

Proceeding Paper

# Uncertainty Quantification with Deep Ensemble Methods for Super-Resolution of Sentinel 2 Satellite Images <sup>†</sup>

David Iagaru <sup>1,2,\*</sup>  and Nina Maria Gottschling <sup>2</sup><sup>1</sup> CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France<sup>2</sup> MF-DAS OP-EO Data Science, Deutsches Zentrum für Luft- und Raumfahrt, 82234 Weßling, Germany

\* Correspondence: david.iagaru@student-cs.fr

<sup>†</sup> Presented at the 42nd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, 3–7 July 2023.

**Abstract:** The recently deployed Sentinel 2 satellite constellation produces images in 13 wavelength bands with a Ground Sampling Distance (GSD) of 10 m, 20 m, and 60 m. Super-resolution aims to generate all 13 bands with a spatial resolution of 10 m. This paper investigates the performance of DSen2, a proposed convolutional neural network (CNN)-based method, for tackling super-resolution in terms of accuracy and uncertainty. As the optimization problem for obtaining the weights of a CNN is highly non-convex, there are multiple different local minima for the loss function. This results in several possible CNN models with different weights and thus implies *epistemic uncertainty*. In this work, methods to quantify epistemic uncertainty, termed *weighted deep ensembles* (WDESen2 and its variants), are proposed. These allow the quantification of predictive uncertainty estimates and, moreover, the improvement of the accuracy of the prediction by selective prediction. They involve a consideration of deep ensembles, and each model's importance can be weighted depending on the model's validation loss. We show that weighted deep ensembles improve the accuracy of prediction compared to state-of-the-art methods and deep ensembles. Moreover, the uncertainties can be linked to the underlying inverse problem and physical patterns on the ground. This allows us to improve the trustworthiness of CNN predictions and the predictive accuracy with selective prediction.

**Keywords:** inverse problems; uncertainty quantification; deep ensemble methods; physics-informed ML; Bayesian inference



**Citation:** Iagaru, D.; Gottschling, N.M. Uncertainty Quantification with Deep Ensemble Methods for Super-Resolution of Sentinel 2 Satellite Images. *Phys. Sci. Forum* **2023**, *9*, 4. <https://doi.org/10.3390/psf2023009004>

Academic Editor: Udo von Toussaint and Roland Preuss

Published: 27 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Satellite images of multiple spectral bands acquired from various satellites, such as MODIS, VIIRS, ASTER, Worldview-3, and Sentinel 2, have different spatial resolutions. The resolution varies by a factor of 2–6 based on the instrumentation. This work focuses on the Sentinel 2 satellites, but the proposed methods for super-resolution, WDESen2 and its variants, are not limited to Sentinel 2 data. The constellation consists of two satellites that capture images of the Earth's surface, where each point on the surface is observed for a period of 5 days. Each image contains information from 13 spectral bands, some of which partially overlap, ranging from ultraviolet/visible to infrared. Band B10 is used for atmospheric corrections and was therefore omitted in the data used for this study. The bands are available in several different resolutions: bands B2, B3, B4, and B8 have a GSD of 10 and will be referred to as the *A bands*; B5, B6, B7, B8a, B10, and B11 have a GSD of 20 m and will be referred to as the *B bands*; while bands B1 and B9 have a GSD of 60 m and will be referred to as the *C bands*. The data can be used in various applications, such as hydrology and water resource management [1–3]. Even though there are reasons for the different spatial resolutions, such as storage and transmission bandwidth restrictions, often it is desirable to have all bands available at the highest spatial resolution. The problem of bringing all bands to a GSD of 10 m can be called

super-resolution. In the work of Lanaras et al. [4], the authors demonstrated that a CNN-based super-resolution method obtained higher accuracy for all super-resolved bands compared to state-of-the-art algorithms. State-of-the-art algorithms for the super-resolution of multi-spectral satellite data can be divided into three types: pan-sharpening per band, as reviewed in [5,6]; inverting an explicit imaging model, as demonstrated in [7,8], with a reduced-rank method applied for Sentinel 2 data in [9]; and machine learning approaches, such as [4]. The approaches in the second category are often based on the reformulation of super-resolution as a convex optimization problem by re-projecting each image in a subspace determined by the corresponding algorithm, such as in [9]. These methods rely on linear sensing model assumptions and are computationally expensive, as the super-resolution of each image requires a separate optimization. While being computationally more efficient, the CNN-based approach of [4] also better preserves spectral characteristics, achieves a favorable computational speed when run on GPUs, and is globally applicable for Sentinel 2 data without retraining. Thus, our work focuses on the DSen2 method from [4], which combines the information of the bands with lower and higher resolutions for super-resolution. The method relies on the assumption of the *self-similarity* of the upsampling process. This means that upsampling from 20 m to 10 m GSD is assumed to have the same mapping as upsampling from 40 m to 20 m GSD. Based on this assumption, in this work we address the features in B bands that cannot be super-resolved by exploiting the correlation between the bands with higher and lower resolutions. This is achieved in two ways: firstly, by studying extensions of the DSen2 method that predict uncertainty, such as WDESen2, and secondly, by studying the kernel of the inverse problem inherent to the proposed super-resolution method. Specifically, predictive uncertainty is evaluated with deep ensembles, as proposed in [10], and based on the success of weighted versions of deep ensembles for classification problems, as in [11,12], we propose a novel method of weighted deep ensembles for super-resolution. Uncertainty quantification (UQ) methods in DL usually distinguish between epistemic, or model, uncertainty and aleatoric, or data, uncertainty [13]. However, there is an ongoing debate as to whether these two types of uncertainty can be disentangled [14]. Adopting this perspective, we aim at explaining the obtained predictive uncertainties by the underlying physics, i.e., the inverse problem model. Here, we interpret the predictive uncertainties that arise by experimentally relating them to the kernel of the forward operator through projection onto the kernel of the inverse problem model. A different technique based on the generalized kernel has been applied in tomographic image reconstruction to detect hallucinations [15]. Moreover, we evaluate the uncertainties and kernel obtained by selective prediction, which has been applied in classification [16] and regression [17] problems for DL methods.

## 2. Methods

In this section, a brief introduction to Bayesian methods for uncertainty quantification in CNNs is given, followed by a short overview of deep ensembles as a Bayesian method and the extension thereof for super-resolution that is proposed in this work, WDESen2. Thereafter, the architecture and pipeline for the proposed method and its relation to the underlying inverse problem are described in detail.

### 2.1. Bayesian Methods for Uncertainty Quantification in Deep Neural Networks

#### 2.1.1. Bayesian Methods and Deep Ensembles

We denote as  $y \in Y$  the space of the predictions and  $x \in X$  the inputs. In Bayesian approaches for uncertainty quantification in CNNs, instead of obtaining a single set of weights  $\theta$  for the CNN by optimizing a given loss function, the aim is to obtain a posterior distribution of weights,  $p(\theta|D)$ , given the training data  $D = (X, Y)$  [18]. In the following, we give an outline of deep ensembles interpreted as a Bayesian method. Given the posterior distribution, a predictive distribution can be obtained by marginalization over the weights:

$$p(y|x, D) = \int p(y|x, \theta)p(\theta|D)d\theta.$$

Since the posterior distribution,  $p(\theta|D)$ , is intractable, it is common to either approximate it with a distribution or apply sampling methods [13]. Deep ensembles, as introduced in [10], can be interpreted as a sampling method [18]. The approach consists in training a number  $N \in \mathbb{N}$  of ensemble members that are CNNs,  $f_{\theta_i} : X \rightarrow Y$ . Due to the non-convexity of the optimization problem for obtaining the weights and the stochastic gradient descent methods, this results in  $N$  different sets of weights,  $(\theta_i(D))_{i=1}^N$ . Then, the posterior distribution [18] is assumed to be approximated by

$$p(\theta|D) \approx \frac{1}{N} \sum_{i=1}^N \delta(\theta = \theta_i). \tag{1}$$

The resulting prediction of the deep ensemble (DE) is the mean of the predictions,

$$\hat{y}_{DE}(x) = \frac{1}{N} \sum_{i=1}^N f_{\theta_i}(x) \tag{2}$$

and the epistemic uncertainty is given by the variance of the predictions,

$$\hat{\sigma}_{DE}(x)^2 = \frac{1}{N} \sum_{i=1}^N f_{\theta_i}(x)^2 - \hat{y}_{DE}(x)^2. \tag{3}$$

### 2.1.2. Utilizing the Loss-Weighted Deep Ensembles

In the following, we introduce an extension of deep ensembles, weighted deep ensembles, which assign weight to the prediction by the validation loss of each ensemble member. This approach implicitly approximates the posterior distribution in (1) with weights  $(w_i(\theta_i, V))_{i=1}^N$  attributed to each network in the ensemble, as illustrated in Figure 1. The weights  $w_i = w(\theta_i, V)$  depend on the validation loss and validation data  $V$  and the weights of each ensemble member:

$$p(\theta|D) \approx \sum_{i=1}^N w_i \delta(\theta = \theta_i). \tag{4}$$

Correspondingly, the prediction of the weighted deep ensemble (WDE) is given by the weighted mean,

$$\hat{y}_{WDE}(x) = \sum_{i=1}^N w_i f_{\theta_i}(x) \tag{5}$$

and the predictive uncertainty by the variance with respect to the distribution (4),

$$\hat{\sigma}_{WDE}(x)^2 = \sum_{i=1}^N w_i f_{\theta_i}(x)^2 - \hat{y}_{WDE}(x)^2. \tag{6}$$

### 2.1.3. Derivation of Ensemble Weights: Contribution Ratio Control

As shown in previous work, there are many possibilities for computing the weights  $(w_i)_{i=1}^N$ . We chose to compute the weights such that the contribution ratio can be controlled by a hyperparameter. The weights depending on the parameters  $\theta_i$  and the validation loss  $L_{\theta_i} = L_{\theta_i}(V)$  are defined as

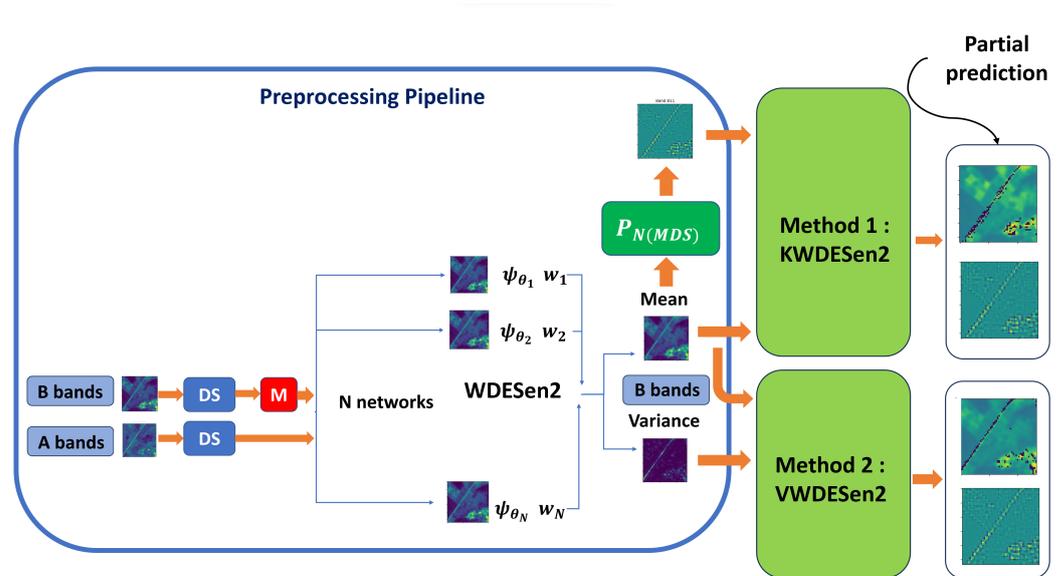
$$w_i(\theta_i, V) = \left( \frac{1}{L_{\theta_i} + \frac{rL_{\max} - L_{\min}}{1-r}} \right) / \left( \sum_{j=1}^N \frac{1}{L_{\theta_j} + \frac{rL_{\max} - L_{\min}}{1-r}} \right) \quad \forall i \in \{1, \dots, N\},$$

where  $L_{\max}(V)$  and  $L_{\min}(V)$  are the maximum and the minimum validation loss, respectively. The above definition ensures that the ensemble member with the lowest validation

loss is weighted as the most important, i.e., has the highest weight. Moreover, the weights satisfy the following conditions:

$$\frac{\min((w_i)_{i=1}^N)}{\max((w_i)_{i=1}^N)} = r \in (0, 1) \quad 0 < w_i(\theta_i, D) < 1 \quad \forall i \in \{1, \dots, N\}, \quad \sum_{i=1}^N w_i(\theta_i, D) = 1.$$

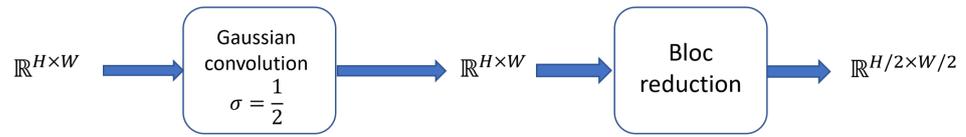
The last two conditions ensure that (4) is a probability distribution, and the first condition allows one to determine the ratio of contribution,  $r$ , a hyperparameter that we fixed at 0.5 in our experiments.



**Figure 1.** Architecture and pipeline of the KWDESen2 and VWDESen2 methods with  $N$  ensemble members.

2.2. Architecture of a Weighted Deep Ensemble for the Super-Resolution of Satellite Data: WDESen2

The weighted deep ensemble (WDESen2) architecture and its extensions, shown in Figure 1, use the architecture of the DSen2 network [4] as a backbone. By ensembling and weighting the predictions, WDESen2 improves the predictive accuracy compared to DSen2. The selective prediction applied to the predictive uncertainties from WDESen2, which are given as variances according to (6), is referred to as variance-based WDESen2 (VWDESen2). The selective prediction applied on the kernel projection is referred to as kernel-based WDESen2 (KWDESen2). The proposed methods are described below. The ensemble members of WDESen2 are the DSen2 networks that super-resolve the B and C bands to 10 m GSD. For this, two networks are used: a CNN,  $\mathbb{S}_{\times 6}$ , which performs the  $6\times$  super-resolution of the C bands from 60 m to 10 m GSD, and another CNN,  $\Psi_{\theta}$ , which performs the  $2\times$  super-resolution of the B bands from 20 m to 10 m GSD. As the  $6\times$  resolution network training procedure is structurally the same as that of the  $2\times$  resolution network, we restrict ourselves to describing the latter. For training a CNN to super-resolve the B bands, there do not exist ground-truth data with a 10 m GSD for the B bands. In order to generate training data, we followed the approach of [4]. The A bands and B bands are downsampled to 20 m and 40 m GSD, respectively. For each band and patch of size  $W/2 \times H/2$ , the linear downsampling operator,  $DS : \mathbb{R}^{W \times H} \rightarrow \mathbb{R}^{W/2 \times H/2}$ , is given below in Figure 2.



**Figure 2.** The  $DS : \mathbb{R}^{W \times H} \rightarrow \mathbb{R}^{W/2 \times H/2}$  operator. Here,  $W$  and  $H$  are the width and the height of a patch.

Then, the downsampled 40 m GSD and 20 m GSD B and A bands are used as inputs, and the 20 m GSD B bands can be used as targets in the training procedure. Based on the assumption of the self-similarity of super-resolution from 40 m GSD to 20 m GSD compared with that from 20 m GSD to 10 m GSD, the resulting network can then be used for  $2 \times$  super-resolution from 20 m to 10 m GSD. We let  $y_A \in \mathbb{R}^{4 \times W \times H}$ , where  $W \times H$  is the patch-size, and  $y_B \in \mathbb{R}^{6 \times W/2 \times H/2}$ , where  $W/2 \times H/2$  is the patch-size, denote the observed intensities for the A and B bands, respectively. The DSen2 method for  $2 \times$  super-resolution is given by

$$\Psi_\theta : \mathbb{R}^{4 \times W \times H} \times \mathbb{R}^{6 \times W/2 \times H/2} \rightarrow \mathbb{R}^{6 \times W \times H},$$

$$(y_A, y_b) \rightarrow x_B,$$

where the B bands, as depicted in Figure 1, are upsampled to the same resolution as the A bands with *bi-linear interpolation*. The associated linear operator,  $M : \mathbb{R}^{H/2 \times W/2} \rightarrow \mathbb{R}^{H \times W}$ , is applied to each B band.

By retraining  $N$  times on the same data splits with the  $\ell^1$  loss, due to the stochasticity of the training procedure, we obtain  $N$  different CNNs, as shown in Table 1. As visualized in Figure 1, the prediction of WDSen2 is composed of a weighted mean, according to (5), and the corresponding variance (6).

The predictions of WDESen2 are post-processed using two different methods, kernel- and variance-based partial prediction, which will be described in the following. For  $6 \times$  super-resolution to have the same dimension inputs in the CNN  $\Psi_\theta$ , the B and C bands are upsampled. However, this implies that there is no ground truth with which to compare the predictions.

### 2.2.1. Kernel Component Exploration

For training, the B band inputs,  $y_B \in \mathbb{R}^{6 \times W/2 \times H/2}$  of  $\Psi_\theta$ , are given by applying a composition of  $M$  and  $DS$ ,  $MDS : \mathbb{R}^{W \times H} \rightarrow \mathbb{R}^{W \times H}$ , as in Figure 1, band-wise to the training labels  $x_B \in \mathbb{R}^{6 \times W \times H}$ . Thus,  $\Psi_\theta$  aims to reconstruct  $x_B^i$  from  $y_B^i = MDSx_B^i$  and the other bands, where  $x_B^i$  is one of the B bands at 20 m GSD. As  $DS : \mathbb{R}^{W/2 \times H/2} \rightarrow \mathbb{R}^{W \times H}$  has a higher input dimension than output dimension, the composition  $MDS$ , depicted in Figure 3, has a non-trivial kernel. Recall that the kernel of a linear map  $A : E \rightarrow F$  is defined as  $\mathcal{N}(A) := \{x \in E | u(x) = 0\}$ . Thus, given an input  $y_i$ , different super-resolved solutions  $x_i$  may exist depending on the effect of correlation. In order to analyze the kernel component of the super-resolved bands, we use the projection operator  $P_{\mathcal{N}(MDS)}$  on the kernel  $\mathcal{N}(MDS)$ :

$$P_{\mathcal{N}(MDS)} = I - (MDS)^\dagger MDS : \mathbb{R}^{W \times H} \rightarrow \mathbb{R}^{W \times H},$$

where  $(MDS)^\dagger$  is the Moore–Penrose inverse of  $MDS$  [19], and this projection is approximated by a truncated singular value decomposition of  $MDS$ . The projection is computed individually per patch and band, as shown on the left-hand side of Figure 4.

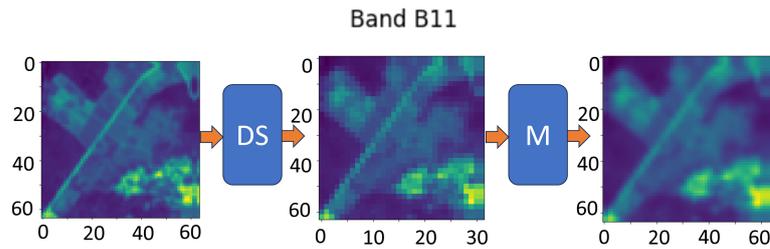


Figure 3. The composition of the DS and M operators,  $MDS : \mathbb{R}^{W \times H} \rightarrow \mathbb{R}^{W \times H}$ .

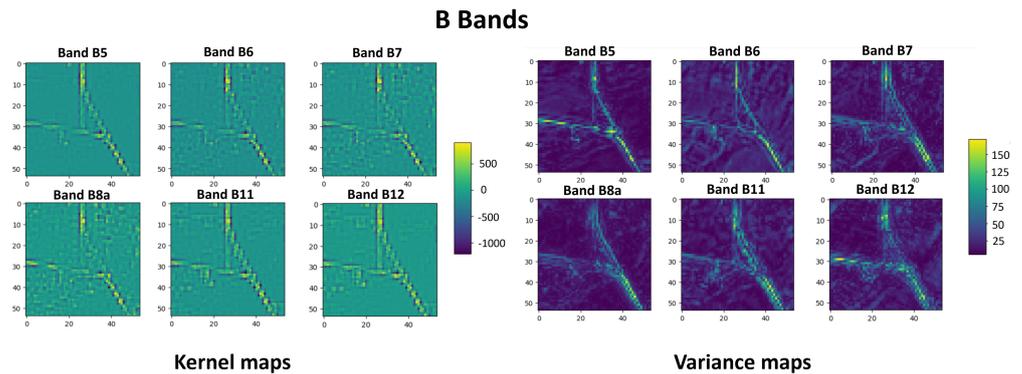


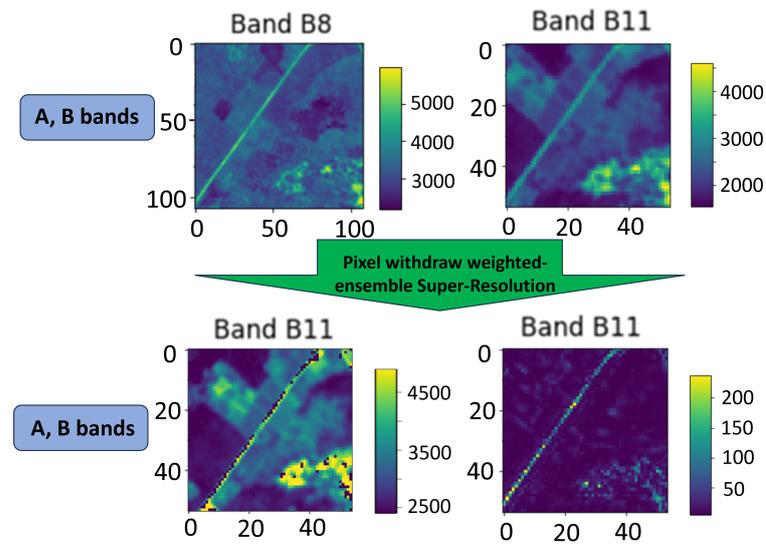
Figure 4. Kernel projection on the mean prediction (left) vs. variance (right) of WDES<sub>en2</sub>. The variance map visualizes the pixel-wise epistemic uncertainty of WDES<sub>en2</sub> with 30 ensemble members. The patch was from the tropical data.

2.3. Evaluation of the Predicted Variance and Kernel Projection of the Predictions of WDES<sub>en2</sub>

In order to utilize the predicted variance of WDES<sub>en2</sub> and the kernel projection, as described in Section 2.2.1, we perform selective prediction, as introduced for classification in [16] and applied to DL for regression problems in [17]. The general idea of selective prediction is to make the model abstain from giving a prediction if the uncertainty surpasses a specific threshold. We apply selective prediction to pixel-wise variances, i.e., uncertainties, from WDES<sub>en2</sub> and pixel-wise values of the kernel projection. The first method, variance-based WDES<sub>en2</sub> (VWDES<sub>en2</sub>), as visualized in Figure 5, abstains from predicting pixels that have the highest variance. In order to maintain comparability between different methods, we withdraw 5% of the pixels from the mean prediction of WDES<sub>en2</sub>. The second method is based on the underlying inverse problem, kernel-based WDES<sub>en2</sub> (KWDES<sub>en2</sub>), where the 5% of the pixels from the mean prediction that have the highest values in the kernel projection of the prediction are withdrawn from the prediction.

2.4. Dataset Splits

In Table 1, the dataset splits for the training, validation, and test datasets are listed. Inference at test time is performed on patches of two images of size  $20 \times 20 \text{ km}^2$ . The  $2 \times$  super-resolution methods are trained and evaluated individually on the “Tropical landscape” and “Mountains landscape” datasets, as these have different spectral characteristics.



**Figure 5.** A and B ground-truth bands (**top**), variance-based withdrawal (**bottom left**) and variance map (**bottom right**) for VWDESen2.

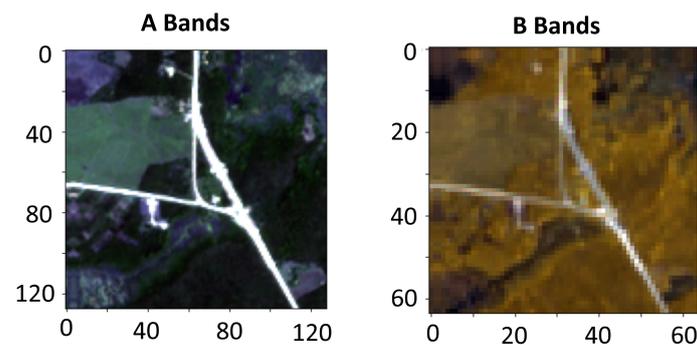
**Table 1.** Data splits for the training, validation, and test datasets. Regarding the images used for inference at test time, the first image will be referred to as “Tropical data” and the second image as “Mountain data”.

Images		Split	Patches
Tropical landscape/farmlands	1	Training	87%
		Validation	8.5%
		Test	4.5%
Mountains landscape	9	Training	89.5%
		Validation	10%
		Test	0.5%

### 3. Results

#### 3.1. Linking Uncertainties and Physical Patterns on the Ground

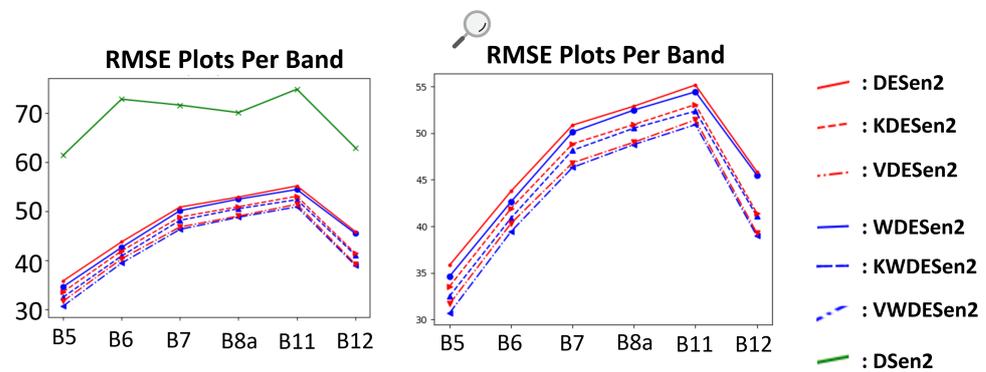
Figure 4 visualizes the kernel projections on the mean predictions and the uncertainty of WDESen2. Pixels of high variance correspond to pixels that have high values in the kernel projections. This indicates a possible relation between pixels where the predictions of the ensemble members diverge and the projection onto the kernel of MDS. Moreover, comparing Figure 4 to the ground truth shows that the pixels of high variance and the pixels with higher values in the kernel projections correspond to the physical patterns shown in Figure 6.



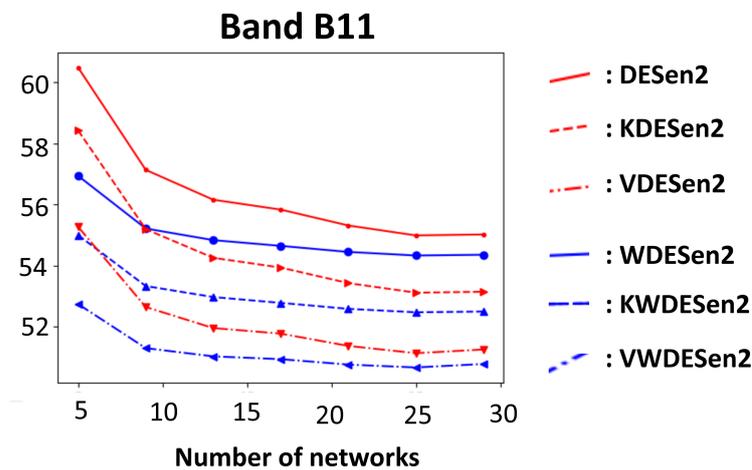
**Figure 6.** The A and B bands of the patch used in Figure 4.

### 3.2. Selective Prediction: Improving Accuracy with Uncertainty and Kernel Projections

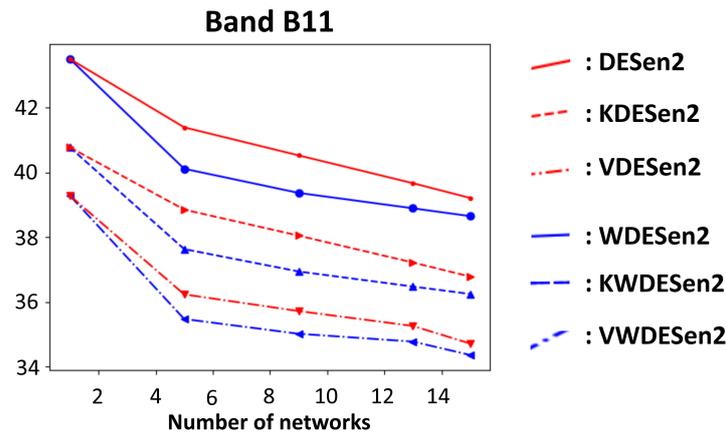
Figures 7–9 enable one to compare the predictions of WDESen2, KWDESen2, and VWDESen2 with those of deep ensembles, where the predictions are given by (2) and the variance by (3). Selective prediction, as in Section 2.3, was applied to deep ensembles DEsen2, resulting in a variance-based DEsen2 method (VDEsen2) and a kernel-based DEsen2 method (KDEsen2). Table 2 shows that removing the most uncertain 5% of the total number of pixels could significantly reduce the band-wise RMSE by up to 15% for VDEsen2, and removing the pixels with a high value in the kernel projection could reduce the band-wise RMSE by 7% for KDEsen2. Moreover, Figures 7–9 show that weighted deep ensembles with and without selective prediction were more accurate in terms of band-wise RMSE than deep ensembles and the baseline model from [4]. Moreover, Figures 8 and 9 indicate that, depending on the dataset, 10–15 ensemble members obtained better accuracy compared to smaller ensembles.



**Figure 7.** Band-wise RMSE of different deep ensemble methods (DEsen2, KDEsen2, VDEsen2, WDEsen2, KWDEsen2, and VWDEsen2). Each ensemble was composed of 30 members, and the baseline of DSen2 corresponds to 1 ensemble member. The RMSE was computed over 300 patches of tropical data.



**Figure 8.** Band B11’s RMSE for different numbers of ensemble members on tropical satellite data. For a given number of ensemble members  $n \in \{1, \dots, 30\}$ , a deep ensemble with  $n$  networks was randomly chosen from the 30 trained models, and this was averaged over 10 times in order to avoid the bias from the network sampling. The results were averaged over the patches.



**Figure 9.** Band B11’s RMSE for different numbers of ensemble members on mountain satellite data. For a given number of ensemble members  $n \in \{1, \dots, 14\}$ , a deep ensemble with  $n$  networks was randomly chosen from the 14 trained models, and this was averaged over 10 times in order to avoid the bias from the network sampling. The results were averaged over the patches.

**Table 2.** RMSE of the bicubic method, DSen2 baseline, deep ensemble of DSen2 (DESen2), weighted deep ensemble of DSen2 (WDESen2), weighted deep ensemble of DSen2 with kernel-based pixel withdrawal (KWDESen2), and weighted deep ensemble of DSen2 with variance-based pixel withdrawal (VWDESen2). All deep ensembles had 30 networks. The RMSE was computed over 300 patches of the tropical image.

	B5	B6	B7	B8a	B11	B12	Average
	<b>RMSE</b>						
Bicubic	93.2	105.1	124.6	133.6	105.4	94.8	109.4
DSen2	44.5	54.3	58.9	60.2	62.5	53.0	55.5
DESen2	35.7	43.6	50.8	52.8	55.1	45.8	47.3
WDESen2	34.6	42.6	50.1	52.4	54.4	45.5	46.6
KWDESen2	30.9	39.4	46.2	48.3	49.7	38.2	42.1
VWDESen2	28.7	37.6	44.0	46.0	48.0	35.8	40.0

#### 4. Conclusions

This work showed that by utilizing uncertainties in predictions, accuracy and interpretability can be improved compared to when other methods for the super-resolution of satellite data, specifically Sentinel 2 data, are applied. The first proposed method, weighted deep ensembles (WDESen2), improved the accuracy compared to state-of-the-art methods and at the same time provided predictive uncertainties. In the second proposed method, VWDESen2, selective prediction was applied to the uncertainties and yielded a lower RMSE compared to state-of-the-art methods. Moreover, the variance could be visually related to the projection onto the kernel of the underlying inverse problem and physical patterns on the ground. The corresponding pixels could be interpreted as being hard to super-resolve, as the withdrawal of such pixels led to a higher accuracy, resulting in the third proposed method, KWDESen2.

**Author Contributions:** Conceptualization and methodology, N.M.G. and D.I.; modeling and validation, D.I.; formal analysis, D.I. and N.M.G.; writing—original draft preparation, D.I. and N.M.G.; writing—review and editing, D.I. and N.M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The Sentinel-2 images used can be downloaded from: <https://dataspace.copernicus.eu/browser/>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mura, M.; Bottalico, F.; Giannetti, F.; Bertani, R.; Giannini, R.; Mancini, M.; Orlandini, S.; Travaglini, D.; Chirici, G. Exploiting the capabilities of the Sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *66*, 126–134. [[CrossRef](#)]
2. Castillo, J.A.A.; Apan, A.A.; Maraseni, T.N.; Salmo, S.G. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 70–85. [[CrossRef](#)]
3. Toming, K.; Kutser, T.; Laas, A.; Sepp, M.; Paavel, B.; Noges, T. First Experiences in Mapping LakeWater Quality Parameters with Sentinel-2 MSI Imagery. *Remote Sens.* **2016**, *8*, 640. [[CrossRef](#)]
4. Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltasvias, E. Super-Resolution of Sentinel-2 Images: Learning a Globally Applicable Deep Neural Network. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 305–319. [[CrossRef](#)]
5. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 2565–2586. [[CrossRef](#)]
6. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [[CrossRef](#)]
7. Brodu, N. Super-resolving multiresolution images with band-independent geometry of multispectral pixels. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4610–4617. [[CrossRef](#)]
8. Dian, R.; Li, S.; Fang, L.; Bioucas-Dias, J. Hyperspectral image super-resolution via local low-rank and sparse representations. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4003–4006.
9. Ulfarsson, M.O.; Palsson, F.; Dalla Mura, M.; Sveinsson, J.R. Sentinel-2 sharpening using a reduced-rank method. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6408–6420. [[CrossRef](#)]
10. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6405–6416.
11. Omisore, O.M.; Akinyemi, T.O.; Du, W.; Duan, W.; Orji, R.; Do, T.N.; Wang, L. Weighting-Based Deep Ensemble Learning for Recognition of Interventionalists’ Hand Motions During Robot-Assisted Intravascular Catheterization. *IEEE Trans. Hum. Mach. Syst.* **2022**, *53*, 215–227 [[CrossRef](#)]
12. Gupta, R.; Bhatnagar, A.S.; Singh, G. A Weighted Deep Ensemble for Indian Sign Language Recognition. *IETE J. Res.* **2023**, . [[CrossRef](#)]
13. Gawlikowski, J.; Tassi, C.R.N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Kruspe, A.M.; Triebel, R.; Jung, P.; Roscher, R.; et al. A Survey of Uncertainty in Deep Neural Networks. *Artif. Intell. Rev.* **2021**, *56*, 1513–1589. [[CrossRef](#)]
14. Gruber, C.; Schenk, P.O.; Schierholz, M.; Kreuter, F.; Kauermann, G. Sources of Uncertainty in Machine Learning—A Statisticians’ View. *arXiv* **2023**, arXiv:2305.16703.
15. Bhadra, S.; Kelkar, V.A.; Brooks, F.J.; Anastasio, M.A. On hallucinations in tomographic image reconstruction. *IEEE Trans. Med. Imaging* **2021**, *40*, 3249–3260. [[CrossRef](#)] [[PubMed](#)]
16. El-Yaniv, R. On the Foundations of Noise-free Selective Classification. *J. Mach. Learn. Res.* **2010**, *11*, 1605–1641.
17. Jiang, W.; Zhao, Y.; Wang, Z. Risk-controlled selective prediction for regression deep neural network models. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
18. Wilson, A.G.; Izmailov, P. Bayesian Deep Learning and a Probabilistic Perspective of Generalization. In Proceedings of the 34th International Conference on Neural Information Processing Systems, Online, 2–6 December 2020; Curran Associates Inc.: Red Hook, NY, USA, 2020.
19. Courriou, P. Fast Computation of Moore-Penrose Inverse Matrices. *Neural Inf. Process. Lett. Rev.* **2008**, *8*, 25–29.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.