

Convolutional Neural Networks for estimating spatially-distributed evapotranspiration

Angel García-Pedrero^a, Consuelo Gonzalo-Martín^a, Mario F. Lillo-Saavedra^b, Dionisio Rodríguez-Esparragón^c, and Ernestina Menasalvas^a

^aCenter for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo, Pozuelo de Alarcón, Spain

^bFaculty of Agricultural Engineering and Water Research Center for Agriculture and Mining, CRHIAM, University of Concepción, Chile

^cEscuela Superior de Ingenieros de Telecomunicaciones, Universidad de la Palmas de Gran Canaria, Spain

ABSTRACT

Efficient water management in agriculture requires an accurate estimation of evapotranspiration (ET). There are available several balance energy surface models that provide a daily ET estimation (ET_d) spatially and temporarily distributed for different crops over wide areas. These models need infrared thermal spectral band (gathered from remotely sensors) to estimate sensible heat flux from the surface temperature. However, this spectral band is not available for most current operational remote sensors. Even though the good results provided by machine learning (ML) methods in many different areas, few works have applied these approaches for forecasting distributed ET_d on space and time when aforementioned information is missing. However, these methods do not exploit the land surface characteristics and the relationships among land covers producing estimation errors. In this work, we have developed and evaluated a methodology that provides spatial distributed estimates of ET_d without thermal information by means of Convolutional Neural Networks.

Keywords: Convolutional Neural Network, Evapotranspiration Estimation, METRIC

1. INTRODUCTION

The growing of the water demand has been a common problem throughout history as the human population has increased. Nowadays, new threats as the climate change can provoke alteration both water availability and water demand, with special effects in specific areas as agriculture.¹ According to the FAO, agriculture uses approximately 70% of the world's freshwater supply.¹ In this regard, one of the main current challenges relative to water use is the increase of water productivity in agriculture.² For this, it is crucial to have access to accurate information on agricultural water requirements.

Evapotranspiration (ET) represents the total amount of water lost, via transpiration and evaporation, from the canopy and soil in an area where crops are growing, its estimation provides noteworthy information regarding these requirements. However, the natural heterogeneity and complexity of agricultural and natural land surfaces make difficult the measuring and modeling of daily evapotranspiration (ET_d) values. The complexity of the problem increases when spatial and temporarily distributed (ET_d) information is needed for managing wide areas in different date. Some approaches to it can be found in the literature based on information gathered from aircraft or satellite platforms. These approaches can be grouped into two main types: i) methods that estimate the skin temperature from a vegetation index (VI), defined through visible and near infrared spectral bands, and the surface radiative temperature³⁻⁵ and ii) residual methods, which use the surface energy balance. These latter methods calculate ET_d by subtracting sensible heat and soil heat fluxes from net radiation.⁶ Of these, METRIC (Mapping Evapotranspiration at high Resolution using Internalized Calibration) model⁷ is one of the most widely used.

Further author information: (Send correspondence to A.G.-P.)

A.G.-P: E-mail: angel.garcia@ctb.upm.es, Telephone: +34 913 36 4664.

METRIC is based almost entirely on SEBAL model (Surface Energy Balance Algorithm for Land), developed by Bastiaanssen et al.⁸ Both models estimate crop ET_d , by solving the surface energy balance using spectral information from multispectral satellite images in the optical, near infrared and thermal ranges. Unfortunately, most of current operational remote sensors do not supply thermal band and consequently, the imagery provided can not be used as input to these models. Therefore, it is crucial for agricultural water management purposes the research of alternative methodologies that allow mapping ET_d , avoiding the need of thermal information.

One of the most promising area of research for forecasting distributed ET_d on space and time is Machine Learning. However, only few approaches can be found in the literature and all of them use standard Machine Learning algorithms as Supported Vector Machine⁹ or Relevance Vector Machine.¹⁰ However, these methods do not exploit the land surface characteristics and the relationships among land covers, producing large errors in estimates.

Recently deep learning-based methods have demonstrated excellent performance on different artificial intelligence tasks. In this regard, Convolutional Neural Networks (CNNs) play a major role for processing visual-related problems such as image classification, object detection, and face recognition. In particular, different works in the area of remote sensed imagery can be found in the literature.^{11–13} There are two main approaches to exploit the hierarchical analysis capacity of CNN for image analysis: feature extraction and classification. In feature extraction approach, pre-trained CNN models are used to automatically extract image features that later are analysed by traditional Machine Learning methods.¹⁴ In classification approach, a CNN is trained from scratch using a large set of images,¹⁵ which requires high performance equipment for processing (e.g., graphics processing units). Empirical studies have demonstrated that these methods often provided better results than traditional Machine Learning methods.^{16,17}

Therefore our work hypothesis states in which it is possible to train Convolution Neural Network (CNN) to generate a spatial distributed estimates of ET_d , without thermal spectral data. With the aim of validating this hypothesis, the goal of this paper is to evaluate the ability of CNN to estimate spatially distributed ET_d when thermal band is not available, and consequently the surface energy balance models can not be applied.

This paper is organized as follows. In Section 2 a brief introduction regarding CNN is included. The proposed methodology and the data set used for this study are described in section 3. The results are presented and discussed in Section 4. Finally, Section 5 summarizes the conclusions derived from the results.

2. CONVOLUTIONAL NEURAL NETWORKS

The concept of CNN is not new, in fact, it was first proposed in 1980 by¹⁸ with the name of NeoCognitron, and later refined by.¹⁹ CNNs are biological-inspired variants of feed-forward neural networks, where each layer is a non-linear feature detector performing local processing of contiguous features within each layer.²⁰ The output of each layer is a set of features of the image, which are obtained by the spatial convolution of a particular learned filter with the input-layer. This is the reason why the CNN architecture is able to exploit the strong spatially local correlation present in natural images by enforcing a local connectivity pattern between neurons of adjacent layers. Finally, this leads to a higher conceptual representation as information moves up to the output layer. In this context, a CNN is able to generate patterns from an image in a hierarchical manner similar to that of the mammalian visual cortex.

CNN architecture typically comprises several layers of different types:²¹

Convolutional layers. They compute the convolution of the input image with the weights of the network.

These layers are characterized by few parameters: the filter size, the filter spatial support, the step between different windows and an optional zero-padding which controls the size of the layer output. The analysis of the image is performed at different scales in the different layers. As the number of layers increases, the features extracted from the image are higher-level.

Pooling layers. These layers reduce the size of the input layer through some local non-linear operations. Their most important parameters are the support of the pooling window and the step between different windows.

Normalization layers. Their mission is improving the generalization capacity of the CNN by using inhibition schemes inspired in the real neurons of the brain. Neurons typically used in these layers are sigmoid type.

Fully-connected layers. These layers are typically used in the last levels of the network. They have the capacity of abstracting the low-level information generated in previous layers for a final decision.

3. DATA AND METHODS

3.1 Data description

The study site is located in the Central Valley in Chile, with central coordinates of $36^{\circ} 35' S$ and $72^{\circ} 00' W$. The scene is composed of a city (Chillán), rivers, mainly different annual crops and orchards, and alluvial soils, which allows a high production for different crops. Four different land covers are presented in the scene: urban, agricultural vegetation, forest and bare soil. The climate is warm temperate, with an annual mean temperature of $14^{\circ} C$, a short dry season and an annual rainfall ranging from 1,000 to 1,300 mm. Six images were captured by the ETM+ sensor on board the Landsat-7 satellite (path 233, row 85). They were downloaded from the USGS Glovis official site*, with an 1T preprocessing level of standard field correction. Table 1 shows the details (dates and names) of the images from the summer season that have been used in this research. The label used to name each image corresponds to the year–day in the Julian calendar. The size of the six Landsat-7 scenes is 947×702 pixels. Each pixel represents an area of 30 m x 30 m for all spectral bands, except for the thermal band. In this case each pixel represents an area of 60 m x 60 m. An RGB composite of the image registered on 12 January 2012 is displayed in Figure 1.

In addition, data from an Eddy Covariance System (EC), an automatic weather station, as well as, crop evapotranspiration ET_c and reference evapotranspiration ET_o values have been used in this study to characterize the training patterns. Besides data obtained from the satellite images, other input data to the METRIC model were obtained from a weather station. Then, the METRIC model has been applied to the whole dataset to obtain a ET_d map for each date.

Table 1: List of dates used in the study. The first row shows the year. The second row shows month/day. And the name of each scene is displayed in the third row as year-Julian day.

Year	2012		2013	2014		2015
Date	01/12	01/28	01/31	02/02	18/02	05/02
Name	2012-012	2012-028	2013-031	2014-033	2014-049	2015-036

3.2 Methodology

The workflow of the proposed methodology is shown in Figure 2. All images used have been preprocessed in order to ensure the quality of the data used for training the models. Outlier pixels with NaN and Inf values and the regions covered by clouds have been eliminated. Pixels in these maps with outlier ET_d values ($ET_d < 0$ and $ET_d > 11 \text{ mm} \cdot \text{day}^{-1}$) have been also eliminated. A particular CNN architecture is training in a standard pixel-approach. In this approach, during training phase, datasets are broken down into overlapping patches, where each patch is centered on a pixel which provides the response for the whole patch. In this work, the training dataset have been characterized by features obtained from different dates, but only patches centered on agricultural areas are considered. Thus for each date, the following data has been used to create patches that will feed the CNN:

- NDVI, LAI, and SAVI indexes;
- Blue, Green, Red, and NIR spectral bands; and

*<http://glovis.usgs.gov>

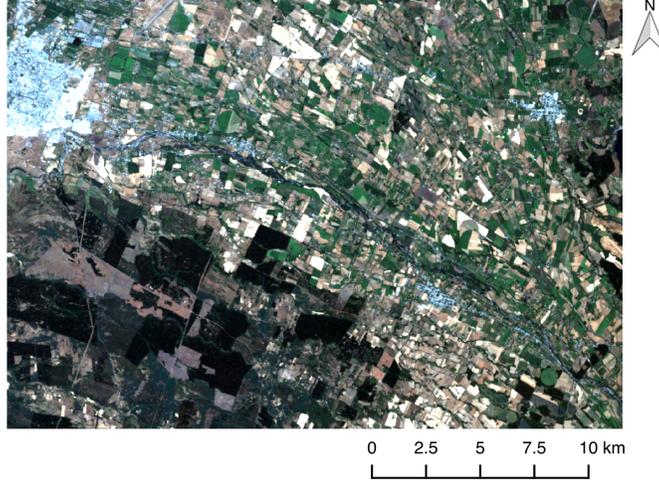


Figure 1: RGB composite of a Landsat-7 scene registered on 12 January 2012.

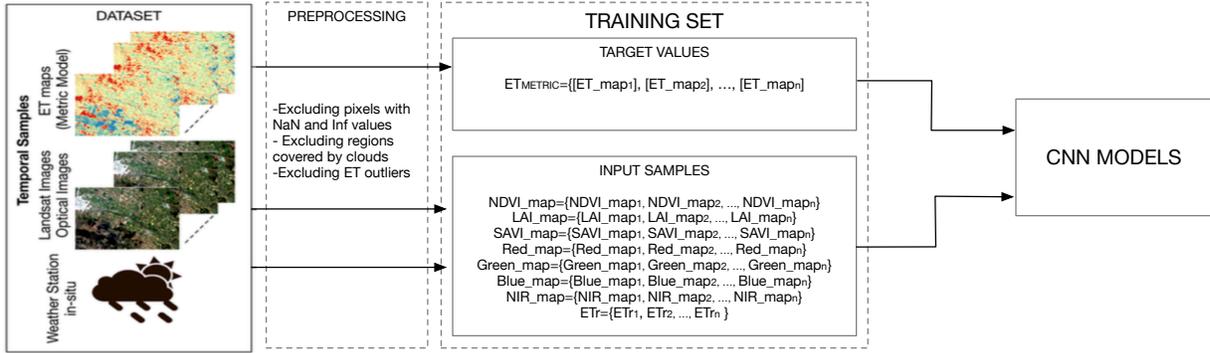


Figure 2: Overview of the CNN training process.

- ET_r values, obtained from a weather station.

Each patch is associated with its corresponding daily evapotranspiration target value provided by METRIC (ET_{METRIC}). In order to have a good representation of the data during training phase, patches are grouped according to their ET values into 5 bins. Then training and validation sets are creating by randomly selecting patches of each bin. Once a CNN is trained with the defined patches, it is applied to test images using a window-sliding approach in order to provide a ET_d value to each pixel in the test image or in other words a ET_d map.

3.2.1 Convolutional Neural Networks architecture

The CNN architecture built in this work is inspired on the approach proposed by Paisitkriangkrai et al.²² A scheme of this architecture is shown in Figure 3. It is defined by the alternation of three convolutional layers, which compute the convolution between the input of each layer and a set of learned filters; and other additional layers (max-pooling layers), which apply a non-linear transformation (rectified linear unit - ReLU) and subsamples the output of the corresponding convolutional layer, respectively. The role of these additional layers is to improve the robustness of the network to distortions and small translations. Moreover, two fully-connected layers are included at the end of the architecture. To reduce overfitting in the fully-connected layers, the dropout

method has been used.²³ The output of these layers feeds the output layer with only one output. A Euclidean loss layer is used to evaluate the error of the network. This error function computes the sum of squares of differences of its two inputs (i.e. target and estimated values), which is useful for regression problems. In Figure 3 the expression $i \times j \times k$ under each convolutional layer represents the size of the kernels associated to this layer, where the number of kernels for the three first layers are 32, 64, and 128, respectively. Each of these kernels generates a feature map for feeding the next layer.

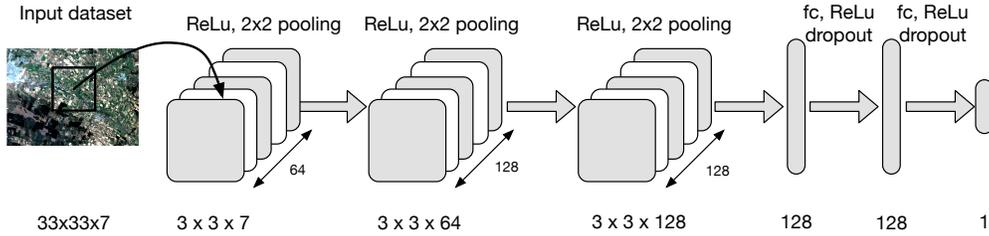


Figure 3: CNN architecture used during the experiments.

To prove the ability of the CNNs to estimate ET values from an unseen image (real-world case) 6 different training datasets have been generated using the different available images. These datasets have been built by removing a certain date, which is used for testing, and using the remaining image series for training the models.

4. RESULTS AND DISCUSSION

By applying the methodology described in the previous section, 6 different CNN models have been obtained. The absolute differences between the ET_{METRIC} maps and the maps generated by CNN models are shown in Figure 4. As might be expected, estimates in agricultural areas are better than in other areas such as urban areas. This is because the models were generated especially for agricultural applications, but similar results can be expected if the entire image is used during training phase.

Table 2 summarizes the assessment of all these models in terms of the Root Mean Squared Error (RMSE) and the R squared (R^2). As can be seen, the largest error, in terms of RMSE, is 3.293 while the mean RMSE value is 2.122 ± 0.591 . On the other hand, models have a mean R^2 value of 0.591 ± 0.100 . The low RMSE values, in most of the cases, indicate the goodness of fit of the CNN models. Only some dates present a low R^2 values (e.g. 2013-031). In fact, the difference between target (METRIC) and estimated ET maps (CNN), in terms of evapotranspiration estimation, is almost negligible in comparison with the error of some traditional methods²⁴ (e.g. vegetation index, Crop Coefficient relationship).

Table 2: Assessment of the CNN models. RMSE and R^2 are shown for each estimated date.

Date	RMSE	R^2
2012-012	1.984	0.713
2012-028	1.958	0.706
2013-031	1.145	0.467
2014-033	2.740	0.594
2014-049	1.612	0.544
2015-036	3.293	0.523

Figure 5 shows, through scatterplots, the relationship between ET_{METRIC} and the obtained by CNNs models. It seems, from them, that CNN models tend to over-estimated ET values. In particular, CNN models have problems when target values are highly concentrated in a narrow range, i.e. from 0 to 6 mm/day. This may be because CNN models did not have sufficient training patterns with the same characteristics as the estimated date, which is a common issue in machine learning applications, models are not able to predict things that they have not seen before.

All computations were performed using an NVIDIA GTX 1080 card. CNN codes are implemented in python using Caffe framework.²⁵

5. CONCLUSIONS

In this work, the use of CNN models for estimating spatially distributed ET have been explored. These models has been trained using remote sensing data without using thermal data, which it opens up an opportunity to use various operational satellites, and to increase the resolution of the ET maps which can contribute to improve the agricultural irrigation and the yield of the fields. From obtained results, it can be concluded that CNN models are able to estimate satisfactorily ET maps. Although in some cases discrepancies between METRIC and CNN estimates are in average of 2.11 mm/day , these errors are tolerable for most agricultural applications.

Two main aspects will be improved in a future research: (i) decreasing the estimation errors by using more deeper CNN models and (ii) including other type of information such as precipitations. Other aspect to consider is increasing the number of images for training, however this will increase also training times. Additional research is necessary to explore the generalization of the obtained models in similar data from other satellite platforms with a higher resolution such as WorldView-2.

ACKNOWLEDGMENTS

This work has been funded by the Agencia Estatal de Investigación (AEI) of Spain and the European Regional Development Fund (ERDF) through the project ARTeMISat-2: Advanced Processing of Remote Sensing Data for Monitoring and Sustainable Management of Marine and Terrestrial Resources in Vulnerable Ecosystems (CTM2016-77733-R).

REFERENCES

- [1] Wang, X.-j., Zhang, J.-y., Shahid, S., Guan, E.-h., Wu, Y.-x., Gao, J., and He, R.-m., "Adaptation to climate change impacts on water demand," *Mitigation and Adaptation Strategies for Global Change* **21**(1), 81–99 (2016).
- [2] Snyder, R., Pedras, C., Montazar, A., Henry, J., and Ackley, D., "Advances in et-based landscape irrigation management," *Agricultural Water Management* **147**, 187–197 (2015).
- [3] Jiang, L. and Islam, S., "An intercomparison of regional latent heat flux estimation using remote sensing data," *Int. J. Remote Sens.* **24**(11), 2221–2236 (2003).
- [4] Nemani, R. R. and Running, S. W., "Estimation of regional surface resistance to evapotranspiration from ndvi and thermal-ir avhrr data," *J. Appl. Meteorol.* **28**(4), 276–284 (1989).
- [5] Gillies, R. R., Kustas, W. P., and Humes, K. S., "A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the normalized difference vegetation index (ndvi) and surface e," *International Journal of Remote Sensing* **18**(15), 3145–3166 (1997).
- [6] Gowda, P. H., Chavez, J. L., Colaizzi, P. D., Evett, S. R., Howell, T. A., and Tolk, J. A., "Et mapping for agricultural water management: present status and challenges," *Irrigation science* **26**(3), 223–237 (2008).
- [7] Allen, R., Tasumi, M., and Trezza, R., "Satellite-based energy balance for mapping evapotranspiration with internalized calibration (metric). model," *Journal of Irrigation and Drainage Engineering* **133**(4), 380–394 (2007).
- [8] Bastiaanssen, W., Meneti, M., Feddes, R., and Holtslag, A., "A remote sensing surface energy balance algorithm for land (sebal). formulation," *Journal of Hydrology* **212-213**, 198–212 (1998).
- [9] Kaheil, Y. H., Rosero, E., Gill, M. K., McKee, M., and Bastidas, L. A., "Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines," *IEEE Transactions on Geoscience and Remote Sensing* **46**(9), 2692–2707 (2008).
- [10] Bachour, R., Walker, W. R., Ticlavilca, A. M., McKee, M., and Maslova, I., "Estimation of Spatially Distributed Evapotranspiration Using Remote Sensing and a Relevance Vector Machine," *Journal of Irrigation and Drainage Engineering* **140**(8), 04014029 (2014).
- [11] Zhang, L., Zhang, L., and Du, B., "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine* **4**(2), 22–40 (2016).

- [12] Gonzalo-Martin, C., Garcia-Pedrero, A., Lillo-Saavedra, M., and Menasalvas, E., “Deep learning for superpixel-based classification of remote sensing images,” (2016).
- [13] Hu, F., Xia, G.-S., Hu, J., and Zhang, L., “Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery,” *Remote Sensing* **7**(11), 14680–14707 (2015).
- [14] Sharif Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S., “Cnn features off-the-shelf: an astounding baseline for recognition,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*], 806–813 (2014).
- [15] Krizhevsky, A., Sutskever, I., and Hinton, G. E., “Imagenet classification with deep convolutional neural networks,” in [*Advances in neural information processing systems*], 1097–1105 (2012).
- [16] Larochelle, H., Bengio, Y., Louradour, J., and Lamblin, P., “Exploring strategies for training deep neural networks,” *The Journal of Machine Learning Research* **10**, 1–40 (2009).
- [17] Salakhutdinov, R. and Hinton, G. E., “Deep boltzmann machines,” in [*International conference on artificial intelligence and statistics*], 448–455 (2009).
- [18] Fukushima, K., “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological cybernetics* **36**(4), 193–202 (1980).
- [19] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D., “Backpropagation applied to handwritten zip code recognition,” *Neural computation* **1**(4), 541–551 (1989).
- [20] Biem, A., “Neural networks: A review,” in [*Data Classification: Algorithms and Applications*], Aggarwal, C. C., ed., ch. 8, 206–236, Chapman and Hall/CRC (2014).
- [21] Castelluccio, M., Poggi, G., Sansone, C., and Verdoliva, L., “Land use classification in remote sensing images by convolutional neural networks,” *arXiv preprint arXiv:1508.00092* (2015).
- [22] Paisitkriangkrai, S., Sherrah, J., Janney, P., and Hengel, A., “Effective semantic pixel labelling with convolutional networks and conditional random fields,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*], 36–43 (2015).
- [23] Hinton, G., Deng, L., Yu, D., Dahl, G. E., r. Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., and Kingsbury, B., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Processing Magazine* **29**(6), 82–97 (2012).
- [24] Tang, Q., Peterson, S., Cuenca, R. H., Hagimoto, Y., and Lettenmaier, D. P., “Satellite-based near-real-time estimation of irrigated crop water consumption,” *Journal of Geophysical Research: Atmospheres* **114**(D5), n/a–n/a (2009).
- [25] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., “Caffe: Convolutional architecture for fast feature embedding,” *arXiv preprint arXiv:1408.5093* (2014).

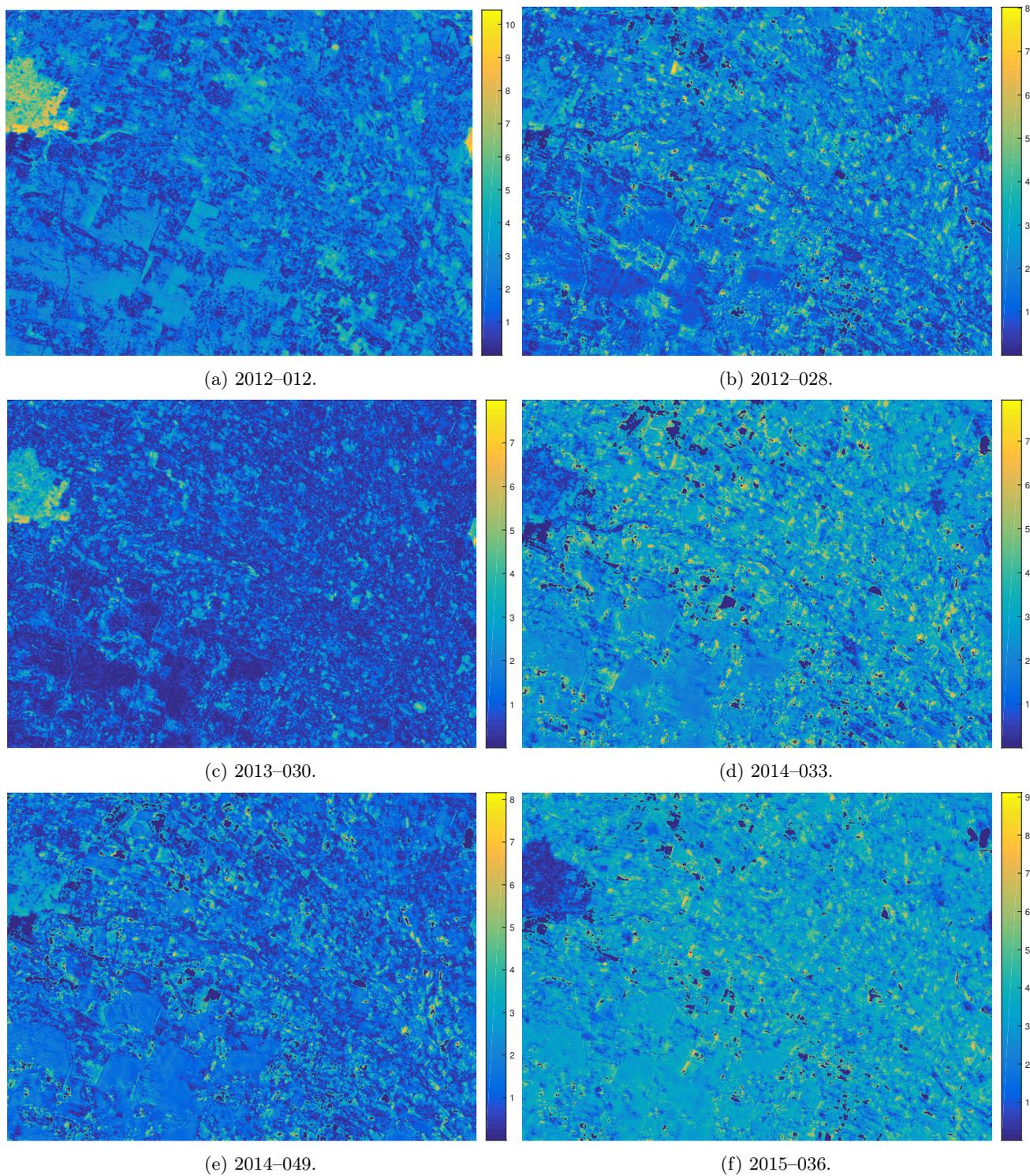
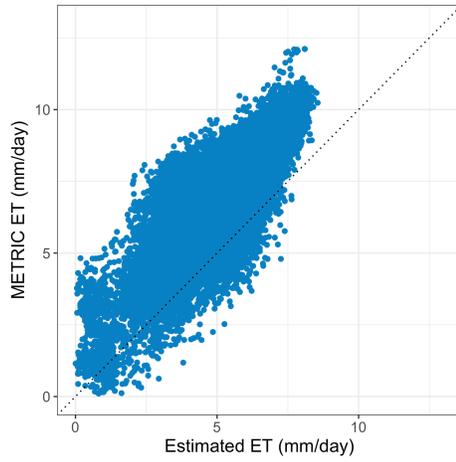
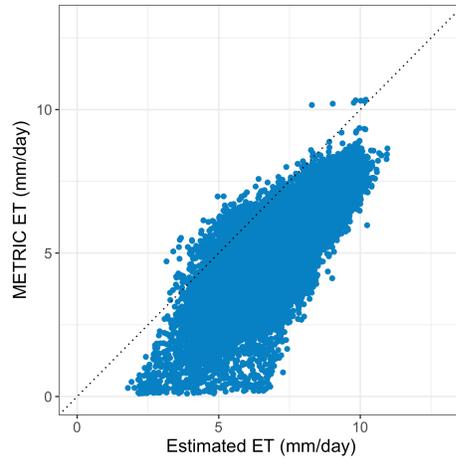


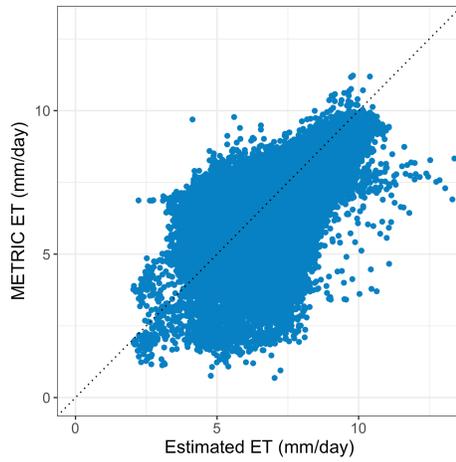
Figure 4: Maps representing the absolute difference between the actual ET_{map} and the ET_{map} estimated by CNNs.



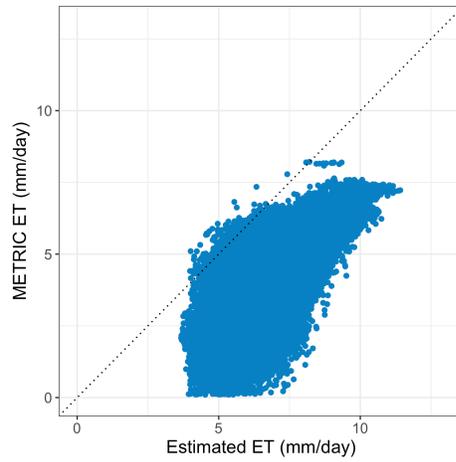
(a) 2012-012.



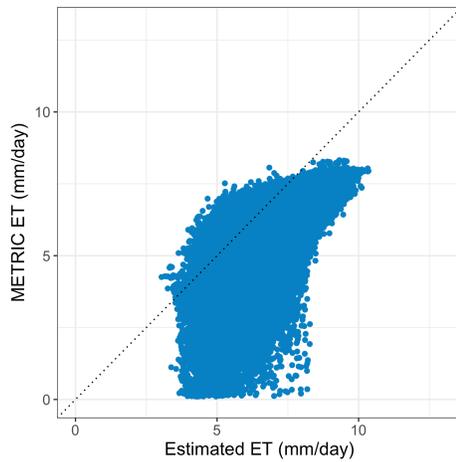
(b) 2012-028.



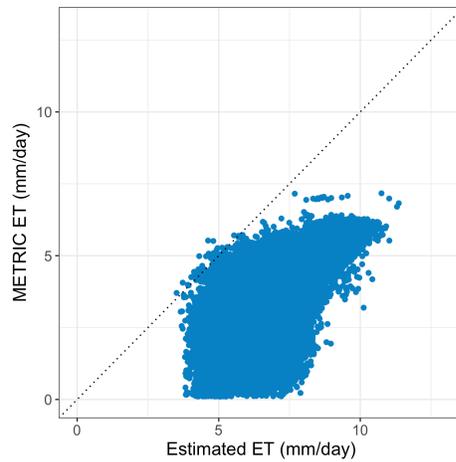
(c) 2013-030.



(d) 2014-033.



(e) 2014-049.



(f) 2015-036.

Figure 5: Scatterplots representing the distribution of evapotranspiration values provided by METRIC (actual) and CNN (estimated).