2h^2
D3 (CCF)	6.67	3.72	1.71
D4 (wavelet PSD)	11.20	8.54	2.89

The performance metrics are the coefficient of linear correlation between the MSDNet prediction and the CORRA estimates (target), and the distances D1, D3 and D4 between the MSDNet prediction and the CORRA estimates. All the performance metrics are computed for the 1000 scenes of the testing dataset.

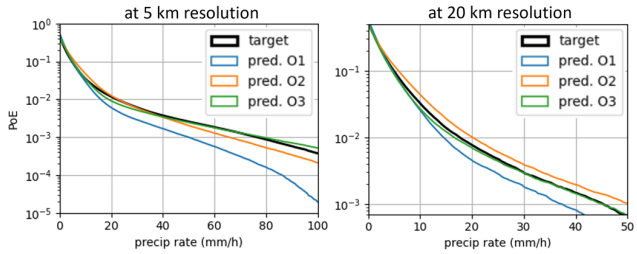

Fig. 3. PoE of the precipitation rates predicted by the three MSDNets trained with the O_1, O_2, and O_3 objective functions, along with the PoE for the CORRA algorithm (target), at 5-km resolution, and aggregated at 20-km resolution. All PoEs are computed over the 1000 scenes of the testing dataset.

in Fig. 1. As expected, the training run with the objective function $O_1 = D_1$ converges to a lower MSE value (around 12 mm$^2 \cdot$ h^2) than the other two runs (around 13.5 mm$^2 \cdot$ h^2). However, the runs with the objective functions O_2 and O_3, which both comprise a constraint on the statistical distribution of the rain rates, converge to a significantly lower D_3 value (distance to the target CCF) and to a standard deviation of the prediction closer to that of the target. The D_4 distance between the wavelet power spectra decreases much more rapidly and reaches a much lower value with the O_3 objective function than with the other two objective functions O_1 and O_2.

Focusing now on the performance of the fully trained MSDNets over the testing dataset, Table I shows the coefficients of linear correlation and the distances D_1, D_3, and D_4 between the three MSDNet predictions and the CORRA reference estimates. Again, as expected, the MSDNet trained with the O_1 objective function achieves the lowest MSE; however, all three MSDNet predictions show comparable coefficients of linear correlation with the CORRA estimates (0.76–0.77). When it comes to the preservation of the statistical distribution (CCF) and of the wavelet power spectrum, the MSDNet trained with the O_3 objective function shows dramatically better performance than the other two. It is worth noting that, with the constraint on the CCF only (objective function O_2), the preservation of the multiscale structure as quantified by the wavelet power spectrum is only marginally improved as compared to the MSE-optimal prediction (objective function O_1). Fig. 2 shows the three MSDNet predictions for two selected scenes among the 1000 scenes of the testing dataset. When visually comparing the predictions to the CORRA reference fields, the smoothness of the O_1 prediction is salient. The O_2 prediction produces higher local maxima, effectively showing a greater dynamical range than the O_1 prediction. However, only the O_3 prediction reproduces the sharp spatial gradients and the texture of the CORRA fields.

Fig. 3 shows the statistical distribution of the precipitation rates, through the PoE, at 5- and 20-km resolutions for the three MSDNet predictions over the testing data. As compared to CORRA, the O_1 prediction underestimates the occurrence of precipitation rates higher than 15 mm/h both at the 5- and 20-km resolutions. At 5-km resolution, the O_2 and O_3 predictions better represent high precipitation rates as compared to the O_1 prediction. At the 20-km resolution, the O_3 prediction well reproduces the statistical distribution of precipitation rates over the full dynamical range; the O_2 prediction, however, appears to overestimate the occurrence of precipitation rates higher than 6 mm/h. This reveals that a constraint on the statistical distribution applied at the pixel level only, without any control on the multiscale variance, can eventually lead to a deteriorated statistical distribution at a coarser aggregated scale. In this case, the overestimated occurrence of high precipitation rates at the 20-km resolution comes from the excessive short-range spatial correlation of the O_2 prediction.

Fig. 4 (left) shows the PoE of the spatial gradients, computed as the absolute values of the differences between two adjacent pixels at 5-km resolution, for the three MSDNet predictions. We can see that the O_3 prediction much better reproduces the distribution of the CORRA gradients as compared to the O_1 and O_2 predictions, particularly for the

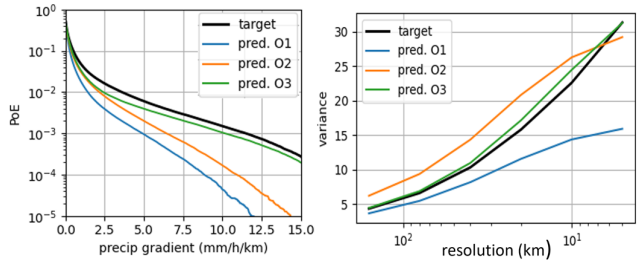

Fig. 4. (Left) PoE of the absolute values of the spatial precipitation gradients (differences between two adjacent pixels at 5-km resolution) predicted by the three MSDNets trained with the O_1, O_2, and O_3 objective functions, along with the PoE for the CORRA algorithm (target). (Right) Multiscale variance of the precipitation fields predicted by the three MSDNets trained with the O_1, O_2, and O_3 objective functions for the testing dataset, along with the multiscale variance of the precipitation fields from the CORRA algorithm (target). The PoEs and multiscale variances are computed over the 1000 scenes of the testing dataset.

strongest gradients (higher than 3 $\text{mm}\cdot\text{h}^{-1}\cdot\text{km}^{-1}$). In spite of being trained with a constraint on the statistical distribution of precipitation rates, the O_2 prediction only shows a marginal improvement in terms of the distribution of spatial gradients as compared to the O_1 MSE-optimal prediction. The right panel of Fig. 3 shows the multiscale variance of the different predictions. The O_1 prediction underestimates the variance of precipitation rates at every scale, and the variance deficit gets stronger with finer scales. As mentioned before, this multiscale variance-compression effect is inevitable for MSE-optimal predictions. The O_3 prediction nearly perfectly reproduces the CORRA multiscale variance, which is not surprising as the preservation of the multiscale variance has been explicitly constrained through the wavelet PSD during the training. While the O_2 prediction slightly underestimates the variance at the finest 5-km scale, it overestimates it at every other coarser scale, which again reveals the excessive spatial correlation of the O_2 prediction (the decrease rate of the variance with spatial aggregation is lower for fields that have stronger spatial correlation).

V. Conclusion

With the example of the retrieval of surface precipitation rates from passive microwave satellite measurements with a deep convolutional neural network, we illustrated the fact that MSE-optimal estimates are generally smoother than the true target variable, with compressed dynamical range, reduced variance, and underrepresentation of extremes, at all scales, but more so at the finest scales. This side effect of the MSE minimization is particularly salient for precipitation estimation from passive measurements, because of the strongly skewed heavy-tail distribution of rain rates at subhourly subdegree resolutions, and because of the relatively large magnitude of the residual errors inherent to the limited information content of the measured radiances. With deep-learning algorithms, this smoothing effect can be avoided by explicitly constraining the statistical distribution of the target variable and its multiscale structure through a discrete wavelet PSD in the objective function during the training. With such constraints, the statistical distribution, texture, and multiscale variance of the variable of interest are significantly better preserved, at the cost of a marginally increased MSE (about 10%–15% increase in our example, see Table I). It must be noted that the improvement is moderate when only the statistical distribution at the pixel scale is constrained, while the addition of the constraint on the wavelet PSD allows a much greater improvement. More generally, for any deep-learning application, the results presented here foster the use of purposedly designed objective functions, adapted and tuned for each specific prediction task, over the reliance on generic "all-purpose" objective functions such as the MSE.

References

[1] E. Akleman, "Deep learning," *Computer*, vol. 53, no. 9, p. 17, Sep. 2020, doi: 10.1109/MC.2020.3004171.

[2] S. Pfreundschuh, P. J. Brown, C. D. Kummerow, P. Eriksson, and T. Norrestad, "GPROF-NN: A neural-network-based implementation of the Goddard profiling algorithm," *Atmos. Meas. Techn.*, vol. 15, no. 17, pp. 5033–5060, Sep. 2022, doi: 10.5194/amt-15-5033-2022.

[3] R. Rahimi, S. Vahedizadeh, and A. Ebtehaj, "A deep learning architecture for passive microwave precipitation retrievals using CloudSat and GPM data," 2022, *arXiv:2212.02236*.

[4] P. Kirstetter, N. Karbalaee, K. Hsu, and Y. Hong, "Probabilistic precipitation rate estimates with space-based infrared sensors," *Quart. J. Roy. Meteorolog. Soc.*, vol. 144, no. S1, pp. 191–205, Nov. 2018, doi: 10.1002/qj.3243.

[5] C. Guilloteau, E. Foufoula-Georgiou, P. Kirstetter, J. Tan, and G. J. Huffman, "How well do multi-satellite products capture the space–time dynamics of precipitation? Part I: Five products assessed via a wavenumber-frequency decomposition," *J. Hydrometeorol.*, vol. 22, no. 1, pp. 2805–2823, Aug. 2021.

[6] C. Guilloteau, E. Foufoula-Georgiou, P. Kirstetter, J. Tan, and G. J. Huffman, "How well do multisatellite products capture the space–time dynamics of precipitation? Part II: Building an error model through spectral system identification," *J. Hydrometeorol.*, vol. 23, no. 9, pp. 1383–1399, Aug. 2021, doi: 10.1175/JHM-D-21-0075.1.

[7] S. Ringerud, M. S. Kulie, D. L. Randel, G. M. Skofronick-Jackson, and C. D. Kummerow, "Effects of ice particle representation on passive microwave precipitation retrieval in a Bayesian scheme," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3619–3632, Jun. 2019, doi: 10.1109/TGRS.2018.2886063.

[8] C. Guilloteau and E. Foufoula-Georgiou, "Beyond the pixel: Using patterns and multiscale spatial information to improve the retrieval of precipitation from spaceborne passive microwave imagers," *J. Atmos. Ocean. Technol.*, vol. 37, no. 9, pp. 1571–1591, Sep. 2020, doi: 10.1175/JTECH-D-19-0067.1.

[9] D. B. Kirschbaum et al., "NASA's remotely sensed precipitation: A reservoir for applications users," *Bull. Amer. Meteorolog. Soc.*, vol. 98, no. 6, pp. 1169–1184, Jun. 2017, doi: 10.1175/BAMS-D-15-00296.1.

[10] Y. Tao, X. Gao, A. Ihler, K. Hsu, and S. Sorooshian, "Deep neural networks for precipitation estimation from remotely sensed information," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 1349–1355, doi: 10.1109/CEC.2016.7743945.

[11] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 2, pp. 254–259, Jan. 2018, doi: 10.1073/pnas.1715832114.

[12] D. W. Draper, D. A. Newell, F. J. Wentz, S. Krimchansky, and G. M. Skofronick-Jackson, "The global precipitation measurement (GPM) microwave imager (GMI): Instrument overview and early on-orbit performance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3452–3462, Jul. 2015, doi: 10.1109/JSTARS.2015.2403303.

[13] G. L. Stephens and C. D. Kummerow, "The remote sensing of clouds and precipitation from space: A review," *J. Atmos. Sci.*, vol. 64, no. 11, pp. 3742–3765, Nov. 2007, doi: 10.1175/2006JAS2375.1.

[14] M. Grecu et al., "The GPM combined algorithm," *J. Atmos. Ocean. Technol.*, vol. 33, no. 10, pp. 2225–2245, Oct. 2016, doi: 10.1175/JTECH-D-16-0019.1.

[15] C. Kidd, G. Huffman, V. Maggioni, P. Chambon, and R. Oki, "The global satellite precipitation constellation: Current status and future requirements," *Bull. Amer. Meteorolog. Soc.*, vol. 102, no. 10, pp. 1844–1861, 2021, doi: 10.1175/BAMS-D-20-0299.1.

[16] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling," *J. Hydrol.*, vol. 377, nos. 1–2, pp. 80–91, Oct. 2009, doi: 10.1016/j.jhydrol.2009.08.003.

[17] V. Venugopal and J. M. Wallace, "Climatology of contribution-weighted tropical rain rates based on TRMM 3B42," *Geophys. Res. Lett.*, vol. 43, no. 19, pp. 1–9, Oct. 2016, doi: 10.1002/2016GL069909.