

# Supplementary Materials

## Joint Supervised Classification and Reconstruction of Irregularly Sampled Satellite Image Times Series

Alexandre Constantin, *Student Member, IEEE*, Mathieu Fauvel, *Senior Member, IEEE*, and Stéphane Girard

### CONTENTS

<b>I</b>	<b>Estimation of the parameters</b>	0
<b>II</b>	<b>Numerical implementation and Convergence</b>	1
<b>III</b>	<b>Mean and Covariance functions</b>	1
<b>IV</b>	<b>Consequences of the independence assumption</b>	1
<b>V</b>	<b>Supervised Classification</b>	3
<b>VI</b>	<b>Time-series reconstruction</b>	6
	<b>References</b>	6

### I. ESTIMATION OF THE PARAMETERS

The identifiability of the model depends on the ability to compute  $\alpha_{b,c}$ . In (1) below, we remind that the optimal parameter is computed using the inverse of the matrix  $\mathbf{G}_c$ , defined as the sum of terms involving the design and precision matrices:

$$\alpha_{b,c} = \mathbf{G}_c^{-1} \left[ \sum_{i|Z_i=c} \mathbf{B}^{i\top} \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} \mathbf{y}_{i,b} \right], \quad \text{where } \mathbf{G}_c = \sum_{i|Z_i=c} \mathbf{B}^{i\top} \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} \mathbf{B}^i. \quad (1)$$

The goal of this Section is to establish a necessary and sufficient condition for  $\mathbf{G}_c$  to be non-singular. Let  $T^c$  be the total number of unique temporal acquisitions for a given class  $c$ . Let us also introduce the  $T_i \times T^c$  matrix  $\mathbf{N}_i$  (with  $T_i \leq T^c$  for all  $i$  such as  $z_i = c$ ) and the global design  $T^c \times J$  matrix  $\mathbf{B}_c$  such that  $\mathbf{B}^i = \mathbf{N}_i \mathbf{B}_c$ . The matrix  $\mathbf{N}_i$  is composed of ones and zeros to select the observed samples in signal  $i$  from the global design matrix. Then, the matrix  $\mathbf{G}_c$  can be rewritten as

$$\mathbf{G}_c = \sum_{i|Z_i=c} \mathbf{B}_c^\top \mathbf{N}_i^\top \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} \mathbf{N}_i \mathbf{B}_c = \mathbf{B}_c^\top \left[ \sum_{i|Z_i=c} \mathbf{N}_i^\top \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} \mathbf{N}_i \right] \mathbf{B}_c = \mathbf{B}_c^\top \mathbf{M}_c \mathbf{B}_c, \quad (2)$$

where we set  $\mathbf{M}_c = \sum_{i|Z_i=c} \mathbf{N}_i^\top \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} \mathbf{N}_i$ . In view of (2),  $\mathbf{G}_c$  is non-singular if and only if  $\mathbf{M}_c$  is non-singular and  $\mathbf{B}_c$  is injective.

- 1) Let us first prove that  $\mathbf{M}_c$  is non-singular. To this end, consider  $\mathbf{v}^* \in \mathbb{R}^{T^c}$  such that  $\mathbf{M}_c \mathbf{v}^* = \mathbf{0}$ . This implies that  $\mathbf{v}^{*\top} \mathbf{M}_c \mathbf{v}^* = 0$ . From the definition of  $\mathbf{M}_c$ , and since  $\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}$  is a positive definite matrix,  $\mathbf{v}^{*\top} \mathbf{M}_c \mathbf{v}^*$  is a sum of non-negative terms. Consequently, this entails that  $(\mathbf{N}_i \mathbf{v}^*)^{\top} \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} (\mathbf{N}_i \mathbf{v}^*) = 0$  for all  $i$  such that  $z_i = c$ . Using again the positive definiteness of  $\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}$  yields  $\mathbf{N}_i \mathbf{v}^* = \mathbf{0}$  for all  $i$  such that  $z_i = c$  or equivalently  $\mathbf{N} \mathbf{v}^* = \mathbf{0}$  where  $\mathbf{N} := [\mathbf{N}_1^\top, \dots, \mathbf{N}_{n_c}^\top]^\top$  and  $n_c$  is the number of signals in class  $c$ . Up to a lines permutation,  $\mathbf{N}$  can be rewritten as  $\mathbf{N} = [\mathbf{I}_{T^c}, \tilde{\mathbf{N}}]^\top$  where  $\mathbf{I}_{T^c}$  denotes the  $T^c \times T^c$  identity matrix. Then,  $\mathbf{N} \mathbf{v}^* = \mathbf{0}$  implies  $\mathbf{v}^* = \mathbf{0}$ . Hence, the result.
- 2) Second,  $\mathbf{B}_c$  is injective if and only if  $\mathbf{B}_c$  has full rank, *i.e.*  $\text{rank}(\mathbf{B}_c) \geq J$ .

As a conclusion, a necessary and sufficient condition for the existence of the inverse in (1) is  $\text{rank}(\mathbf{B}_c) \geq J$ . This condition involves constraints both on the orthogonal basis and on the temporal acquisition points. It can be checked numerically on each experiment. Finally, let us note that, in all situations,  $T^c \geq J$  is a necessary condition.

This work is supported by the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02) and by the Centre National d'Etudes Spatiales (CNES).

A. Constantin and S. Girard are with Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France (e-mail: alexandre.constantin@inria.fr; stephane.girard@inria.fr).

M. Fauvel is with CESBIO, Université de Toulouse, CNES/CNRS/INRAe/IRD/UPS, 31000 Toulouse, France (e-mail: mathieu.fauvel@inrae.fr).

## II. NUMERICAL IMPLEMENTATION AND CONVERGENCE

The model has been implemented with Python and inherits from Scikit Learn the Gaussian Process and Kernel classes [1]. The optimization problem is solved using L-BGFS-B subroutine [2] that allows additional constraints or bounds (*i.e.* positivity or minimal/maximal values) on the covariance function parameters.

Minimization of (7) in the main document is known to converge to a local minimum, which depends on initial values of the hyperparameters. Figure 2 illustrates the evolution of the parameter  $h$  of the RBF kernel as well as the marginal log-likelihood with respect to the number of iterations. Here, a Fourier basis of dimension 19 was used to estimate the mean function.

## III. MEAN AND COVARIANCE FUNCTIONS

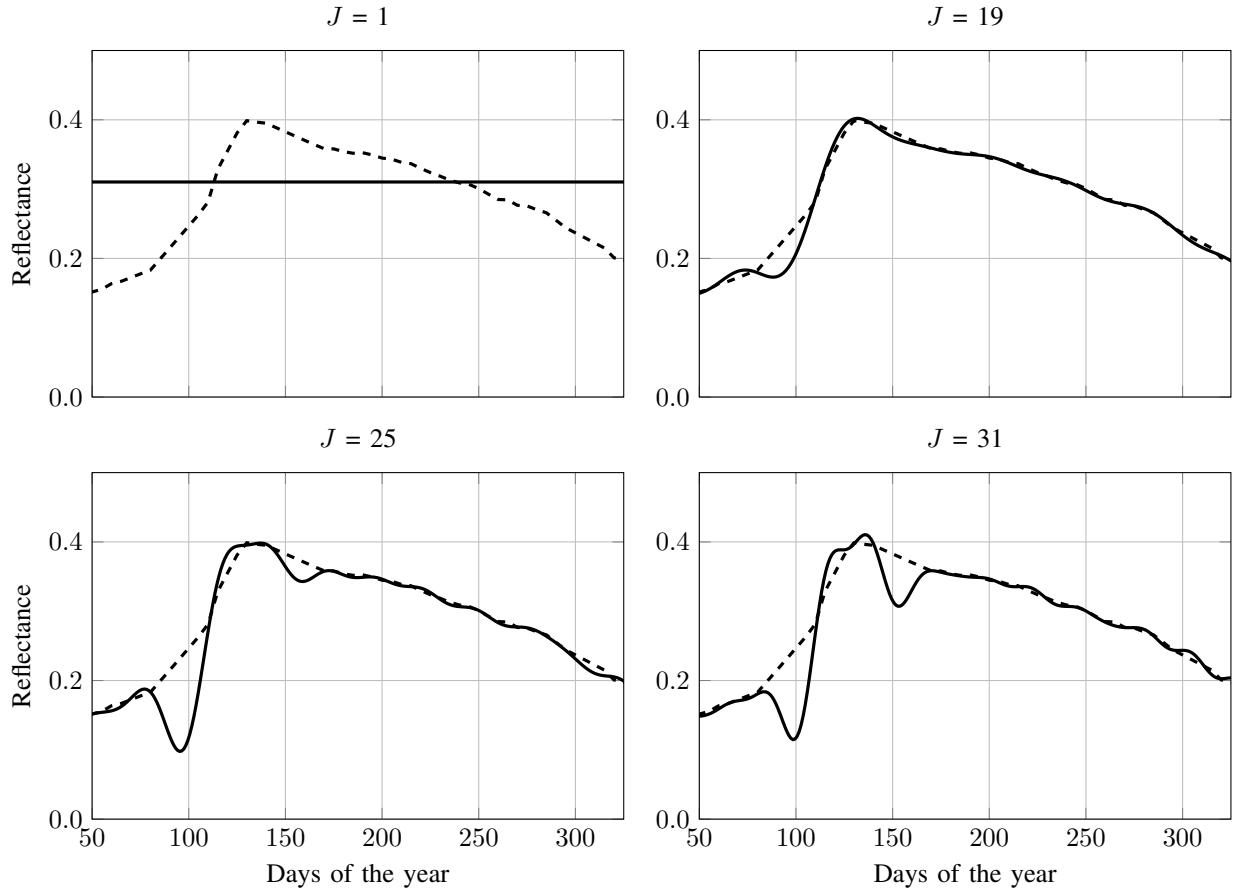


Fig. 1: Reconstructed mean near-infrared (IR) time-series from broad-leaved forest class with parameters  $\alpha$  (continuous line) estimated on the raw data, for different dimensions of basis. The dashed line corresponds to the gap-filled data mean from QDA on the resampled grid.

Figure 1 represents the estimated mean function in the infra-red for class “broad-leaved forest” for different dimensions of the Fourier basis. For  $J > 25$ , some oscillations appear in the reconstruction, due to a possible overfit.  $J = 19$  seems to be the most appropriate dimension, even though, in terms of classification,  $J = 25$  yields a slightly better accuracy score. In practice, a compromise should be found between a good reconstruction and a better classification score (see Figure 8 in the main document), depending on the final objective.

Figure 3 presents the RBF kernel computed as  $\exp\left\{-\frac{(t-180)^2}{2h^2}\right\}$ ,  $t \in \{1, \dots, 365\}$  on the infrared wavelength from three different classes. It appears that the continuous urban fabric has a higher temporal correlation than Broad-Leaved forests and Water bodies. Indeed, natural elements reflectance such as vegetation evolve along the year (e.g. phenology) while man-made material reflectance does not evolve along the year and thus exhibits longer temporal correlation.

## IV. CONSEQUENCES OF THE INDEPENDENCE ASSUMPTION

To illustrate the impact of the independence assumption on the classification accuracy, two versions of Quadratic Discriminant Analysis (QDA) have been implemented and compared. The first version, referred to as u-QDA is based on the assumption that all the wavelengths are uncorrelated. More specifically, a QDA model is fitted independently to each wavelength. The

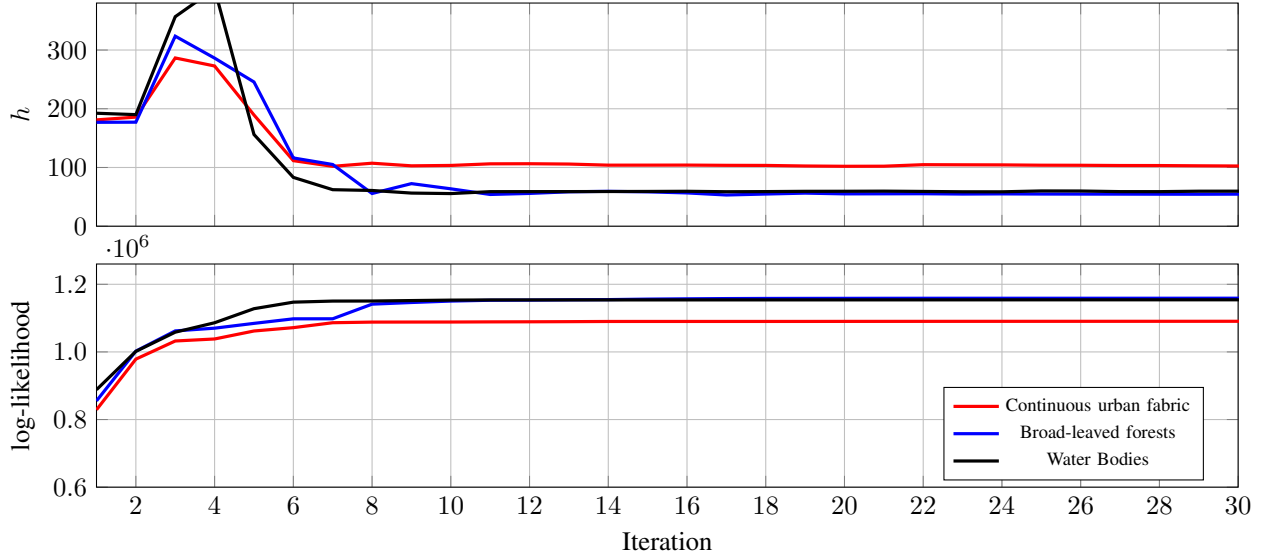


Fig. 2: Evolution of the length-scale  $h \in \theta$ , given in (4) in the main document, and associated log-likelihood for the IR wavelength and three different areas: artificial (continuous urban fabric), semi-natural (broad-leaved forests) and water (water bodies) areas.

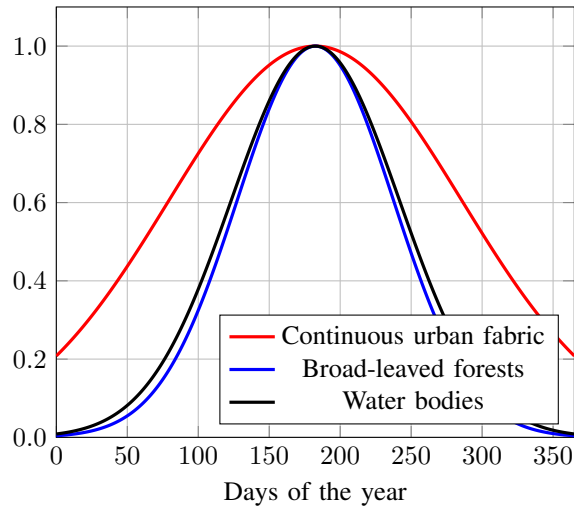


Fig. 3: Estimated RBF kernels for the IR wavelength and three different areas: artificial (continuous urban fabric), semi-natural (broad-leaved forests) and water (water bodies) areas. It shows the correlation between day of the year 180 and the other days, from day 1 to day 360.

likelihood is computed as the product of the likelihoods associated with each wavelength, and the posterior probability is obtained using Bayes' rule as usual. The second version is the conventional QDA method presented in the main document. The results are reported in Table I. It appears that the classification accuracy of u-QDA is significantly smaller than the one of the usual QDA method. Therefore, one can conclude that the independence assumption has a significant impact on the performance of the classifiers.

TABLE I: Averaged mean  $F_1$  score (mean(%)  $\pm$  standard deviation) associated with u-QDA and QDA computed on three tiles.

	u-QDA	QDA
T31TCJ	26.2 $\pm$ 3.1	36.2 $\pm$ 2.5
T31TDN	21.3 $\pm$ 3.6	30.5 $\pm$ 2.8
T31TGK	29.2 $\pm$ 2.1	38.9 $\pm$ 2.1

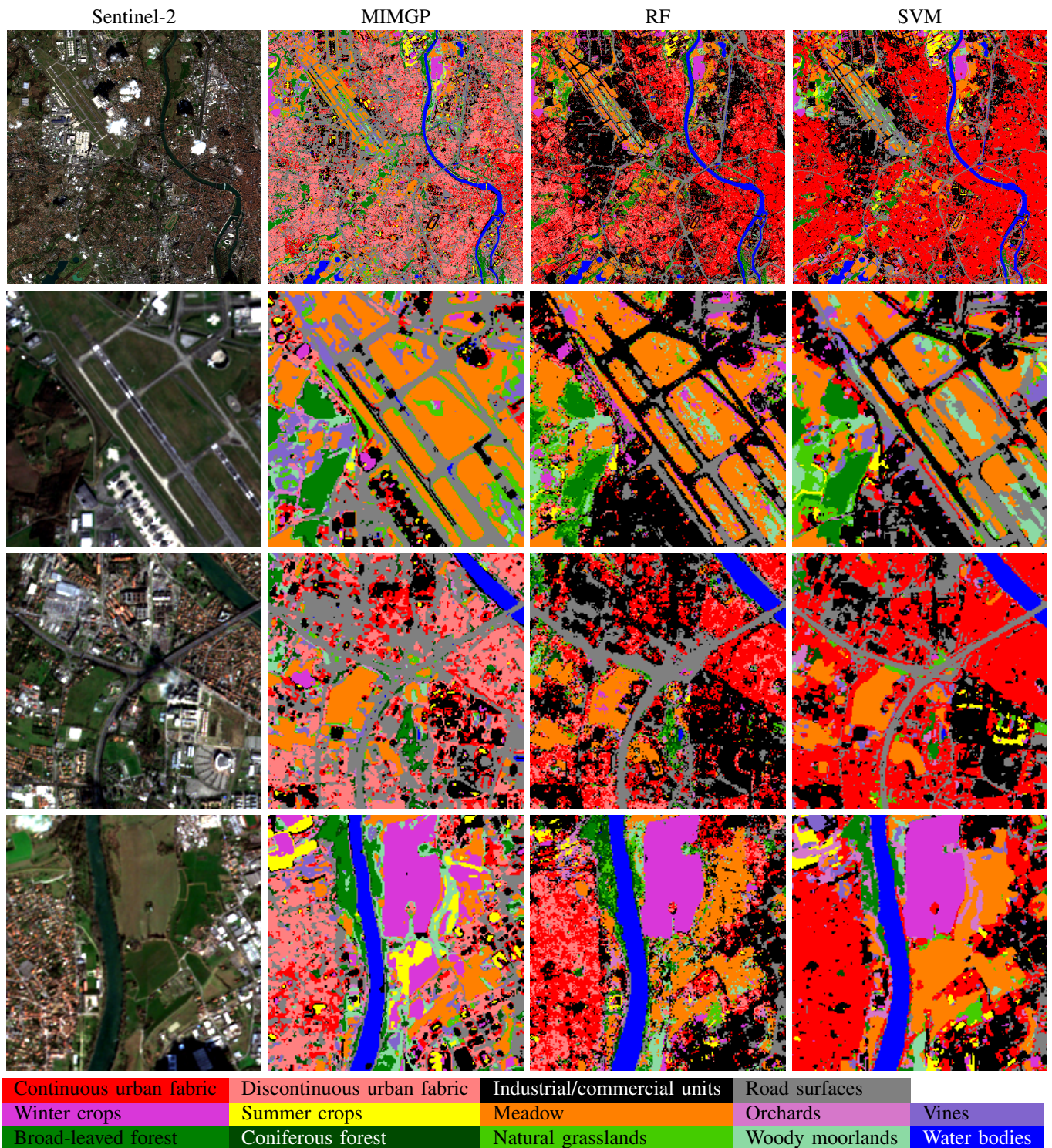


Fig. 4: Thematic maps for 3 sites from tile T31TCJ. The first line is the whole 10km long side square while the other lines are enlargements (2km length) of some extracts.

## V. SUPERVISED CLASSIFICATION

This section provides an extract of the thematic classification maps obtained with RF, MIMGP and SVM on tiles T31TCJ (Figure 4), T31DN (Figure 5) and T31TGK (Figure 6). The original tiles are too large to be correctly displayed, we thus choose to extract the center of each tile. All extracts have  $1000 \times 1000$  pixels. We also provide a zoom ( $200 \times 200$  pixels) on several extracts.

The  $F_1$  scores for each class and for each tiles, averaged over the 10 independent runs, are provided in tables II, III and IV.

TABLE II:  $F_1$  scores for tile T31TCJ.

Class	QDA	RF	SVM	MIMGP
Continuous urban fabric	23.0 ± 11.8	53.6 ± 01.1	<b>56.9 ± 01.2</b>	27.8 ± 03.6
Discontinuous urban fabric	24.2 ± 05.6	<b>53.7 ± 02.7</b>	38.9 ± 10.1	51.2 ± 01.1
Industrial or commercial units	13.1 ± 03.7	<b>55.5 ± 00.9</b>	47.8 ± 03.8	27.7 ± 01.6
Road surfaces	36.7 ± 06.1	<b>85.4 ± 01.7</b>	78.1 ± 04.0	65.7 ± 03.2
Winter crops	40.7 ± 07.9	93.2 ± 00.7	<b>93.4 ± 01.1</b>	86.4 ± 02.0
Summer crops	66.5 ± 08.4	<b>96.6 ± 00.3</b>	95.5 ± 00.9	92.1 ± 01.0
Meadow	25.4 ± 07.7	<b>63.4 ± 02.9</b>	63.2 ± 03.7	53.7 ± 06.2
Orchards	48.8 ± 05.9	<b>79.2 ± 02.7</b>	76.4 ± 03.3	53.1 ± 04.4
Vines	21.6 ± 13.9	74.4 ± 06.6	<b>78.7 ± 07.0</b>	64.6 ± 09.7
Broad-leaved forest	58.6 ± 11.0	<b>85.3 ± 02.4</b>	84.6 ± 02.9	71.7 ± 03.6
Coniferous forest	28.4 ± 16.5	86.1 ± 02.5	<b>86.4 ± 02.4</b>	73.4 ± 03.6
Natural grasslands	06.0 ± 07.1	<b>29.3 ± 15.3</b>	22.5 ± 12.2	21.8 ± 13.0
Woody moorlands	25.9 ± 03.6	54.6 ± 03.5	<b>55.4 ± 04.0</b>	20.6 ± 05.9
Water bodies	87.7 ± 07.7	<b>99.3 ± 00.0</b>	99.2 ± 00.1	95.1 ± 01.9

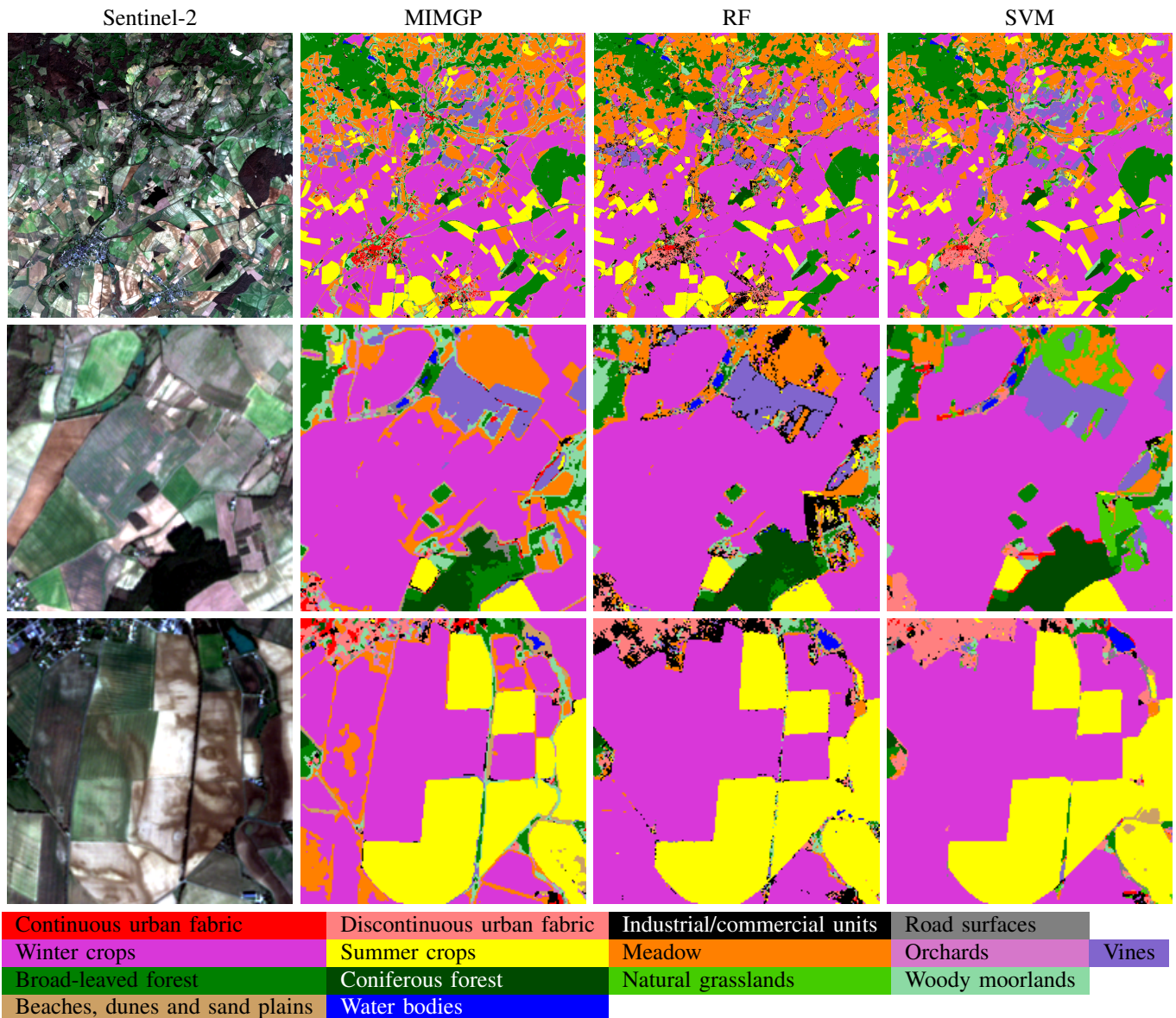


Fig. 5: Thematic maps for 3 sites from tile T31TDN. The first line is the whole 10km long side square while the other lines are enlargements (2km length) of some extracts.

TABLE III:  $F_1$  scores for tile T31TDN.

Class	QDA	RF	SVM	MIMGP
Continuous urban fabric	20.3 $\pm$ 7.7	<b>67.7 <math>\pm</math> 2.3</b>	66.6 $\pm$ 4.0	58.5 $\pm$ 2.8
Discontinuous urban fabric	14.0 $\pm$ 6.5	<b>64.7 <math>\pm</math> 3.6</b>	54.8 $\pm$ 11.0	51.3 $\pm$ 3.2
Industrial or commercial units	13.4 $\pm$ 2.2	<b>55.1 <math>\pm</math> 2.6</b>	48.3 $\pm$ 4.3	38.3 $\pm$ 1.6
Road surfaces	34.9 $\pm$ 6.4	<b>83.3 <math>\pm</math> 1.5</b>	75.6 $\pm$ 2.9	65.2 $\pm$ 2.6
Winter crops	23.0 $\pm$ 5.0	<b>93.4 <math>\pm</math> 0.7</b>	91.4 $\pm$ 1.2	85.3 $\pm$ 1.6
Summer crops	26.5 $\pm$ 9.9	<b>95.5 <math>\pm</math> 0.7</b>	94.2 $\pm$ 1.0	92.9 $\pm$ 0.8
Meadow	32.1 $\pm$ 3.0	85.0 $\pm$ 2.1	<b>85.4 <math>\pm</math> 2.2</b>	71.2 $\pm$ 1.8
Orchards	24.4 $\pm$ 12.9	64.1 $\pm$ 9.6	<b>66.0 <math>\pm</math> 20.1</b>	41.8 $\pm$ 6.8
Vines	42.8 $\pm$ 13.9	<b>89.6 <math>\pm</math> 2.1</b>	88.4 $\pm$ 3.0	85.8 $\pm$ 2.0
Broad-leaved forest	7.6 $\pm$ 10.1	<b>86.0 <math>\pm</math> 1.2</b>	83.5 $\pm$ 2.0	70.9 $\pm$ 2.8
Coniferous forest	61.6 $\pm$ 7.6	<b>91.9 <math>\pm</math> 0.3</b>	91.2 $\pm$ 0.5	80.3 $\pm$ 1.5
Natural grasslands	24.3 $\pm$ 11.6	<b>43.1 <math>\pm</math> 11.3</b>	36.7 $\pm$ 14.3	39.3 $\pm$ 10.0
Woody moorlands	4.9 $\pm$ 4.7	<b>69.2 <math>\pm</math> 2.2</b>	68.3 $\pm$ 3.2	49.8 $\pm$ 2.5
Beaches, dunes and sand plains	43.1 $\pm$ 16.5	<b>80.6 <math>\pm</math> 5.6</b>	78.7 $\pm$ 4.4	59.7 $\pm$ 6.8
Water bodies	84.1 $\pm$ 3.3	<b>94.0 <math>\pm</math> 1.6</b>	92.3 $\pm$ 2.6	86.4 $\pm$ 2.6

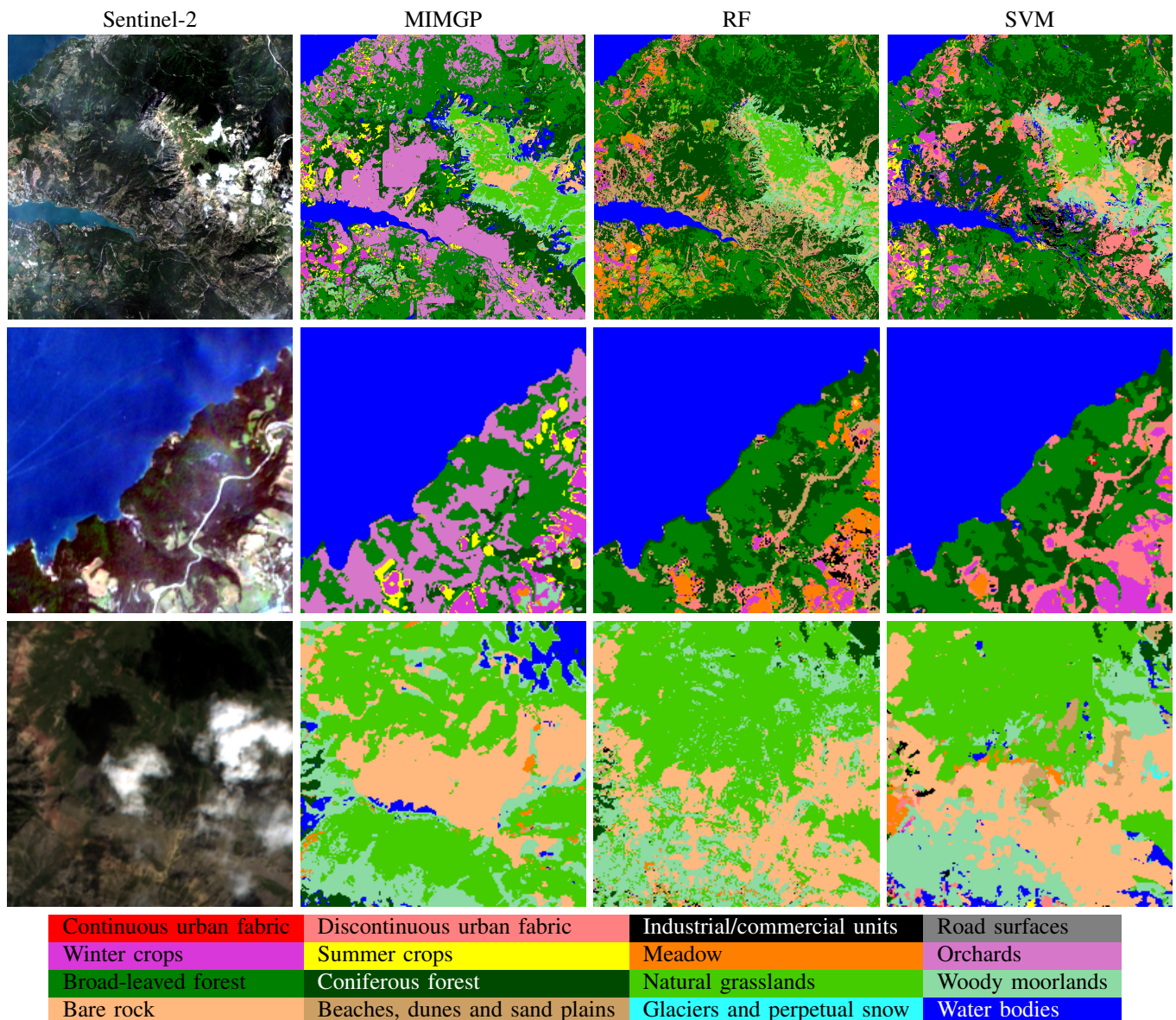


Fig. 6: Thematic maps for 3 sites from tile T31TGD. The first line is the whole 10km long side square while the other lines are enlargements (2km length) of some extracts.

TABLE IV:  $F_1$  scores for tile T31TGK.

Class	QDA	RF	SVM	MIMGP
Continuous urban fabric	0.3 ± 0.9	<b>9.1 ± 3.6</b>	1.7 ± 1.0	8.7 ± 4.2
Discontinuous urban fabric	23.2 ± 18.3	<b>54.6 ± 4.5</b>	45.6 ± 11.0	27.7 ± 15.5
Industrial or commercial units	27.0 ± 6.1	<b>55.5 ± 3.5</b>	50.3 ± 4.1	28.2 ± 4.7
Road surfaces4	33.3 ± 12.2	<b>63.2 ± 8.5</b>	61.3 ± 9.0	42.1 ± 5.8
Winter crops	61.6 ± 9.2	<b>91.3 ± 1.9</b>	88.0 ± 3.7	72.4 ± 3.2
Summer crops	80.8 ± 4.7	<b>94.3 ± 1.4</b>	92.8 ± 2.2	81.9 ± 3.2
Meadow	25.2 ± 4.7	48.1 ± 2.6	<b>53.1 ± 2.2</b>	19.2 ± 3.2
Orchards	84.6 ± 3.6	<b>91.5 ± 1.2</b>	88.9 ± 2.0	73.5 ± 2.3
Broad-leaved forest	20.9 ± 15.8	<b>79.9 ± 5.2</b>	78.5 ± 5.2	56.0 ± 17.0
Coniferous forest	43.8 ± 4.6	<b>79.4 ± 2.5</b>	79.2 ± 2.6	45.4 ± 4.6
Natural grasslands	35.2 ± 5.6	<b>49.3 ± 3.2</b>	45.5 ± 9.3	37.9 ± 4.4
Woody moorlands	46.7 ± 5.3	<b>50.8 ± 2.9</b>	48.4 ± 5.4	33.2 ± 2.3
Bare rock	40.7 ± 7.0	72.3 ± 4.9	<b>73.2 ± 4.4</b>	46.8 ± 4.2
Beaches, dunes and sand plains	47.3 ± 14.3	56.8 ± 9.3	<b>61.3 ± 10.7</b>	33.8 ± 7.5
Glaciers and perpetual snow	76.6 ± 10.2	89.3 ± 7.3	<b>90.9 ± 2.7</b>	78.7 ± 10.8
Water bodies	14.5 ± 11.1	<b>86.3 ± 2.4</b>	83.3 ± 4.1	43.2 ± 13.1

## VI. TIME-SERIES RECONSTRUCTION

This Section displays additional temporal reconstructions.

Figures 7 and 8 present the reconstructed time-series for all days in year 2018 for two pixels on four wavelengths of different classes, summer crops and broad-leaved forests.

Figure 9 compares reconstructions obtained with MIMGP model and Whittaker smoother on the 10 wavelengths. The more intense the colors are (yellow) on the line of equation  $y = x$ , the better the reconstruction is. Both methods have higher density on the  $y = x$  line, confirming that there is no bias in the reconstructions, but Whittaker exhibits higher dispersion. This is especially true for tiles T31TDN and T31TGK.

Finally, MIMGP and Whittaker image reconstructions are compared visually on the three tiles. The four wavelengths natively at a resolution of 10 meters per pixel (Blue, Green, Red and Infrared) are reported in Figures 12, 13 and 14 for, respectively, tile T31TCJ, T31TDN and T31TGK. For the visualization sake, each image of a given tile has been displayed using the same histogram stretching function and identical parameters. The mask is a binary image, with white color for valid pixels and black color for invalid pixels (clouds, saturation, ...).

## REFERENCES

- [1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [2] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software (TOMS)*, vol. 23, no. 4, pp. 550–560, 1997.

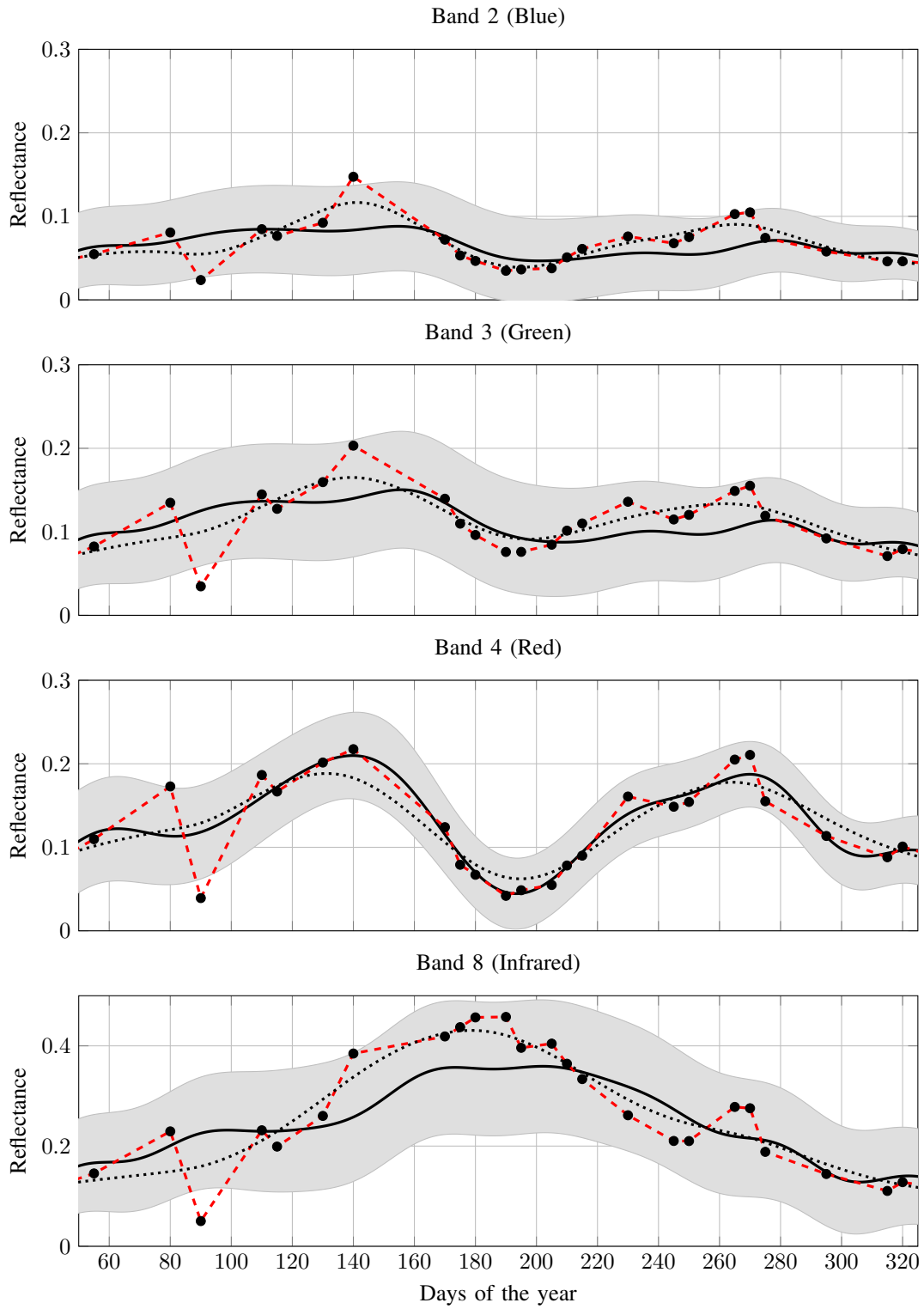


Fig. 7: Blue, Green, Red and IR wavelengths reconstruction for the Summer crops class. The red dashed line represents the linear interpolation, the black dotted line the Whittaker reconstruction and the black continuous line the MIMGP reconstruction.



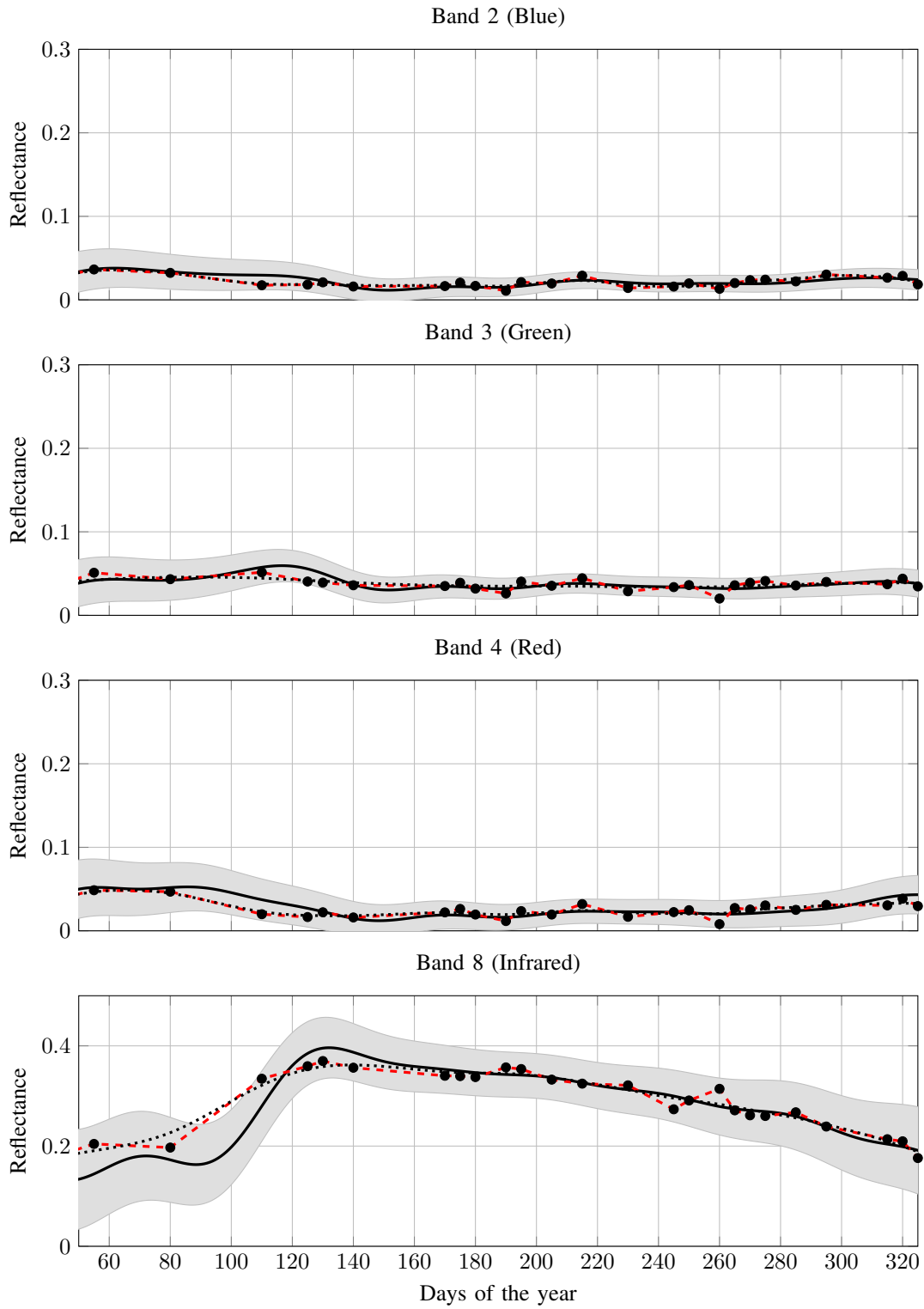


Fig. 8: Blue, Green, Red and IR wavelengths reconstruction for the broad-leaved forests class. The red dashed line represents the linear interpolation, the black dotted line the Whittaker reconstruction and the black continuous line the MIMGP reconstruction.

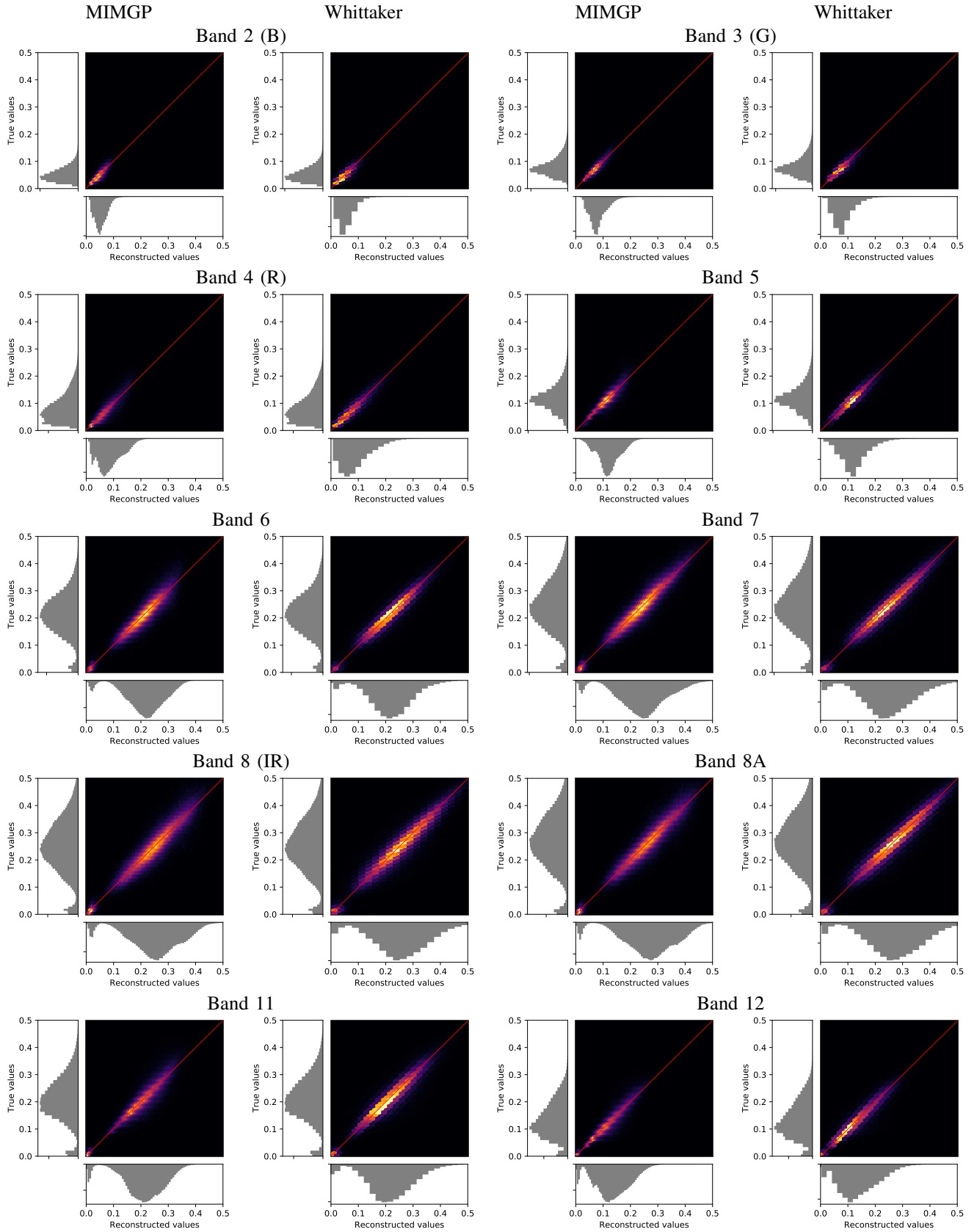


Fig. 9: Scatter plots from tile T31TCJ. They illustrate the link between the true values (x-axis) and the reconstructed ones (y-axis) with MIMGP (left) and with the Whittaker smoother (right).

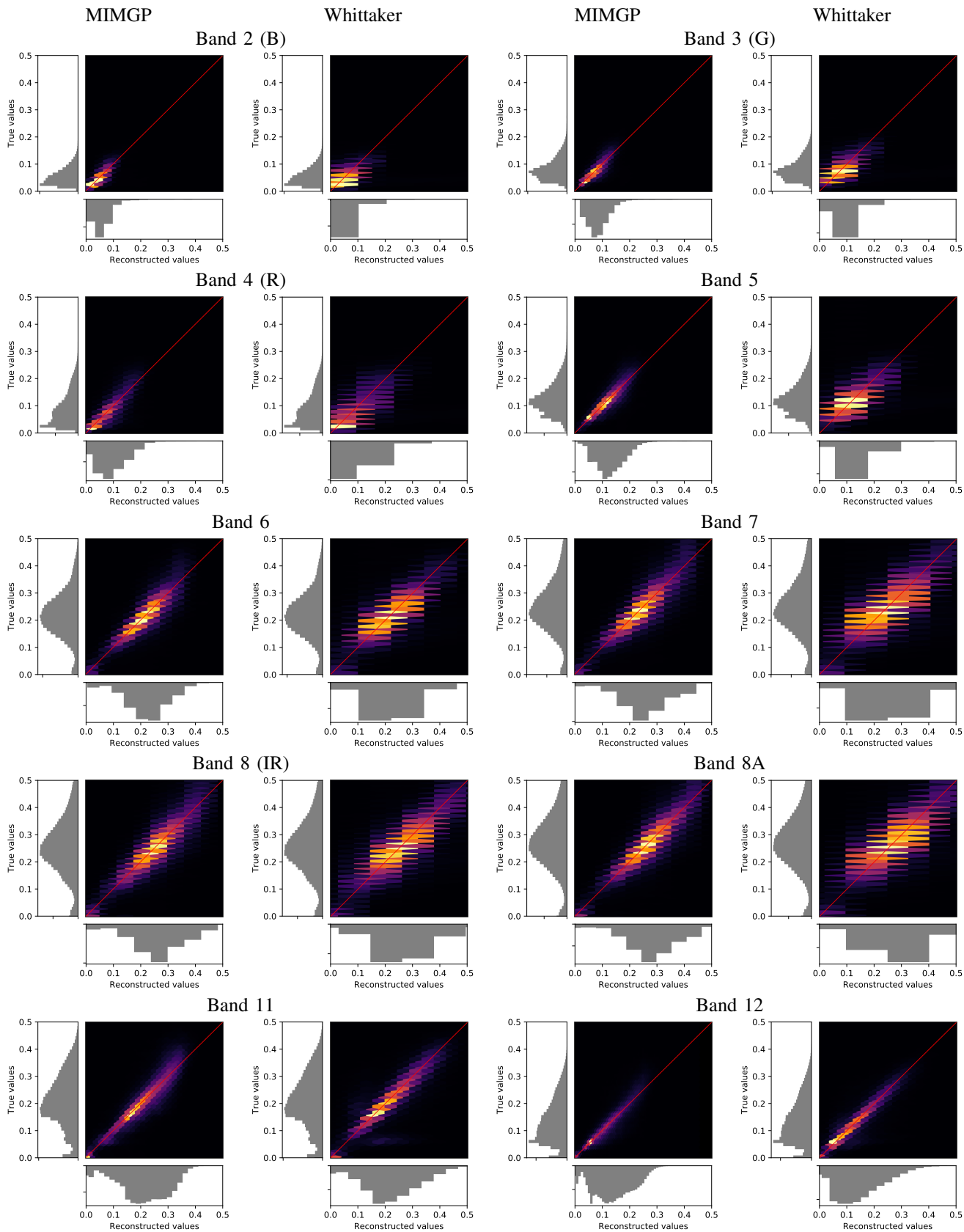


Fig. 10: Scatter plots from tile T31TDN. They illustrate the link between the true values (x-axis) and the reconstructed ones (y-axis) with MIMGP (left) and with the Whittaker smoother (right).

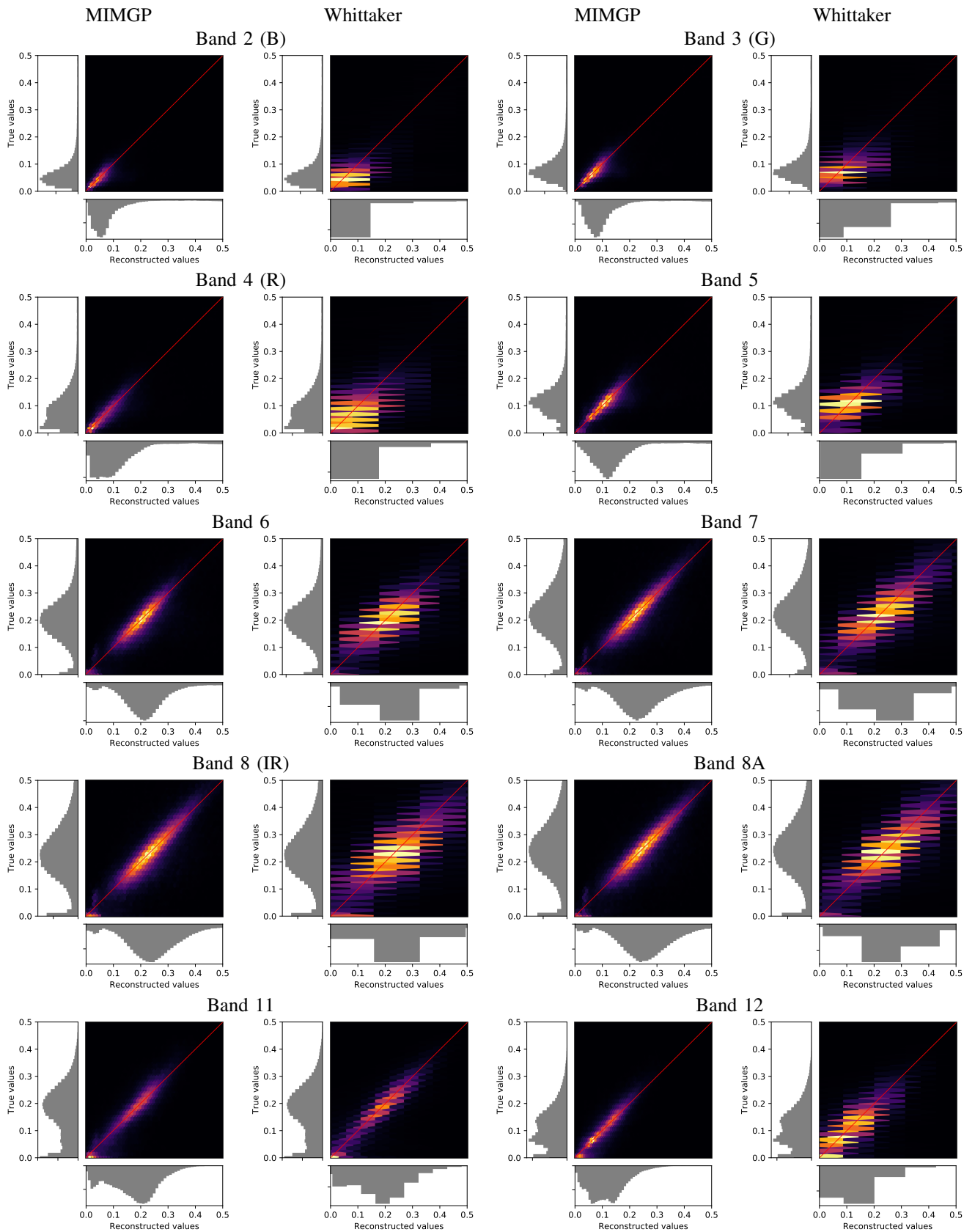


Fig. 11: Scatter plots from tile T31TGK. They illustrate the link between the true values (x-axis) and the reconstructed ones (y-axis) with MIMGP (left) and with the Whittaker smoother (right).

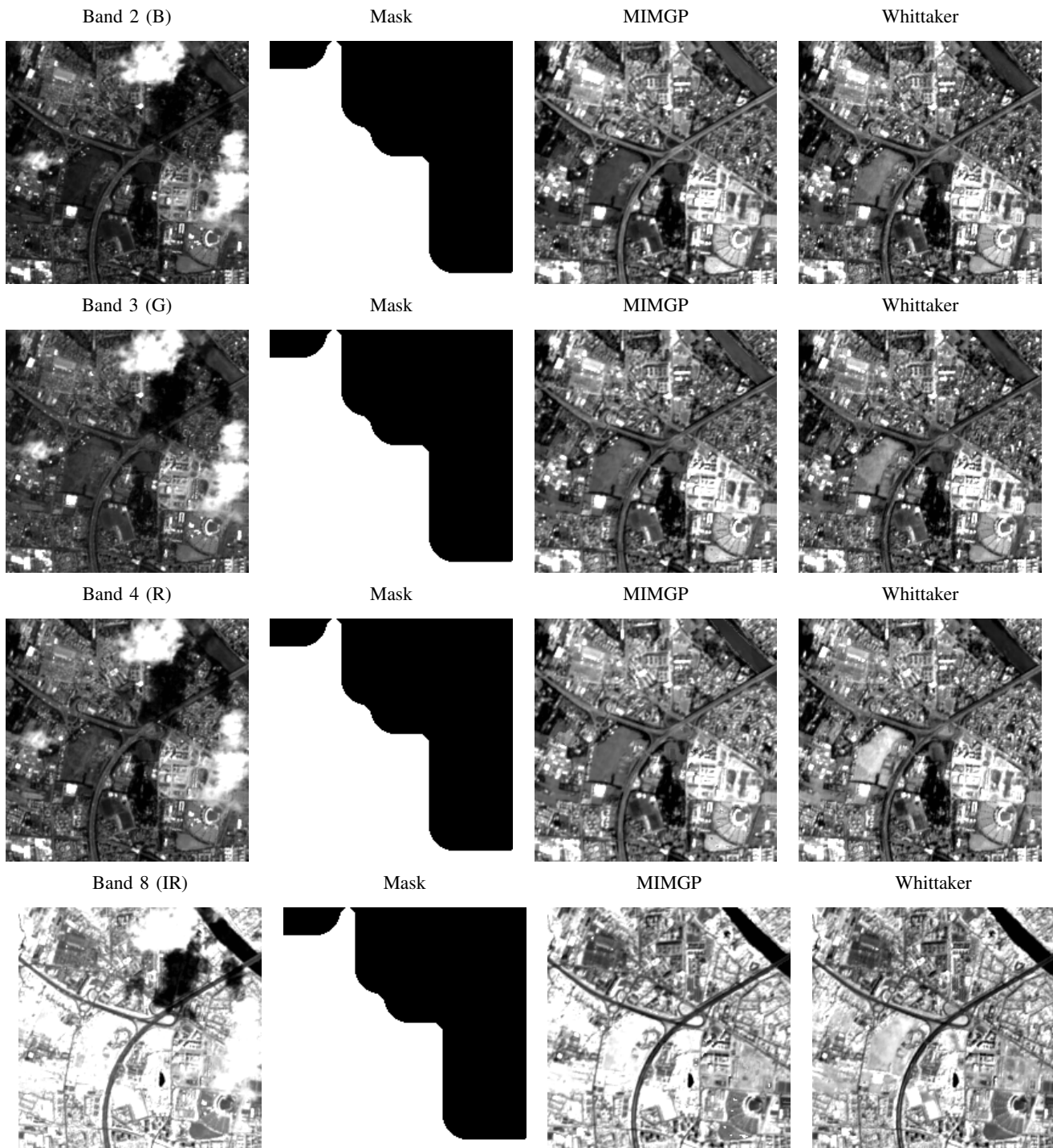


Fig. 12: Comparison of reconstructions on a 2km long side square from T31TCJ tile on July 12, 2018.

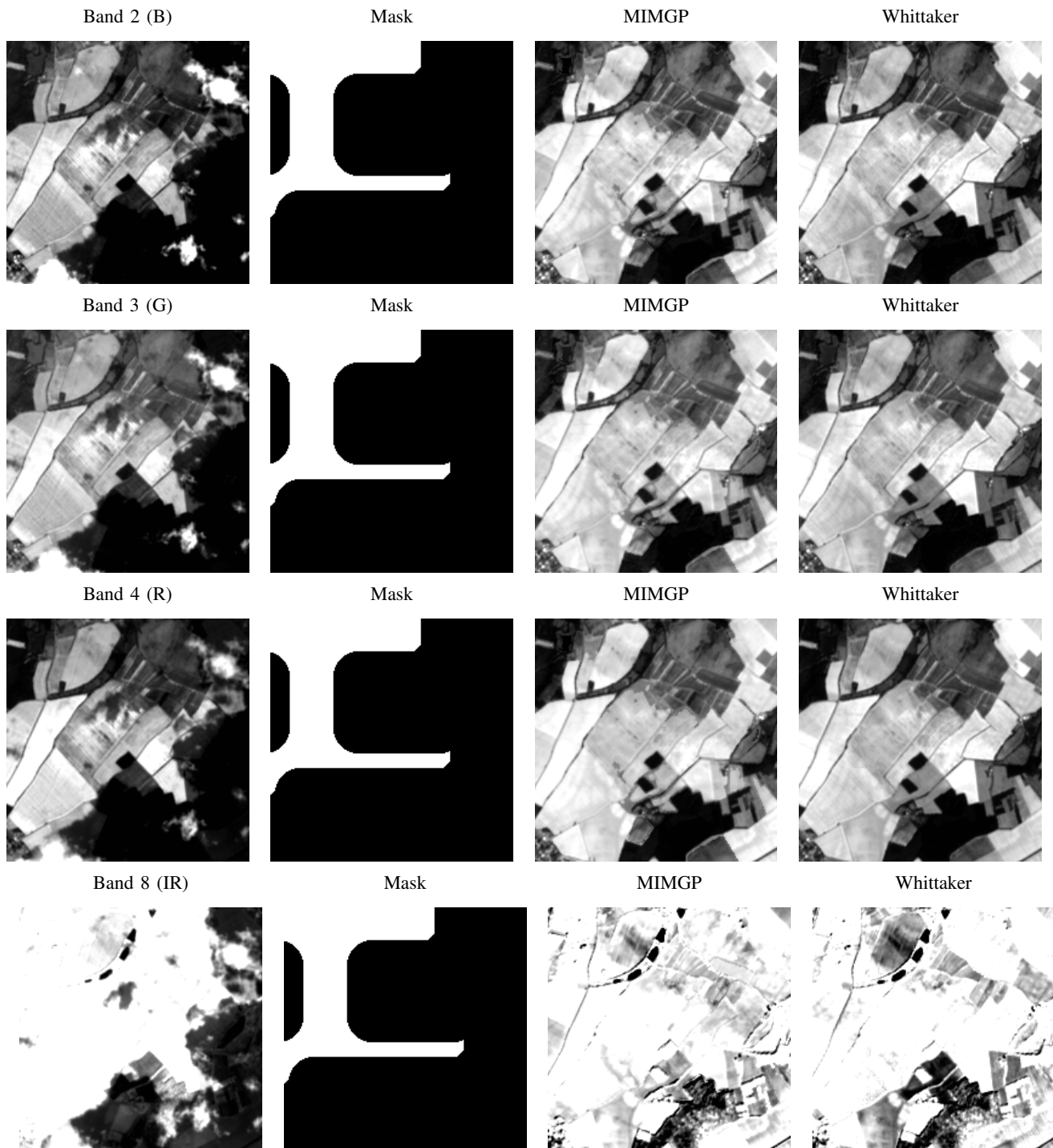


Fig. 13: Comparison of reconstructions on a 2km long side square from T31TDN tile on July 14, 2018.

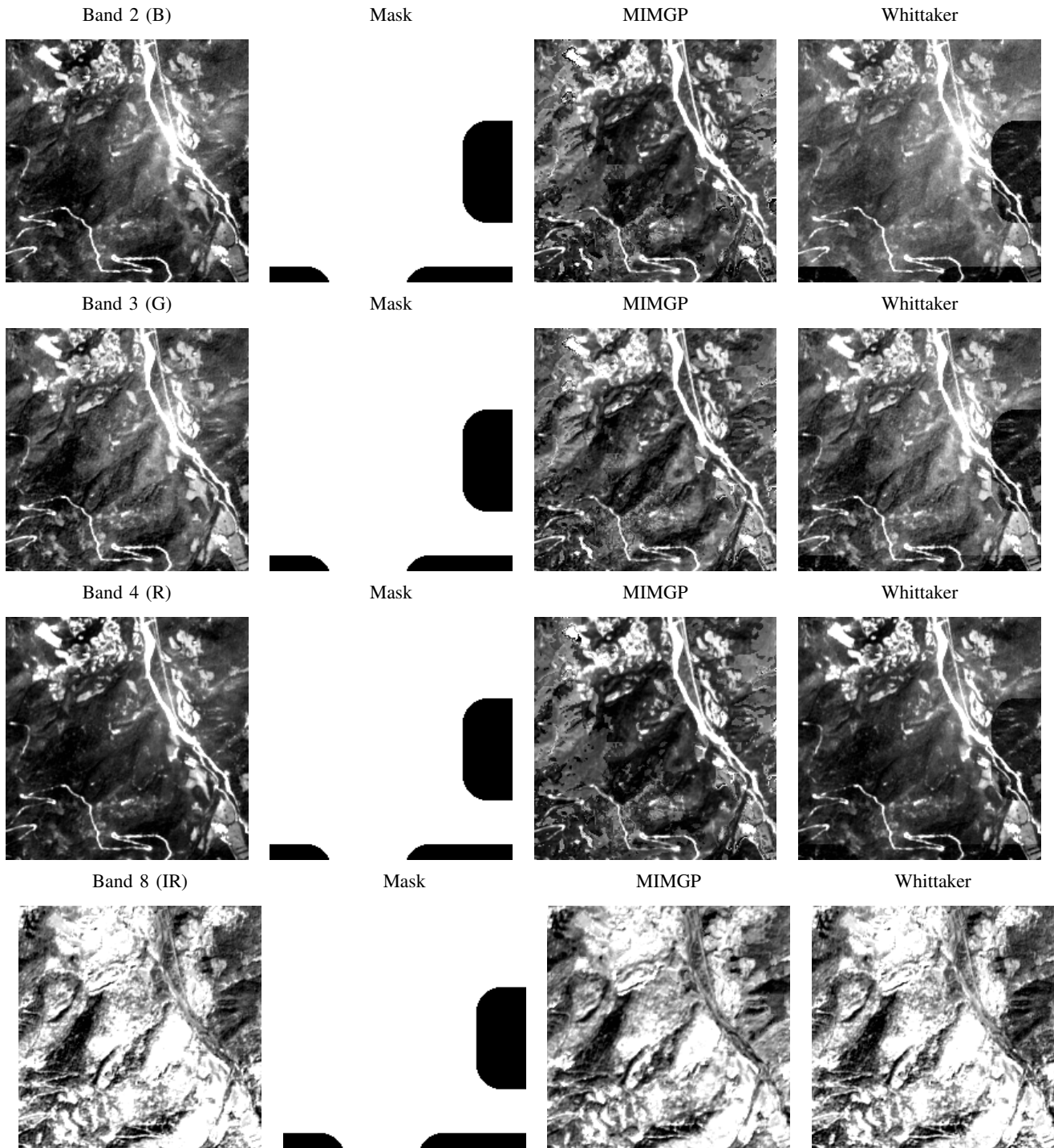


Fig. 14: Comparison of reconstructions on a 2km long side square from T31TGK tile on August 6, 2018.