

Evaluating an energy balance setting and random forest-based downscaling for the estimation of daily ET at sub-kilometer spatial resolution

José Miguel Barrios, Nicolas Ghilain, Alirio Arboleda, Françoise Gellens-Meulenberghs
Royal Meteorological Institute of Belgium
Avenue Circulaire 3
B-1180 Uccle, Belgium
Email: jose.miguel.barrios@meteo.be

Abstract—A large number of processes are impacted by the magnitude and timing of water fluxes. A proper monitoring of these processes would benefit from periodic estimation of evapotranspiration (ET) with enough spatial detail to discriminate the different components of the landscape. However, high temporal and high/moderate spatial resolution are rarely present in the same ET product. In this study, we aimed at merging the strengths of geostationary and polar-orbit satellites to estimate daily ET at the native spatial resolution of Proba-V, 0.3 km. Two modelling paths were examined: the conduction of an energy balance at the target spatial resolution, and; energy balance at a coarser resolution followed by a statistical downscaling procedure based on building a Random Forest. The study area was the upper catchment of the Biebrza river in North Eastern Poland. The comparison of the modelled values with ET derived from *in situ* measurements (eddy covariance) indicate that both approaches performed good (Nash-Sutcliffe efficiency coefficient of 0.7589 and 0.7857 for first and second approach, respectively). The modelled spatial and temporal patterns were similar but the statistical downscaling path tended to generate lower ET estimates (bias equal to 0.0071 and -0.2254 for first and second approach, respectively).

I. INTRODUCTION

Practical applications of ET estimates (e.g. ecosystem health monitoring, agriculture monitoring, yield prediction, etc.) often require daily or weekly estimates with enough spatial detail to discriminate contiguous ecosystems. This aspect has encouraged research towards combining the strengths of geostationary and polar-orbit satellites. In this study we examined two strategies to generate daily estimates of ET at the native resolution of Proba-V (~ 0.3 km). Working at that spatial resolution might open the gate to the use of Sentinel-3 in ET modelling, given the continuity criterion linking both missions.

The first strategy consisted in the pixelwise computation of an energy balance at the target spatial resolution. The second strategy departs from an energy balance conducted the same way as in the first strategy but at a coarser spatial resolution (~ 1 km), followed by a downscaling procedure based on building a random forest to generate the target ET estimates. A major advantage of the latter procedure was its higher computational speed.

The results were compared to *in situ* measurements at an eddy covariance (EC) tower. In addition, the results were displayed in the form of ET maps which allowed the visual assessment of the results and the verification of similarities and differences in spatial pattern between spatial resolutions and methodological approaches.

II. METHODS

A. Study area

The study area comprised over 35000 hectares and was located in the upper catchment of the Biebrza river, in North Eastern Poland. The upper catchment of the Biebrza River is an area of great ecological value. It is a Natura 2000 site and is listed in the RAMSAR convention for the wetlands protection. The maps in Figure 1 illustrate the location of the study area as well as the land cover composition of the study area according to the CORINE land cover map [1]. The studied period was from March to July 2015.

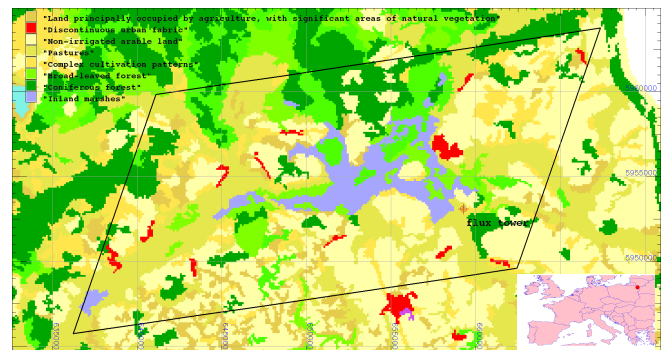


Fig. 1. Delimitation of study area (black line) and land cover according to the CORINE land cover map. The location of the flux tower is indicated. Coordinates in UTM projection (UTM zone 34 North). *Inserted:* Geographic location of the study area in North-Eastern Poland

B. Calculation procedure

Two paths were followed for estimating daily ET at 0.3 km resolution ($ET_{d0.3km}$). They are briefly described in the following sections.

1) *Energy balance*: The basic principle in this approach is the computation of an energy balance. The methods for doing so were based on the principles and procedures behind the LSA-SAF ET product [2], [3]. The LSA-SAF ET product is an operational service delivering ET estimates in near-real time at continental scale (see www.lsasaf.meteo.pt).

The basic form of the energy balance can be written as $R_{net} = LE + H + G$ with R_{net} : net radiation, LE : latent heat flux, H : sensible heat flux; and, G : ground heat flux. This equation was solved iteratively for subpixel homogeneous units (tiles) which are then aggregated to obtain a pixel level estimate. In the LSA-SAF ET product the energy balance is forced by remote sensing products derived from the Meteosat Second Generation (MSG) satellite, a set of meteorological variables obtained from the operational forecast by ECMWF and information and parameters related to land cover obtained from the ECOCLIMAP database [4]. In this study, ~ 1 km and ~ 0.3 km grids were adopted for the computation of half-hourly ET. A higher spatial detail was attained by introducing data from PROBA-V (~ 1 km and ~ 0.3 km), MODIS (MCD43B3 and MCD43A3 resampled to 0.3m) and, CORINE land cover map [1] for the ~ 0.3 km resolution. This procedure is summarized schematically in Figure 2.

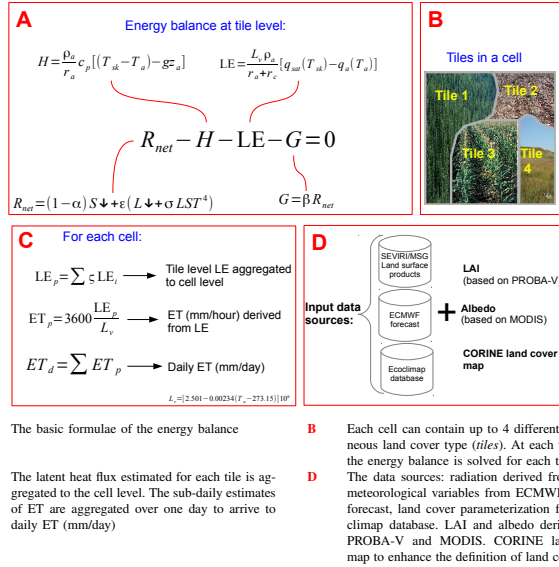


Fig. 2. Basic formulation and main methodological principles of daily ET calculation of the basis of an energy balance

As in the LSA-SAF product, the actual computation of ET occurred at a timestep of 30 minutes and daily ET estimates resulted from aggregating these half-hourly estimates over the duration of one day. At each timestep, estimates of land surface temperature (skin temperature) were generated in the iteration process.

2) *Random forest*: The random forest [5] is a data-driven approach whereby a number of regression trees are grown ($n=500$ in this study) on bootstrap subsamples of a given training dataset. The technique has been successfully used in downscaling (e.g. [6]) and presents the advantage of not requiring beforehand knowledge on the mathematical relation among the considered variables. In this study, the construction of a random forest aimed at deriving $ET_{d0.3km}$ as a function of ET_{d1km} , leaf area index (LAI), albedo and the occupation fraction of land cover classes (tiles) in the 0.3km pixel. Equation 1 summarizes this formulation.

$$ET_{d0.3km} \sim ET_{d1km}, LAI_{0.3km}, albedo_{0.3km}, t_1, \dots, t_{12} \quad (1)$$

where LAI was derived from Proba-V NDVI, albedo was obtained from MODIS (MCD43B3 product); and $t_1 \dots t_{12}$ denoted the fraction of each tile within the pixel at a given date (t_1 = bare soil, t_2 =water, t_3 =C3 crops, etc).

Since no reference $ET_{d0.3km}$ data were available, the random forest was trained with values at ~ 1 km resolution between March and July 2015. The coarser ET_d values acting as independent variable in the model were derived from resampling the ET_{d1km} to a 3 km spatial resolution grid. For the composition of the training dataset a random sampling was performed over the 1km pixels across the study area and the dates in the studied period. The sample size was arbitrarily set to 30% of the 1km pixels and 30% of the dates. Figure 3 presents a schematical view on the downscaling process.

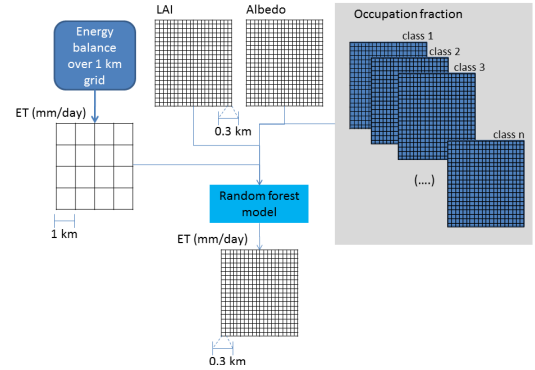


Fig. 3. Methodology for ET downscaling from ~ 1 km to ~ 0.3 km spatial resolution

C. Validation

The flux tower in the study area (N $53^\circ 35' 20''$, E $22^\circ 53' 29''$) supplied half-hourly latent heat values. These values were converted into daily ET and used to assess the accuracy of the ET_{d1km} and $ET_{d0.3km}$ modeled time series for the pixels containing the tower.

III. RESULTS

The agreement between the modelled output and the validation dataset at the location of the EC tower is summarized in the statistical parameters shown in Table I. This table also includes the parameter values for ET_{d1km} .

The values indicated a general good correspondence between modelled and validation values with NSE values above 0.75 and r^2 above 0.89. The $ETd_{0.3km}$ series obtained by random forest downscaling exhibited, however negative bias which might be the effect of propagation of the negative bias obtained in the ETd_{1km} simulations.

TABLE I. STATISTICAL INDICATORS OF AGREEMENT BETWEEN VALIDATION AND MODELLED DATA FOR THE PIXEL AT THE LOCATION OF THE EC TOWER

model	rmse	bias	NSE*	r^2	n
ETd_{1km} (energy balance)	0.5568	-0.1366	0.7931	0.8992	129
$ETd_{0.3km}$ (energy balance)	0.6052	0.0071	0.7589	0.8913	131
$ETd_{0.3km}$ (random forest)	0.5667	-0.2254	0.7857	0.9068	129

*Nash-Sutcliffe Efficiency coefficient [7]

Figure 4 shows the time series of modelled and validation daily ET at the location of the tower (top) as well as ET maps across the study area for two dates, May 30, 2015 and July 9, 2015 (bottom). The signals in the plot indicate that the largest deviation between models occurred during the period june-july 2015. During an important fraction of this period the ETd values derived from random forest downscaling were lower than the validation dataset. In contrast, the $ETd_{0.3km}$ series obtained from the energy balance calculation tended to generate higher values which, for certain dates, exceeded significantly the validation data. This is likely to happen in other locations of the study area as well.

The ET maps in Figure 4 illustrate the difference in spatial detail between ET_{1km} and $ET_{0.3km}$. More importantly, the maps also show that despite the difference in ETd values

between the energy balance and the random forest approaches, both models revealed similar spatial ET patterns.

IV. CONCLUSION

Two paths for deriving daily ET estimates were examined in this study. The comparison with validation data showed that both modelling approaches obtained similar scores. The spatial patterns revealed by both approaches was similar as well. The most remarkable difference in the output of both approaches is that random forest-based estimates tend to be lower.

ACKNOWLEDGEMENT

The research pertaining to these results received financial aid from the Belgian Federal Science Policy according to the agreement of subsidy no. [SR/34/163]. Further research on the topic for wetland ecosystems is currently being conducted by the HiWet project of the STEREO III programme (contract SR/01/301) supported by the Belgian Federal Science Policy.

The authors acknowledge the providers of the EC data.

REFERENCES

- [1] European Environment Agency, "CORINE land cover 2000," <http://www.eea.europa.eu/themes/landuse/clc-download>, 2009.
- [2] N. Ghilain, A. Arboleda, and F. Gellens-Meulenberghs, "Evapotranspiration modelling at large scale using near-real time MSG SEVIRI derived data," *Hydrology and Earth System Sciences*, vol. 15, no. 0, pp. 771–786, 2011.

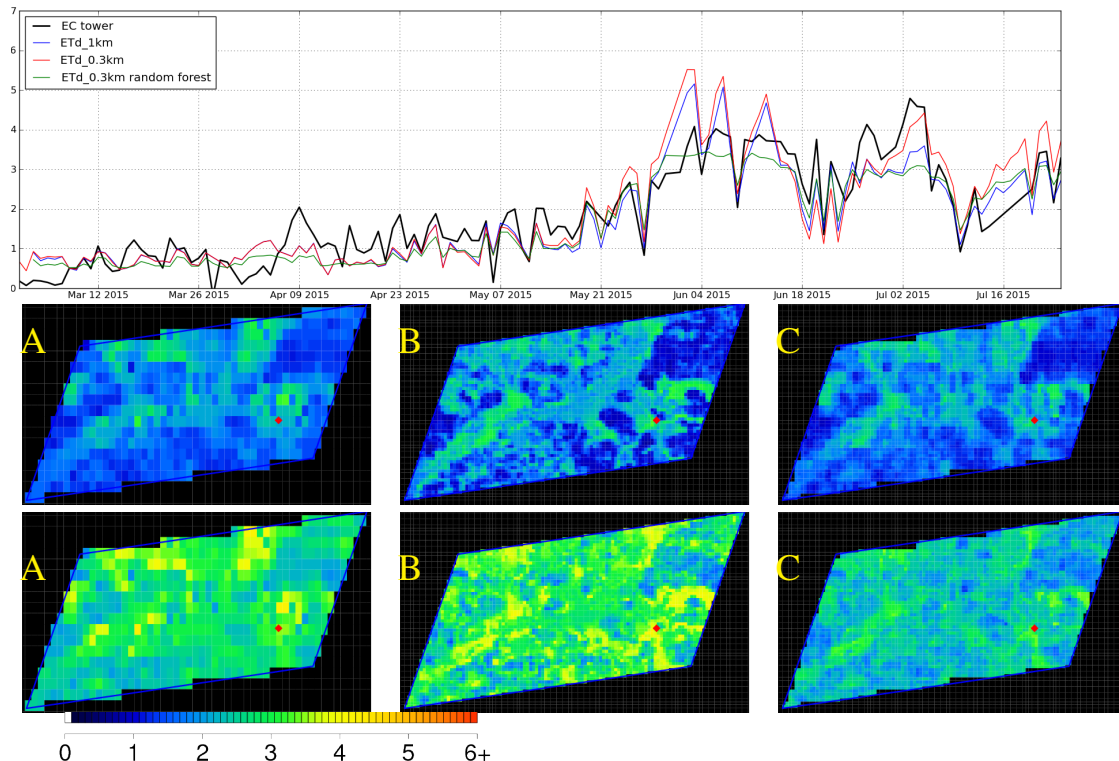


Fig. 4. Daily ET (mm/day) as modelled at 1km (A), 0.3km (energy balance) (B) and 0.3km (random forest)(C) for May 30, 2015 (top) and July 9, 2015 (bottom) over the upper part of the Biebrza catchment. The red diamond symbol represents the location of the EC tower

- [3] Royal Meteorological Institute of Belgium, *Algorithm Theoretical Basis Document- Evapotranspiration (MET) Product - Daily MET (DMET) Product*, LSA-SAF, March 2010, <http://landsaf.meteo.pt>.
- [4] V. Masson, J.-L. Champeaux, F. Chauvin, C. Meriguet, and R. Lacaze, "A global database of land surface parameters at 1-km resolution in meteorological and climate models," *Journal of climate*, vol. 16, no. 9, 2003.
- [5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: <http://dx.doi.org/10.1023/A:1010933404324>
- [6] C. Hutengs and M. Vohland, "Downscaling land surface temperatures from modis data to mesoscale resolution with random forest regression," in *Gemeinsame Tagung 2014 der DGfK, der DGPF, der GfGI und des GiN*, E. Seyfert, E. Glch, C. Heipke, J. Schiewe, and M. Sester, Eds. Publikationen der Deutschen Gesellschaft fr Photogrammetrie, Fernerkundung und Geoinformation (DGPF) e.V., 2014, pp. 1–10.
- [7] J. Nash and J. Sutcliffe, "River flow forecasting through conceptual models part i -a discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970.