

Supplementary Material

Francisco Zambrano

1 Validation of ERA5L variables

1.1 Methods

We compared the ERA5L variables for monthly mean temperature, total precipitation, and volumetric soil water content against values retrieved by weather stations. For temperature and precipitation, we used the weather network from the Ministry of Agriculture of Chile (www.agromet.com) between 2015 and 2023. We used 277 stations located throughout Chile. For soil moisture, we select a private soil network that is owned by the agricultural enterprise Garces Fruit, which has 99 stations in Central Chile, located in cherry fruit crops. The sensors are installed at 30, 60, and 90m and are the model Teros 12 from MeterGroup. To avoid the effect of irrigation on soil moisture, which ERA5L hardly captures, we used daily data for the year 2022 and the months outside the growing season, May to September.

To account for the performance of the ERA5L climatic variables regarding the values measured by the weather stations. We selected the following metrics:

$$\begin{aligned}MAE &= \frac{1}{n} \sum |E - S| \\Bias &= \frac{\sum E}{\sum S} \\ubRMSE &= \sqrt{\frac{\sum [(E_i - \bar{E}) - (S_i - \bar{S})]^2}{n}} \\CC &= \frac{\sum (S_i - \bar{S})(E_i - \bar{E})}{\sqrt{(\sum (S_i - \bar{S})^2)(\sum (E_i - \bar{E})^2)}}\end{aligned}$$

MAE: mean absolute error *bias*: bias *ubRMSE*: unbiased root mean squared error *CC*: coefficient of correlation *S*: value of the variable measure by the weather station *E*: value of the variable measure by ERA5L

1.2 Results

The average metrics of performance of ERA5L over the 266 weather stations were in the case of monthly temperature: $ubRMSE = 1.06^{\circ}C$, $MAE = 1.131^{\circ}C$, and $CC = 0.963$, showing a good agreement, low error, and low overestimation. For cumulative monthly precipitation, $MAE = 28.1\text{ mm}$, $bias = 1.93$, and $CC = 0.845$, showing a high correlation and a 93% bias and being overestimated by ERA5L. In the case of the 97 soil moisture stations, we averaged for the three depths (30, 60, and 90m) and then compared it with volumetric water content at 1m derived from ERA5L. For this case, we made a daily comparison, having a $CC = 0.71$, $RMSE = 0.174\text{ m}^3\text{m}^{-3}$, $MAE = 0.167\text{ m}^3\text{m}^{-3}$, and $bias = 1.74$. The ERA5 soil moisture overestimate is 74%, but it has a kind of good correlation.

2 Land cover macroclasses and validation

2.1 Methods

To analyze the LULCC, we use the IGBP scheme from the MCD12Q1 collection 6.1 from MODIS. This product has a yearly frequency from 2001 to 2022. The IGBP defines 17 classes; from these, we regrouped into ten macroclasses, as follows: classes 1-4 to forest, 5-7 to shrublands, 8-9 to savannas, 10 as grasslands, 11 as wetlands, 12 and 14 to croplands, 13 as urban, 15 as snow and ice, 16 as barren, and 17 to water bodies. Thus, we have a land cover raster time series with the ten classes for 2001 and 2023.

To validate the land cover obtained, we compare the macroclasses with the ones of a more detailed land cover map made by @Zhao2016 for Chile with samples acquired in the years 2013–2014 (LCChile). The later has a spatial resolution of 30 m and three levels of defined classes; from those, we used level 1, which fits with the macroclasses land cover. We chose the years 2013 (IGBP2013) and 2014 (IGBP2014) from land cover macroclasses to validate with LCChile.

We follow the next procedure:

- i) resampled LCChile to the spatial resolution (500m) of the land cover macroclasses using the nearest neighbor method,
- ii) took a random sample of 1000 points within continental Chile and extracted the classes that fell within each point for LCChile, IGBP2013, and IGBP2014; we considered the point extracted from LCChile as the truth and the values from the other two years as predictions.
- iii) calculate a confusion matrix with the classes extracted from the 1000 points for LCChile, IGBP2013, and IGBP2014. Calculate the performance metrics of accuracy and F1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\text{correct classifications}}{\text{all classifications}}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

where TP and FN refer to true positive and false negative, correctly classified classes; TN and FP to true negative and false positive, wrongly classified classes.

2.2 Results

For vegetation, we obtained and use hereafter five macroclasses of land cover from IGBP MODIS: forest, shrubland, savanna, grassland, and croplands. Figure 1 c shows the spatial distribution of the macroclasses through Chile for the year 2022. The validation of IGBP2013 and IGBP2014 with LCChile reached near the same metrics of performance, having an accuracy of ~ 0.82 and a F1 score of ~ 0.66 .