From Points to Areas: Spatial Interpolations

William H. Aeberhard

Swiss Data Science Center (SDSC), ETH Zurich, Switzerland

Fore[st]cast Annual Meeting
Zurich, Switzerland

July 10, 2025

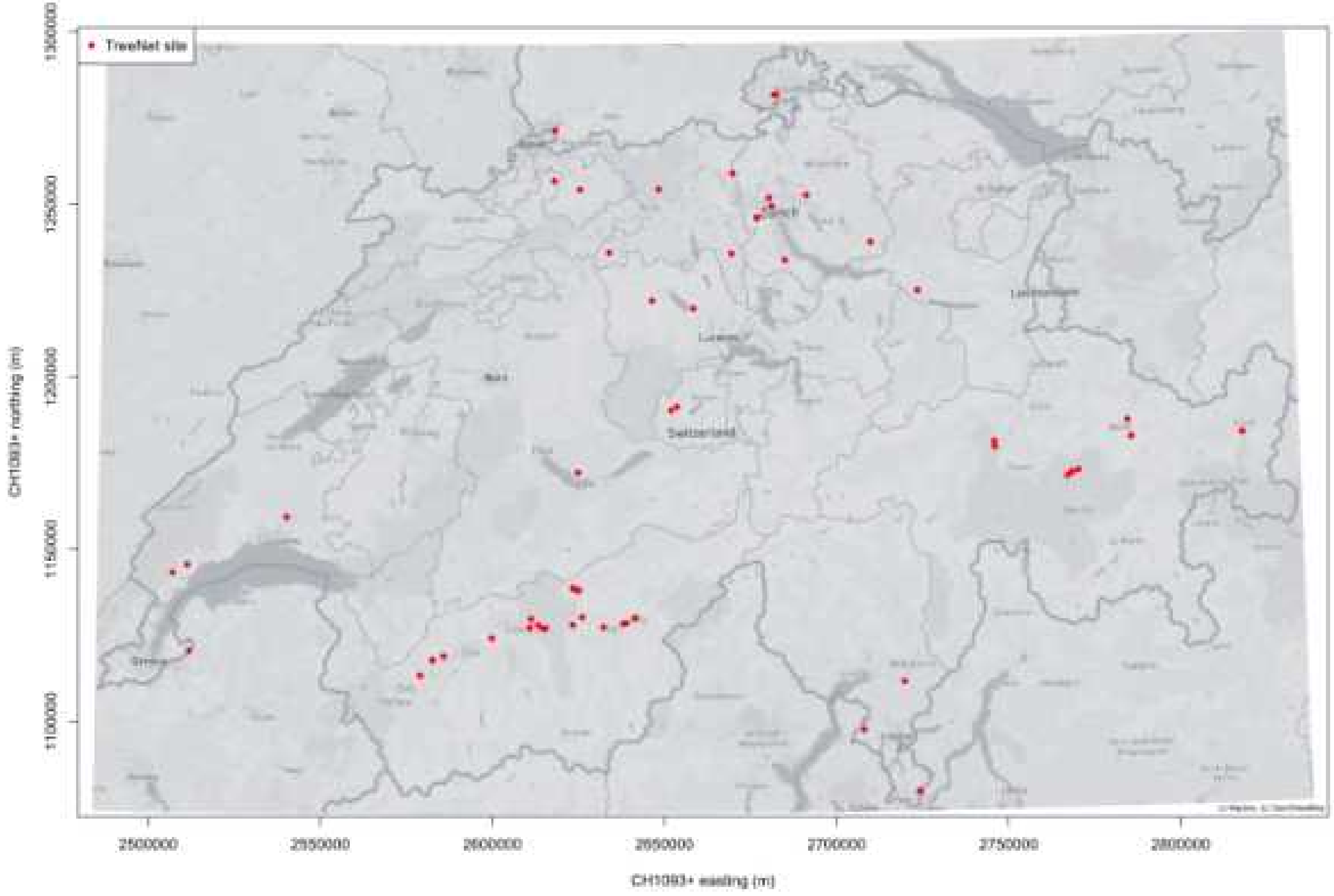
Outline

1. Introduction to TWD Spatial Prediction
2. Data
3. Long Short-Term Memory Recurrent Neural Network
4. Results
5. Discussion

Fore[st]cast: Spatial Prediction Task Overview

- Main goal of Fore[st]cast WP3: create smooth maps of tree water deficit (TWD) covering Switzerland.
- Challenge: TWD measurements only at a few locations. Spatial prediction becomes a difficult extrapolation task.
- That said, we have TWD time series at each location, along with gridded environmental features (from COSMO)
- Solution: transform the spatial prediction problem into a temporal prediction problem. By restricting the predictive model to be non-autoregressive, and rely only on gridded environmental features, it can in principle predict at arbitrary locations.

TreeNet sites in Switzerland



Training Data

- Time window: 2018-01-01 till 2022-12-31 (five years, 1826 days overall). Restrict to series growth period, so the number of available observations ranges from ~ 100 to ~ 1000 for each series, with median around 400.
- Focus on three abundant tree species in Switzerland:
  - European beech (*Fagus sylvatica*)
  - Norway spruce (*Picea abies*)
  - Scots pine (*Pinus sylvestris*)
- Target variable is a "TWD index": start from TWD series normalized at a daily resolution by dividing by the maximum daily shrinkage, then normalize by the series' min and max over training time window. Thus index $\in [0, 1]$, where 0 retains the interpretation of TWD being zero and 1 means that the TWD series achieved its maximum value in the 2018–2022 window.

Training Data (cont'd)

- Site-species median aggregation: 46 "series" over 37 sites for the three species.
- Covariates from COSMO-1E on a regular grid, pointwise mean of 11 ensemble members, daily mean of hourly values:
  - total precipitation
  - air temperature
  - wind speed
  - dew point
  - net short wave radiation flux
  - net long wave radiation flux
- Additional input features:
  - vapor pressure deficit (VPD), computed from air and dew point temperature (fixed atmospheric presure of 101 kPa)
  - soil water potential (SWP) computed from the BROOK90 model (Federer et al., 2003)
  - day of the year (DOY)
  - static features like projected coordinates and elevation

LSTM Neural Network I

For a single time series, the long short-term memory (LSTM; Hochreiter and Schmidhuber, 1997) class of recurrent neural network architectures can be summarized by the following recursive equations, where t denotes the discrete time index (days here):

$$
\begin{aligned}
i_t &= \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) && \text{(input gate)}\\
f_t &= \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) && \text{(forget gate)}\\
\tilde{c}_t &= \tanh(W_{ic}x_t + b_{ic} + W_{hc}h_{t-1} + b_{hc}) && \text{(candidate cell state)}\\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t && \text{(cell state)}\\
o_t &= \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) && \text{(output gate)}\\
h_t &= o_t \odot \tanh(c_t) && \text{(hidden state)}\\
\hat{y}_t &= \sigma(W_{hy}h_t + b_y), && \text{(prediction)}
\end{aligned}
$$

where...

LSTM Neural Network II

- x_t is a vector of p covariate values at (or up to) time t
- $\hat{y}_t$ is the scalar pointwise prediction for the TWD index y_t
- all Ws are weight matrices and all bs are bias vectors (parameters to estimate)
- σ is the (standard logistic) sigmoid function mapping the real line to $(0, 1)$
- tanh is the hyperbolic tangent function mapping the real line to $(-1, +1)$
- $\odot$ denotes the Hadamard (elementwise) product

LSTM Neural Network III

- All activation functions are applied elementwise. The recursive equations are initialized at $t = 1$ with zeroes for h_0 and c_0.
- Main hyperparameter tuning the complexity of the model: dimension d of the hidden state h. The total number of free parameters is $4pd + 4d^2 + 9d + 1$.
- The parameters are estimated by minimizing the mean squared (MSE) loss function $\frac{1}{nT}\sum_{s=1}^{n}\sum_{t=1}^{T}(y_{s,t} - \hat{y}_{s,t})^2$, where s identifies the sites.
- Optimization details: RMSprop with momentum $\mu = 0.5$, slight weight decay with $\lambda = 0.1$ (all features are normalized), learning rate scheduled from 0.1 to 0.001, early stopping with at most 200 epochs

LSTM Neural Network IV

- The observed y_t does not enter any of the LSTM equations $\Rightarrow$ the model is (by design) non-autoregressive.
- The prediction $\hat{y}_t$ need not be calculated for the other six recursive equations to be evaluated at time t $\Rightarrow$ compute and propagate both cell c and hidden h states for time points $t-m$, $t-m+1$, $t-m+2$, up to t to then compute the prediction and add a term to the MSE loss, where m is a maximum "look-back" number of days.
- Thus the model accommodates missing values in the target variable at arbitrary times, each y_t being treated as an independent observation to predict. This also allows the model to predict at arbitrary locations (no TWD measurement required).

LSTM Data-Driven Setup

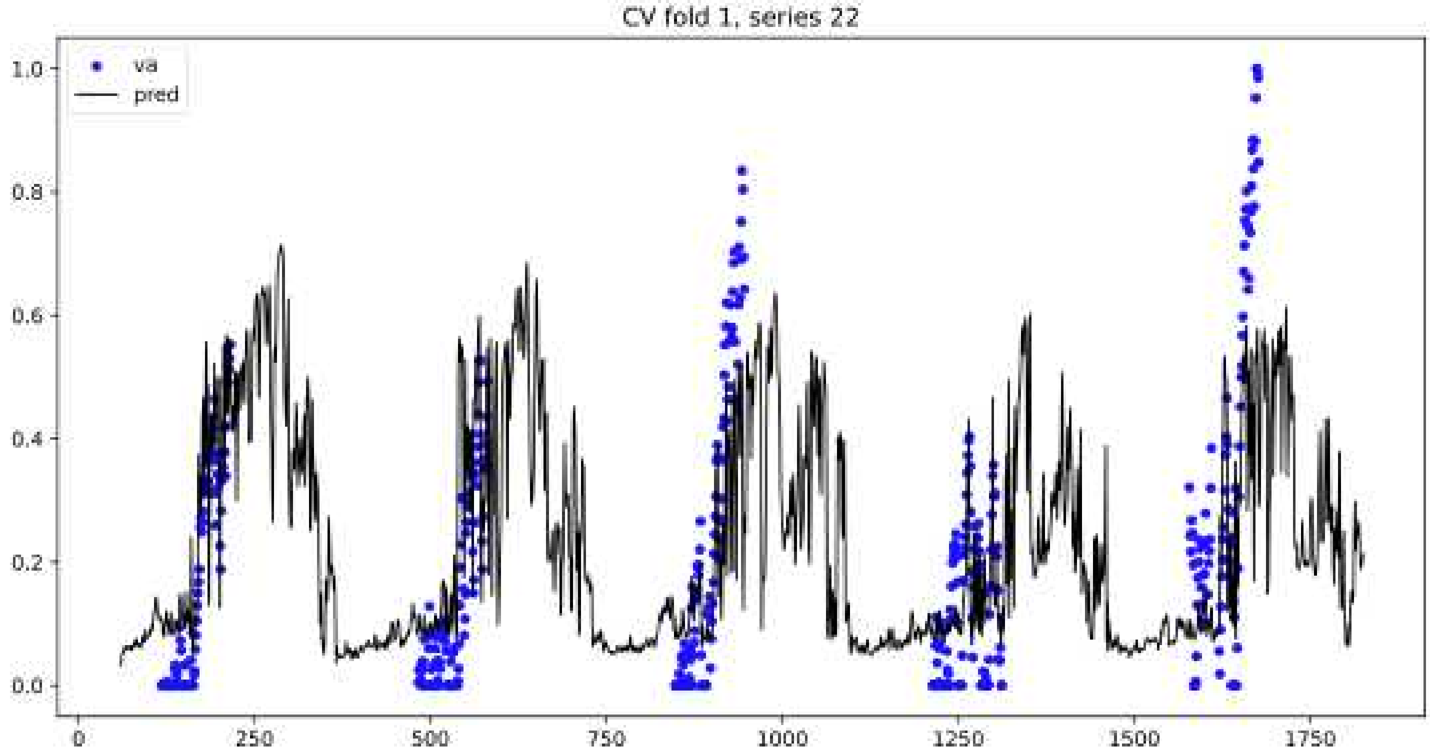
- Out-of-sample and out-of-range (spatial) prediction assessed by a leave-one-series-out cross validation (CV) scheme. Main prediction performance criterion: median prediction R^2 (among other metrics) over CV folds.
- Selected input features by (CV-based) forward stepwise variable selection: SWP, VPD, air temperature, DOY
- We set on $d = 16$ and $m = 60$. Total number of parameters = 1425
- Model fitted to the three species separately and then to all of them jointly.

LSTM fit on *P. sylvestris*

- Full fit training $R^2 = 0.2529$
- CV R^2 over $n = 9$ folds:

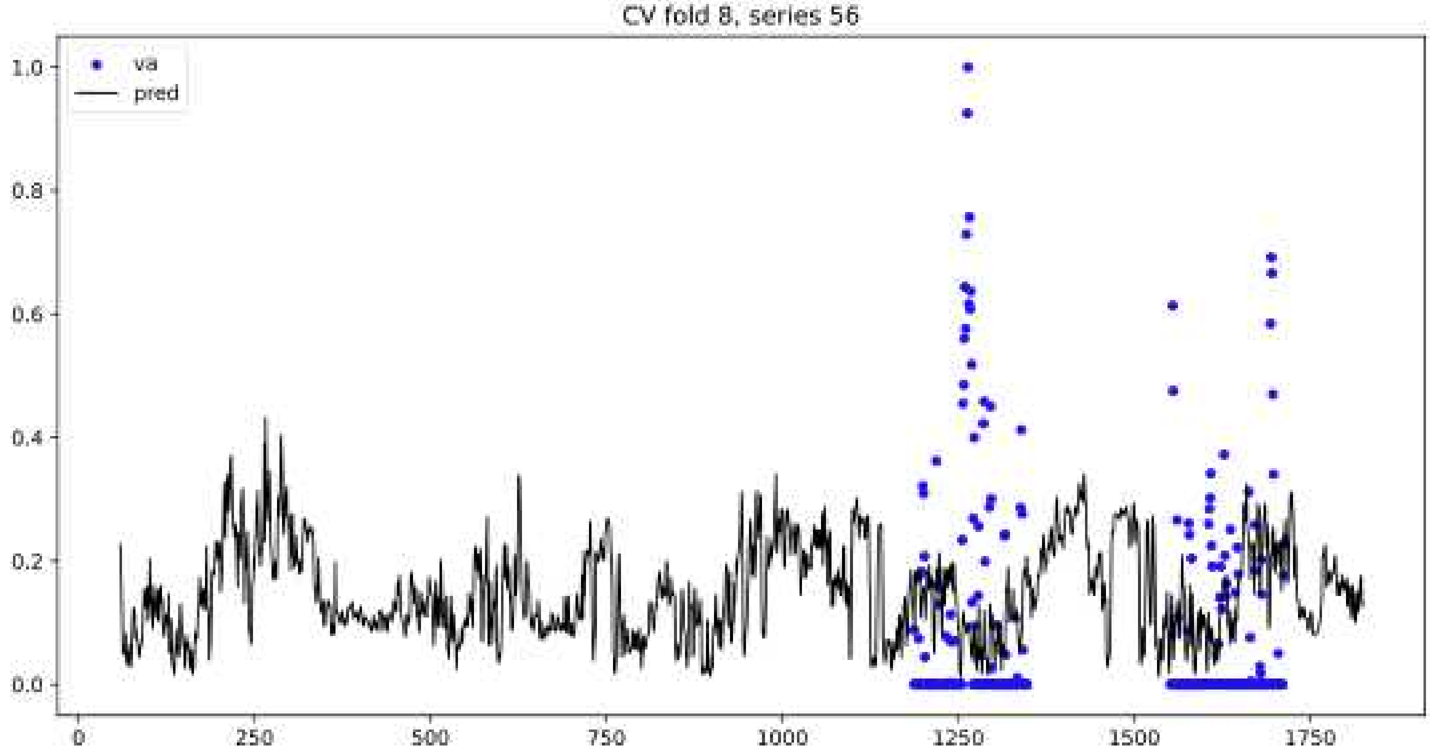
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.2506	0.0834	0.1949	0.1937	0.3901	0.5504

P. sylvestris Prediction Examples I



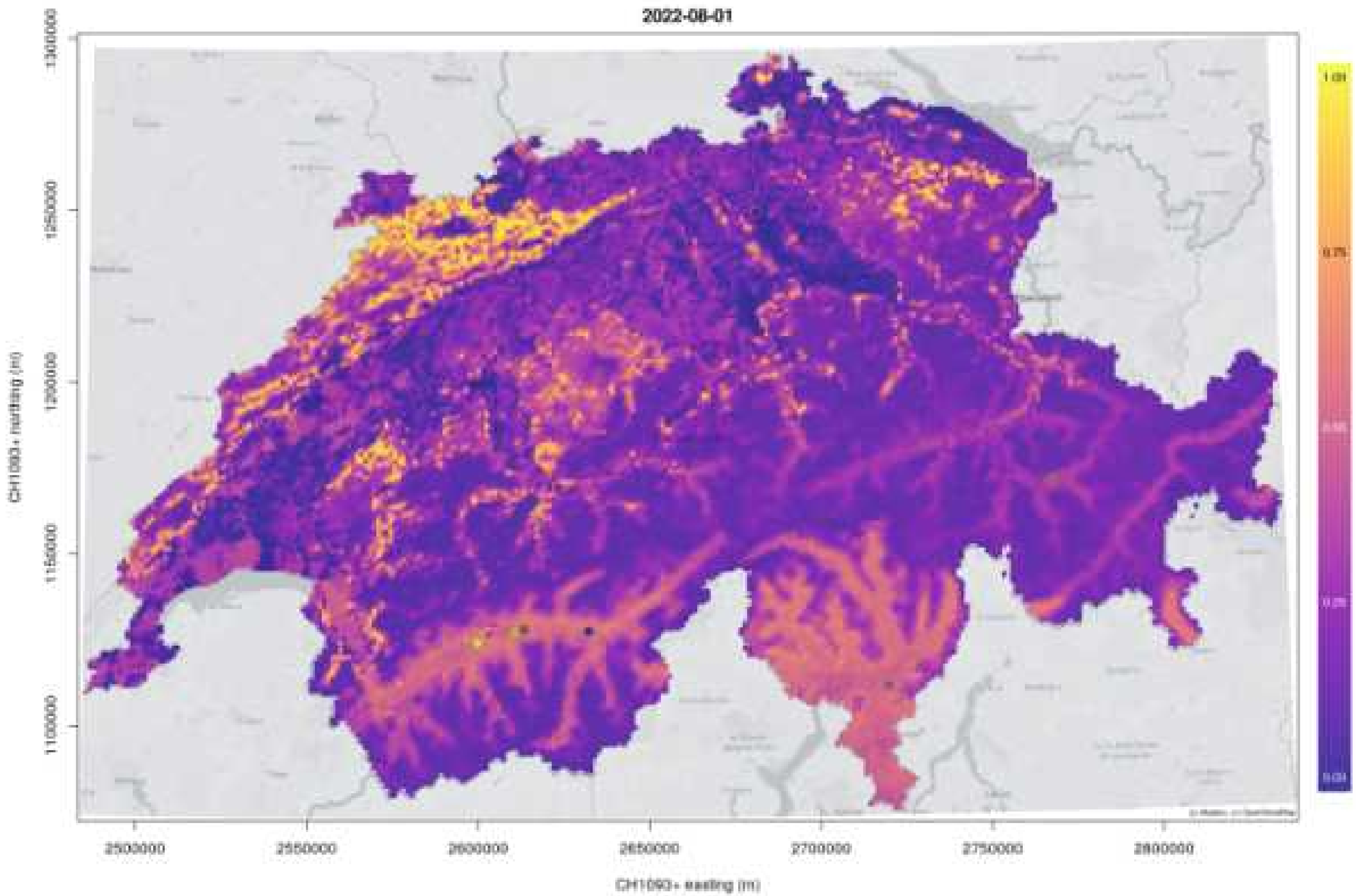
Time series and predictions for series with validation $R^2 = 0.5504$

P. sylvestris Prediction Examples II



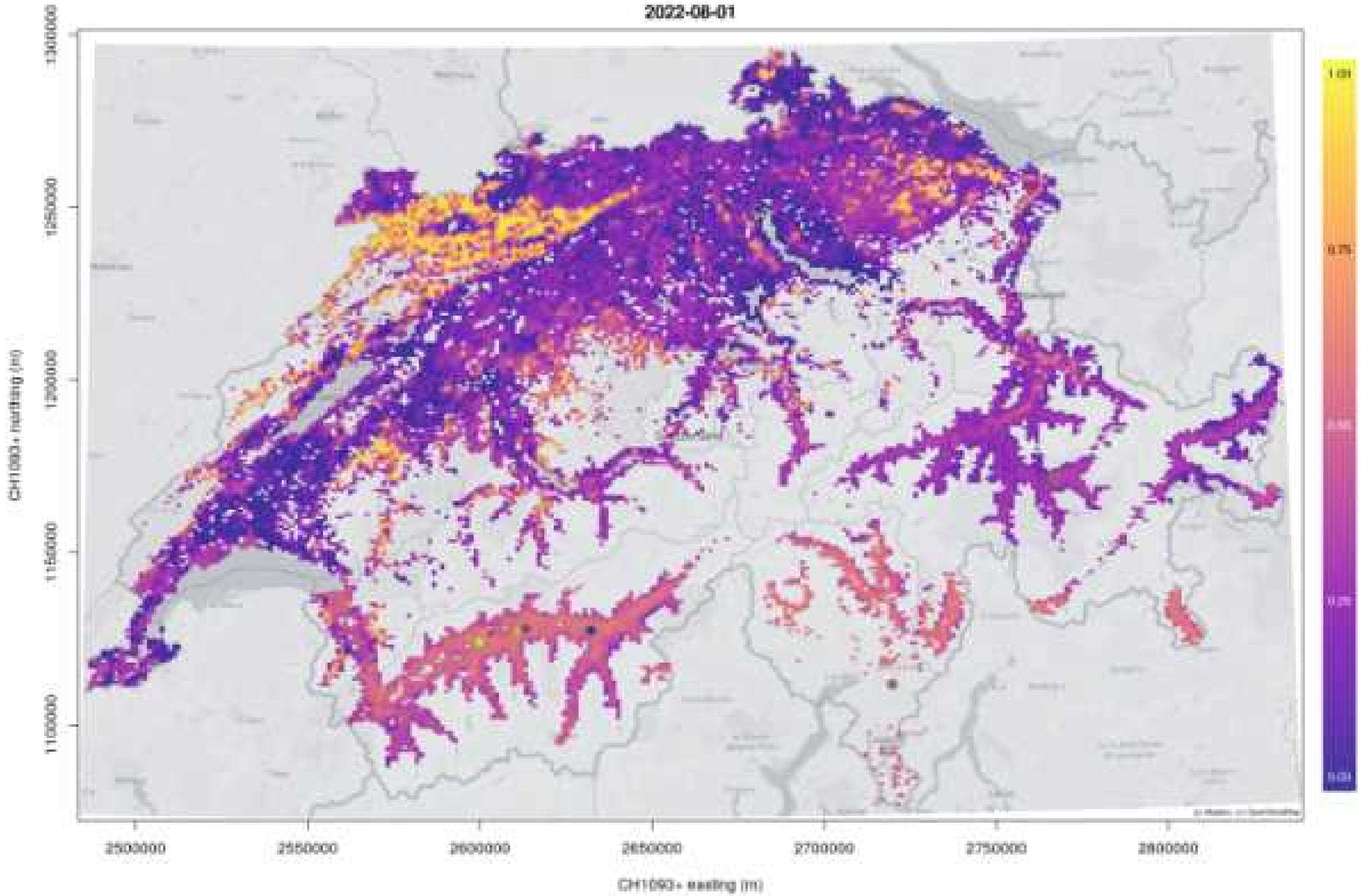
Time series and predictions for series with validation $R^2 = -0.2506$

P. sylvestris Predicted Map I



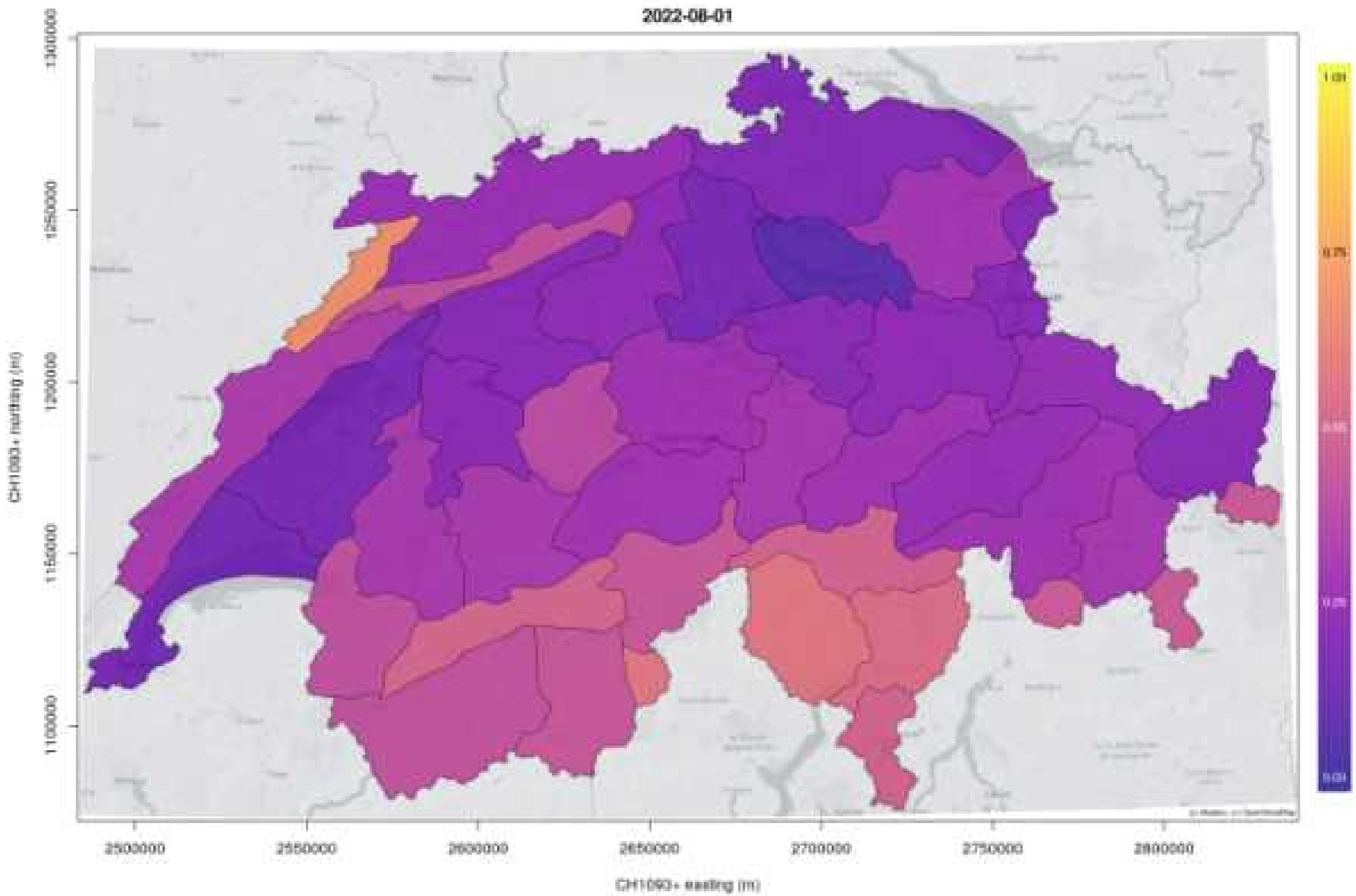
Predicted TWD index map (*P. sylvestris*) for 2022-08-01

P. sylvestris Predicted Map II



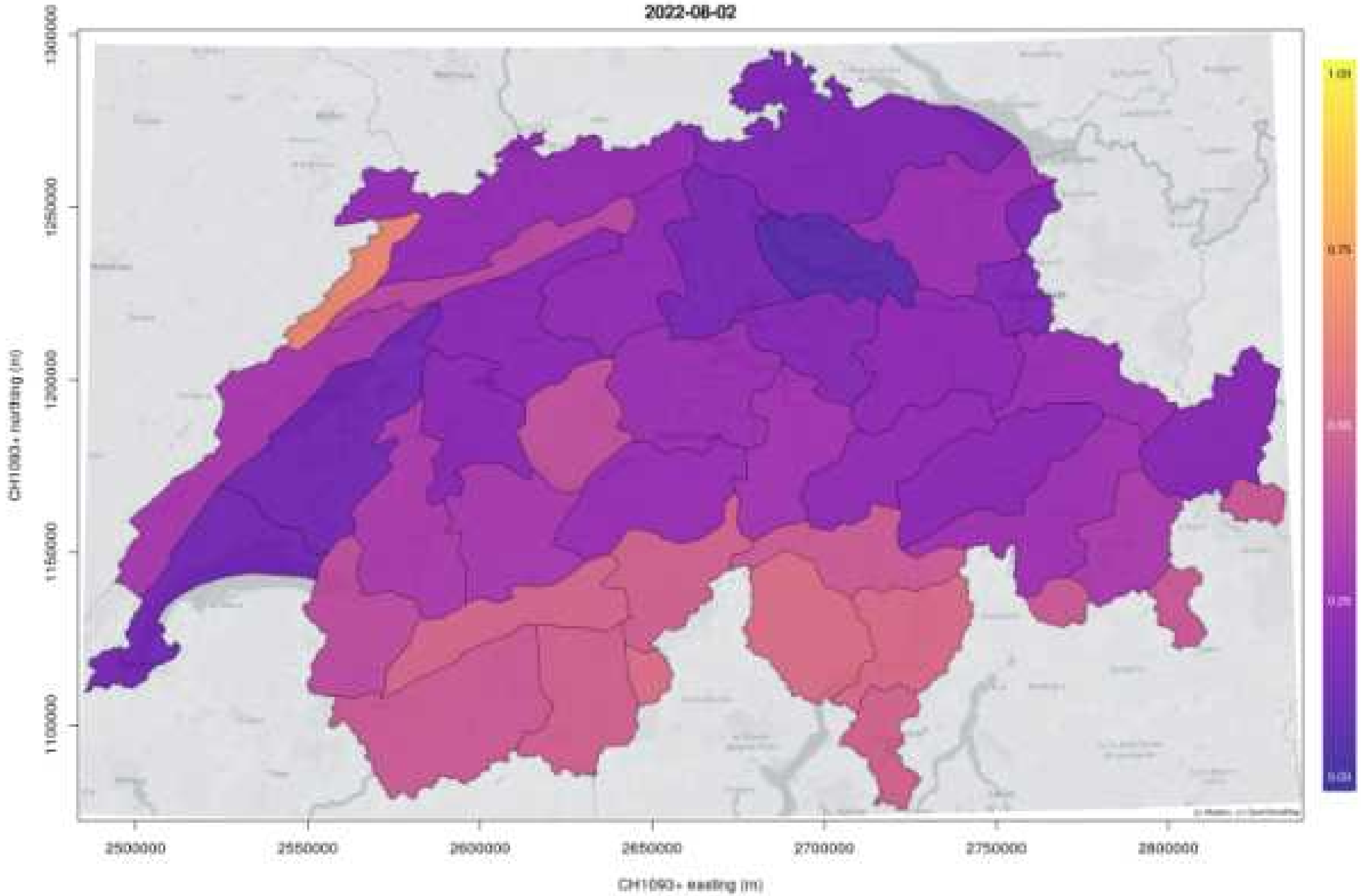
Predicted TWD index map (*P. sylvestris*) for 2022-08-01, with species presence mask

P. sylvestris Predicted Map III



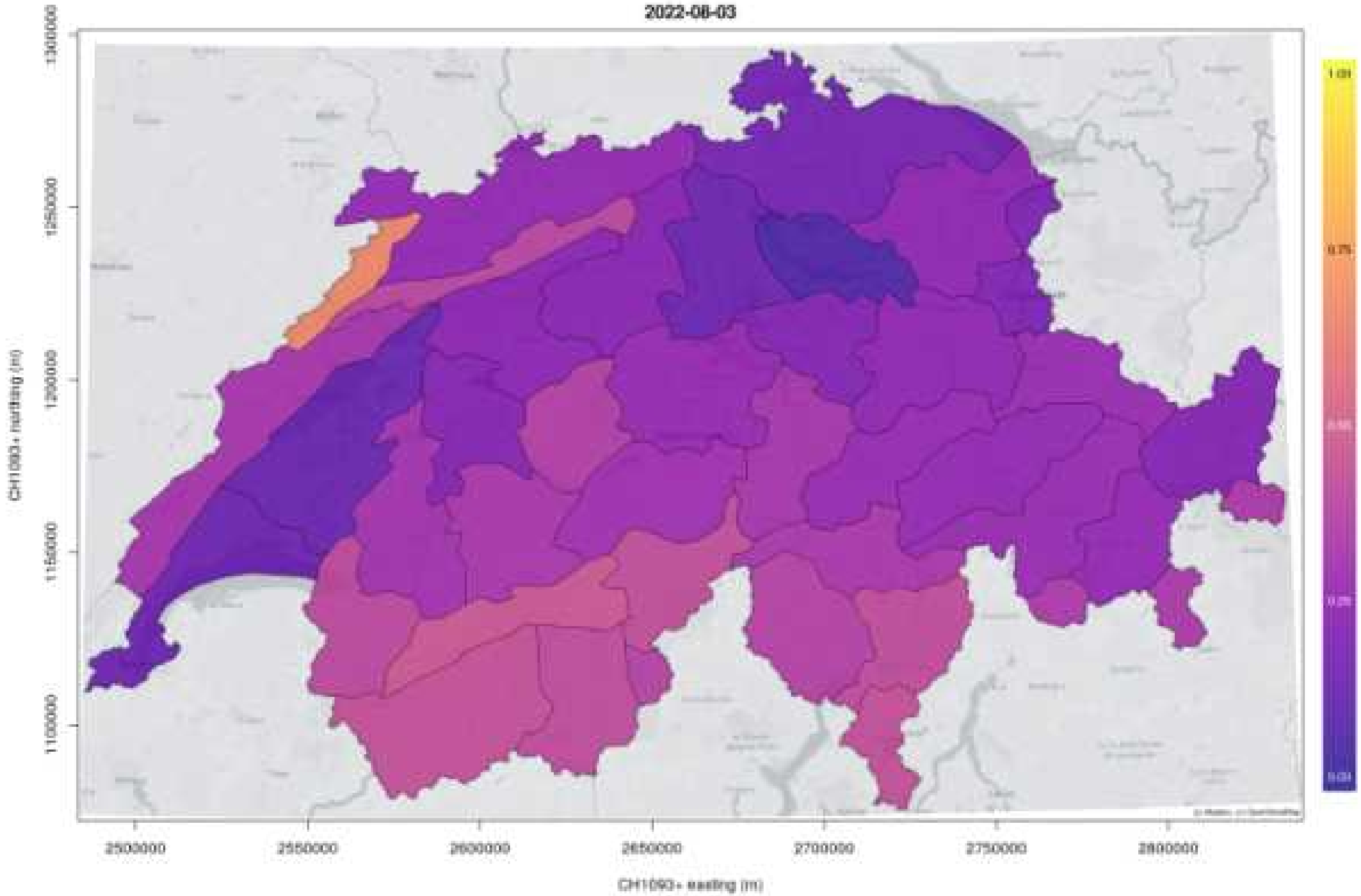
Predicted TWD index map (*P. sylvestris*) for 2022-08-01, median over BAFU "large regions"

P. sylvestris Predicted Map IV



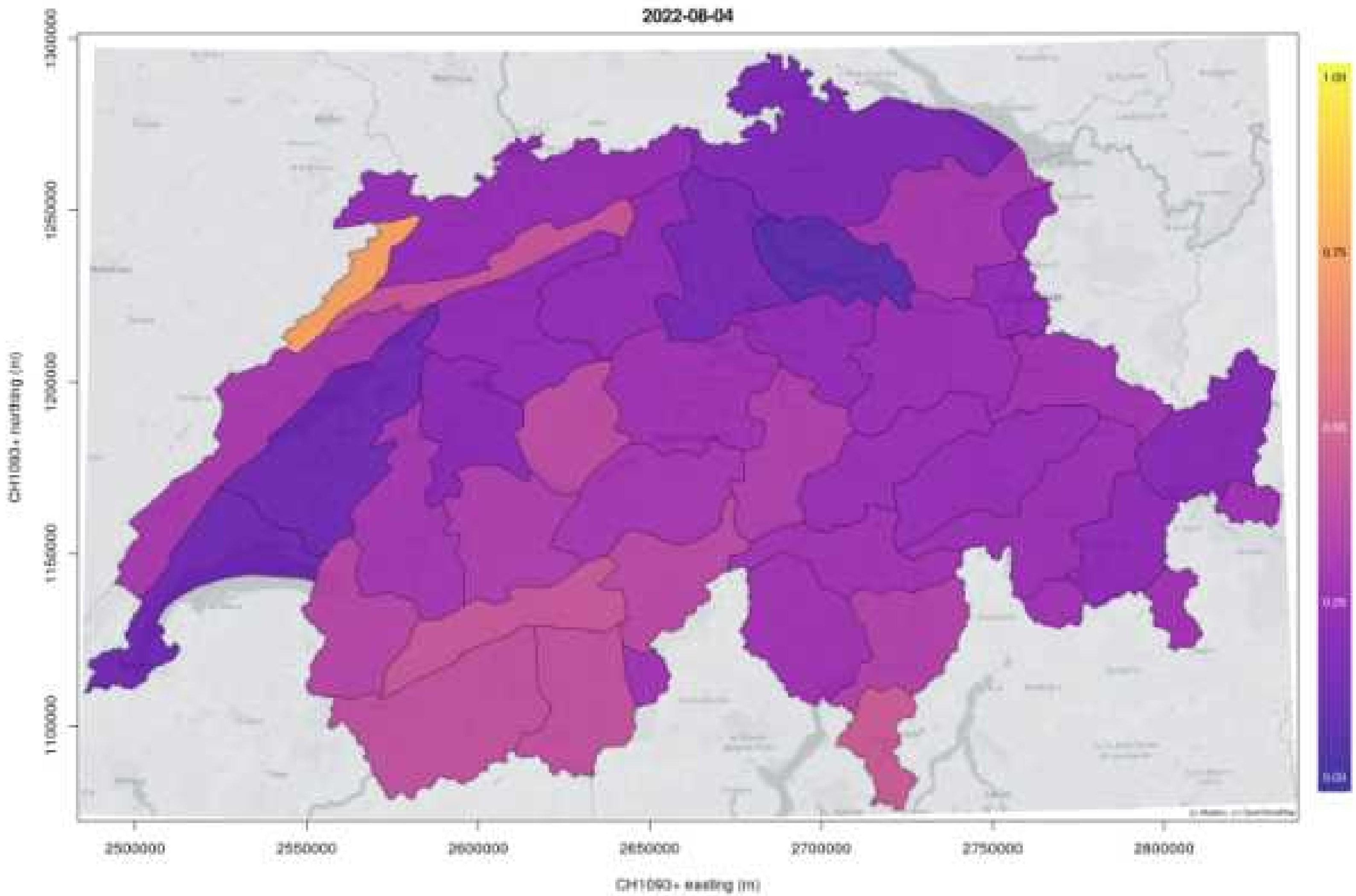
Predicted TWD index map (*P. sylvestris*) for 2022-08-02, median over BAFU large regions

P. sylvestris Predicted Map V



Predicted TWD index map (*P. sylvestris*) for 2022-08-03, median over BAFU large regions

P. sylvestris Predicted Map VI



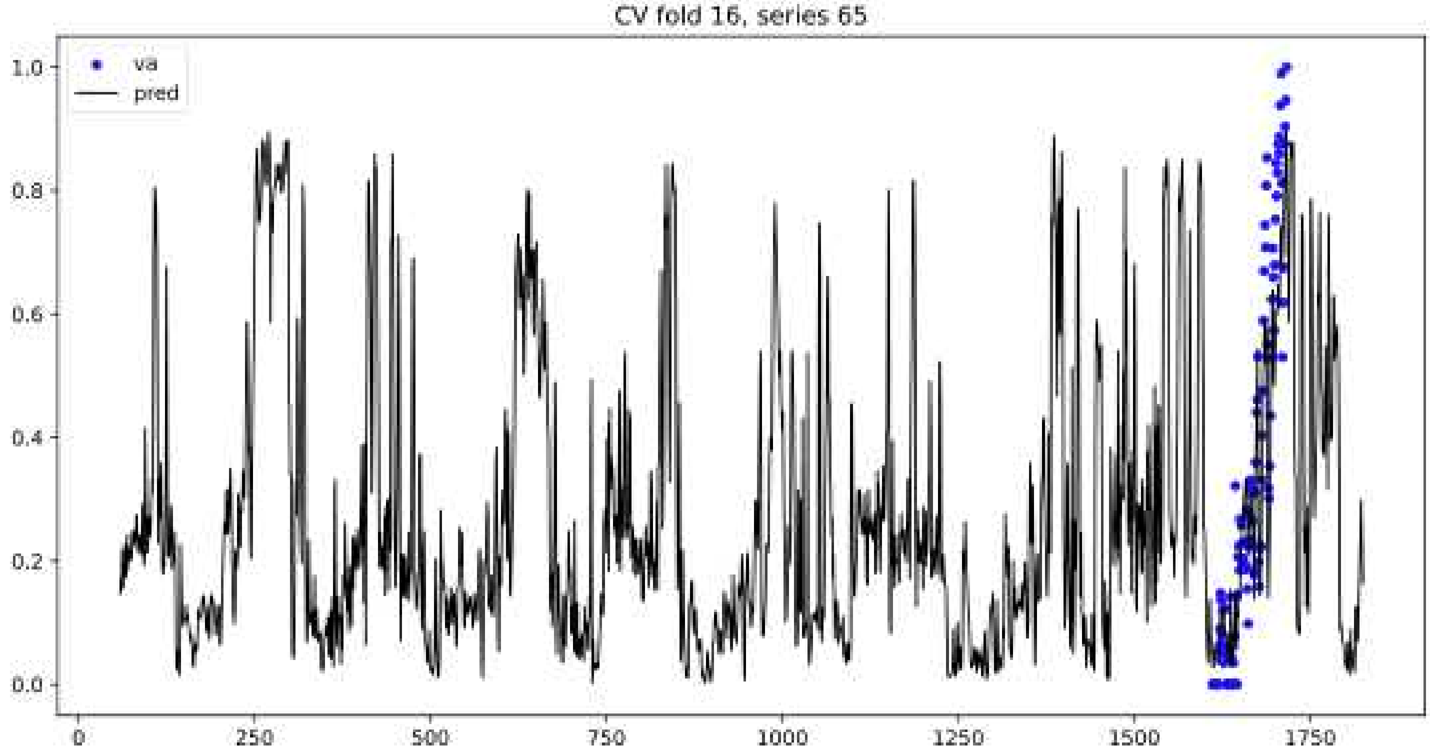
Predicted TWD index map (*P. sylvestris*) for 2022-08-04, median over BAFU large regions

LSTM fit on *P. abies*

- Full fit training $R^2 = 0.5076$
- CV R^2 over $n = 20$ folds:

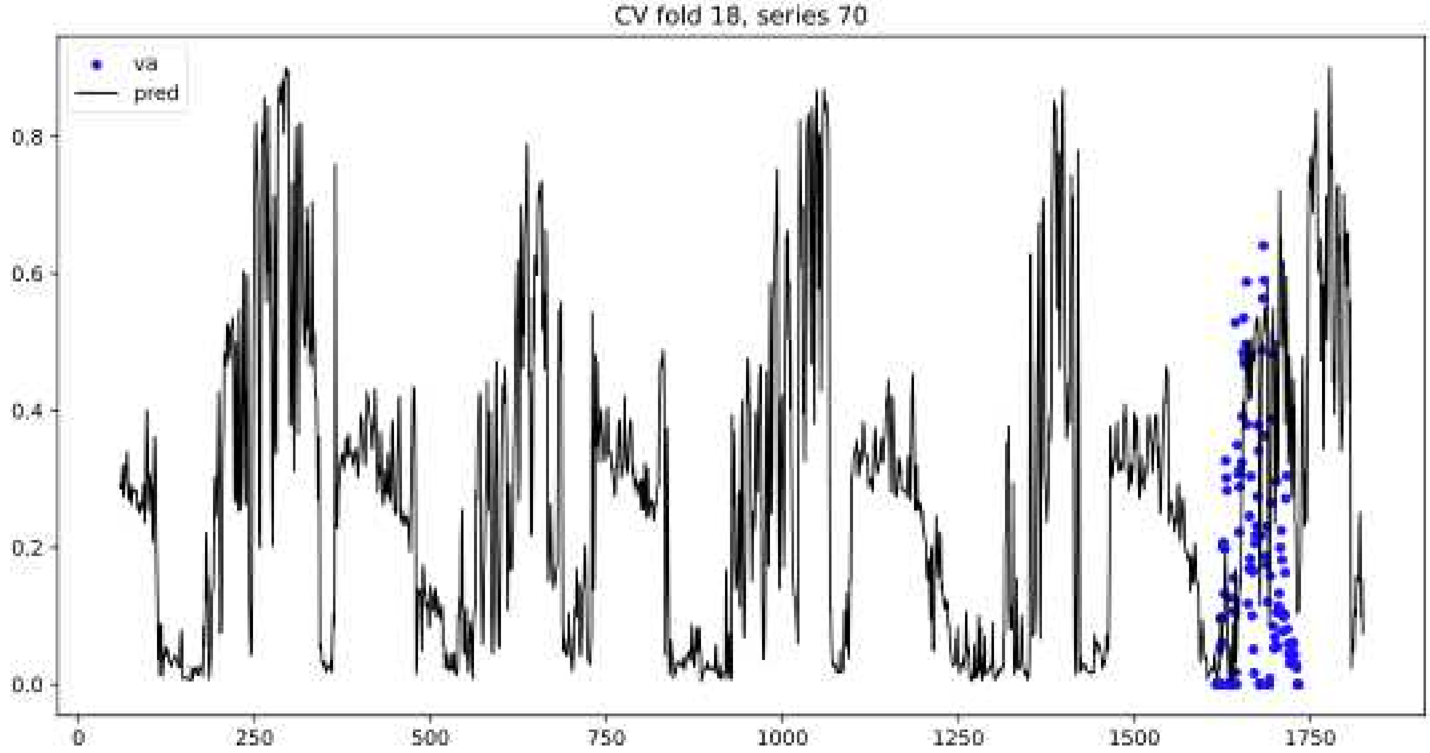
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.1679	0.1450	0.4477	0.2578	0.5505	0.8733

P. abies Prediction Examples I



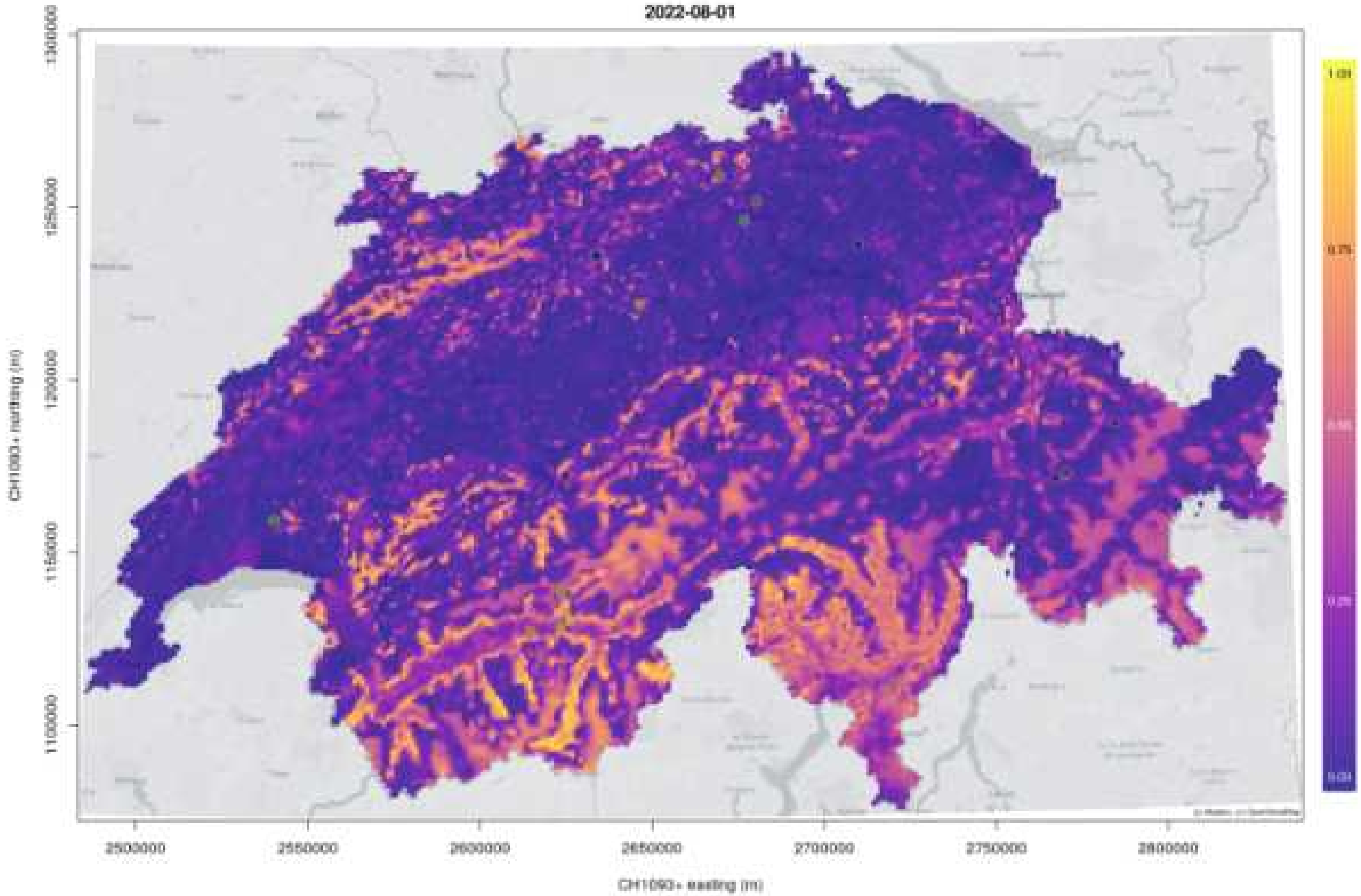
Time series and predictions for series with validation $R^2 = 0.8733$

P. abies Prediction Examples II



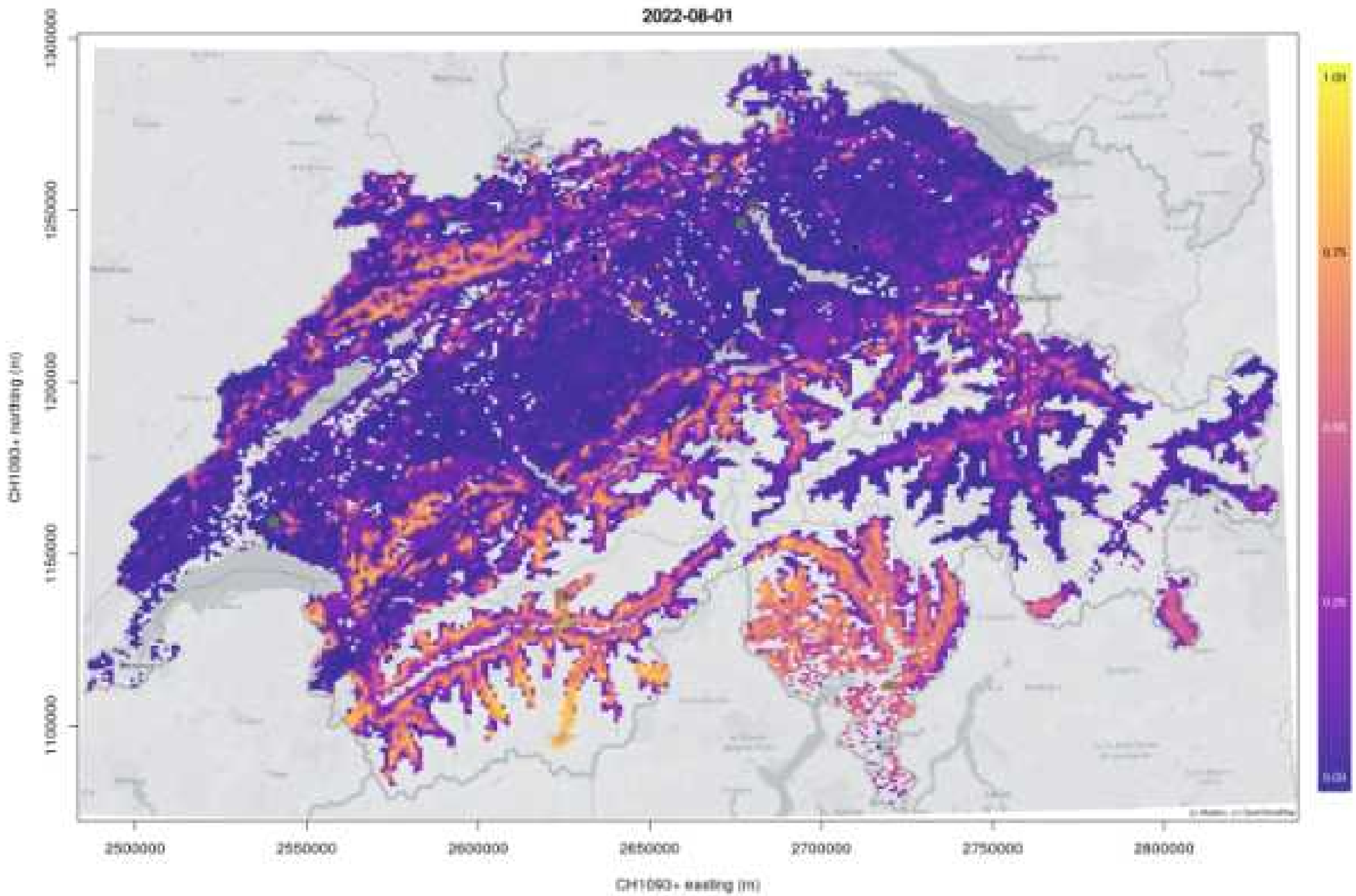
Time series and predictions for series with validation $R^2 = -1.1679$

P. abies Predicted Map I



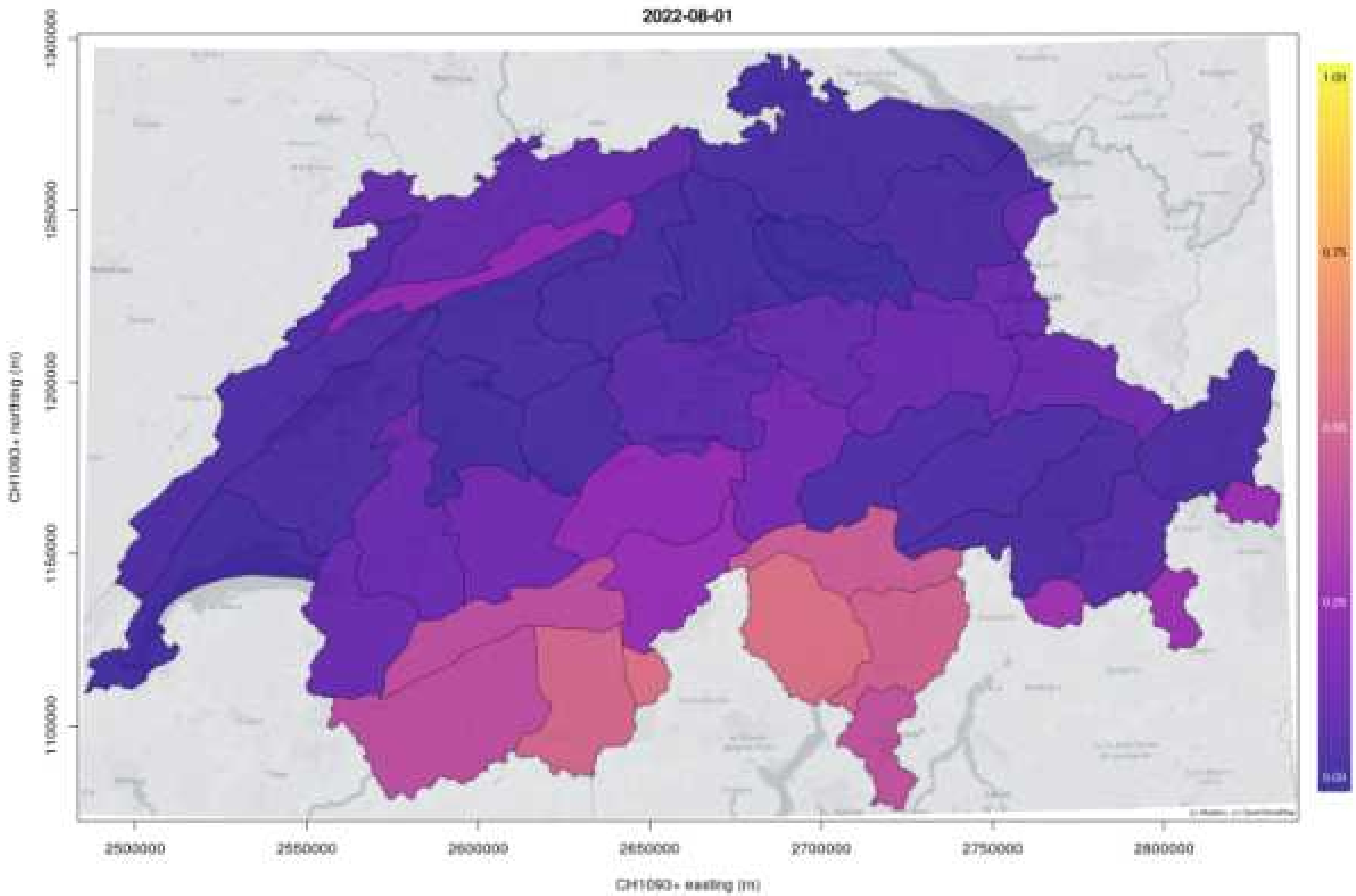
Predicted TWD index map (*P. abies*) for 2022-08-01

P. abies Predicted Map II



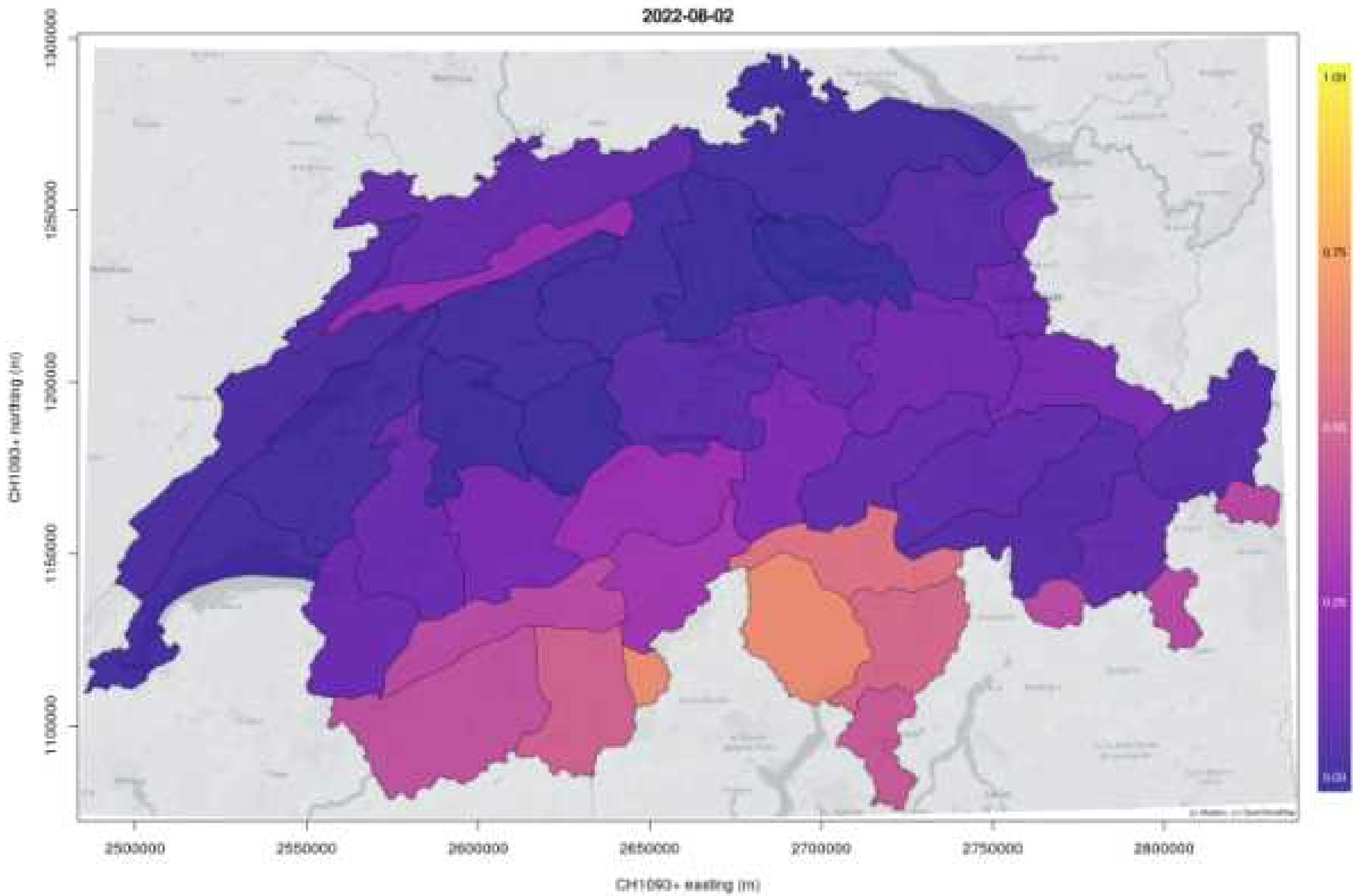
Predicted TWD index map (*P. abies*) for 2022-08-01, with species presence mask

P. abies Predicted Map III



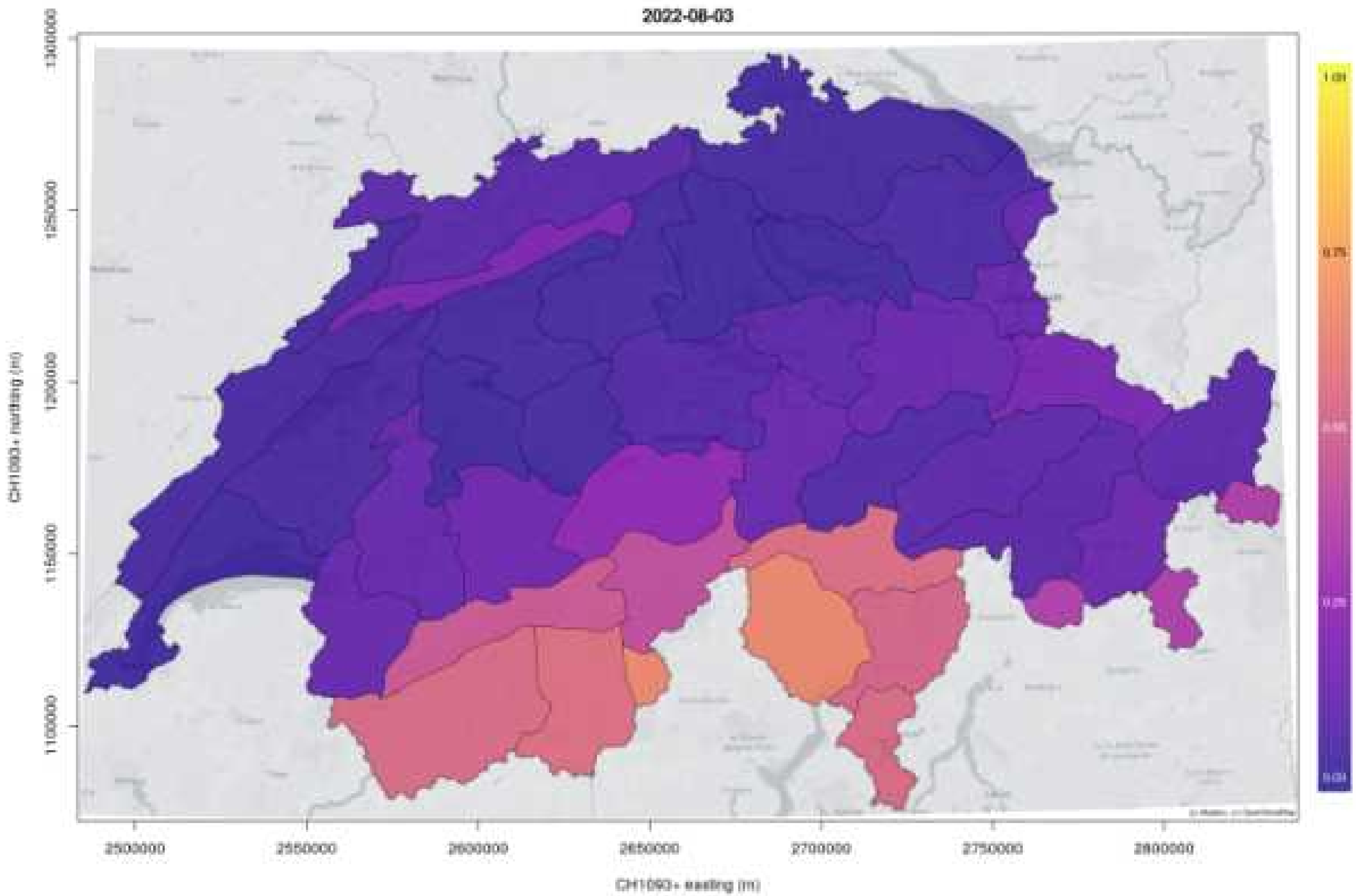
Predicted TWD index map (*P. abies*) for 2022-08-01, median over BAFU large regions

P. abies Predicted Map IV



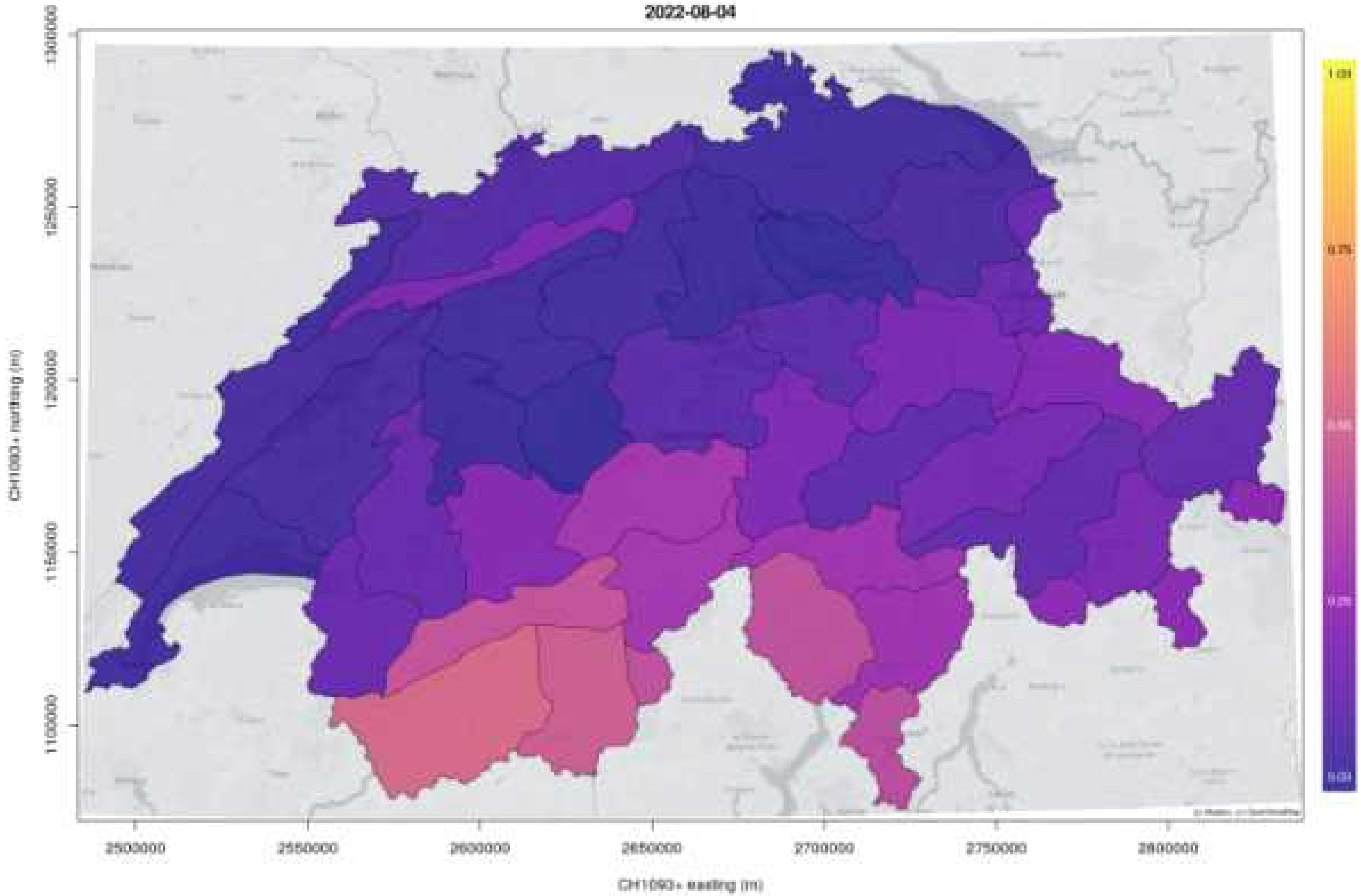
Predicted TWD index map (*P. abies*) for 2022-08-02, median over BAFU large regions

P. abies Predicted Map V



Predicted TWD index map (*P. abies*) for 2022-08-03, median over BAFU large regions

P. abies Predicted Map VI



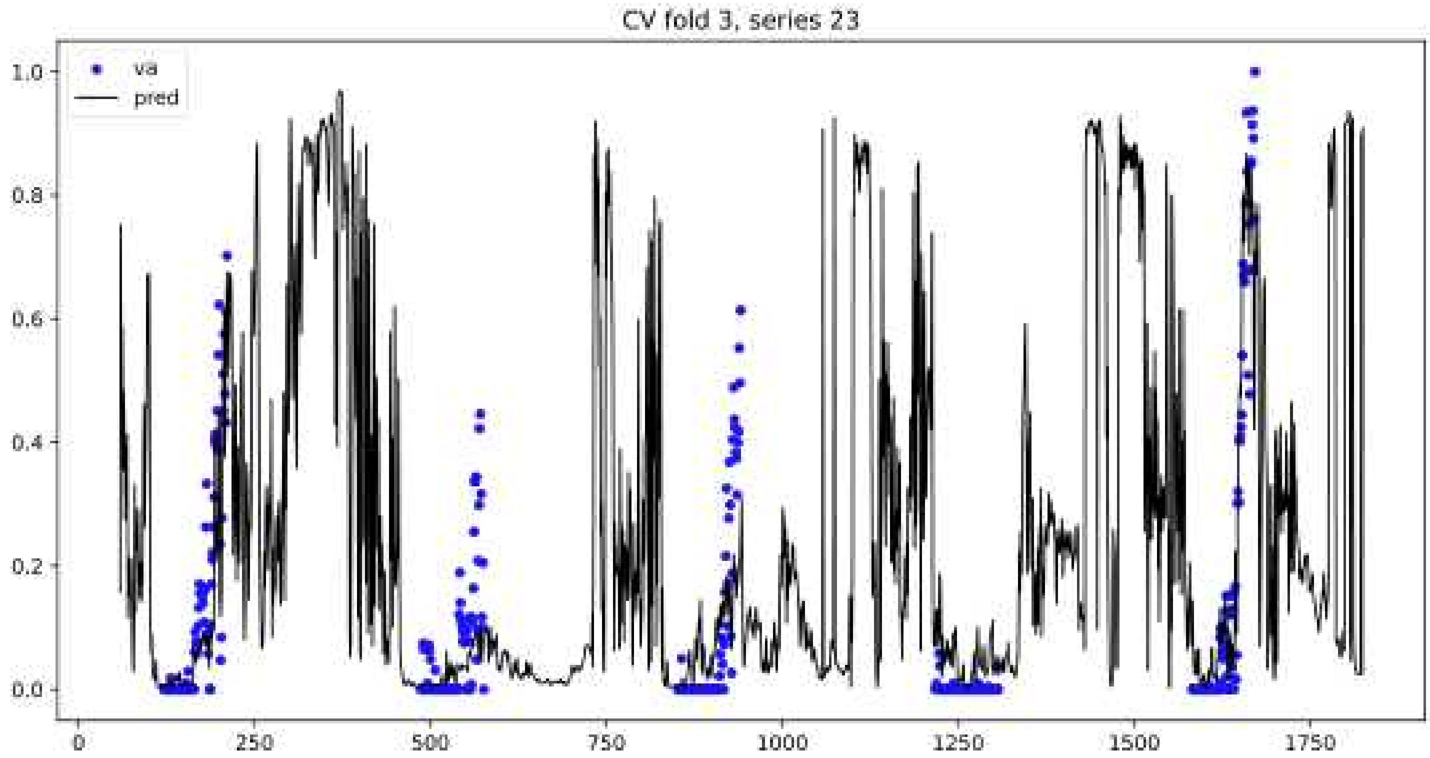
Predicted TWD index map (*P. abies*) for 2022-08-04, median over BAFU large regions

LSTM fit on *F. sylvatica*

- Full fit training $R^2 = 0.7194$
- CV R^2 over $n = 17$ folds:

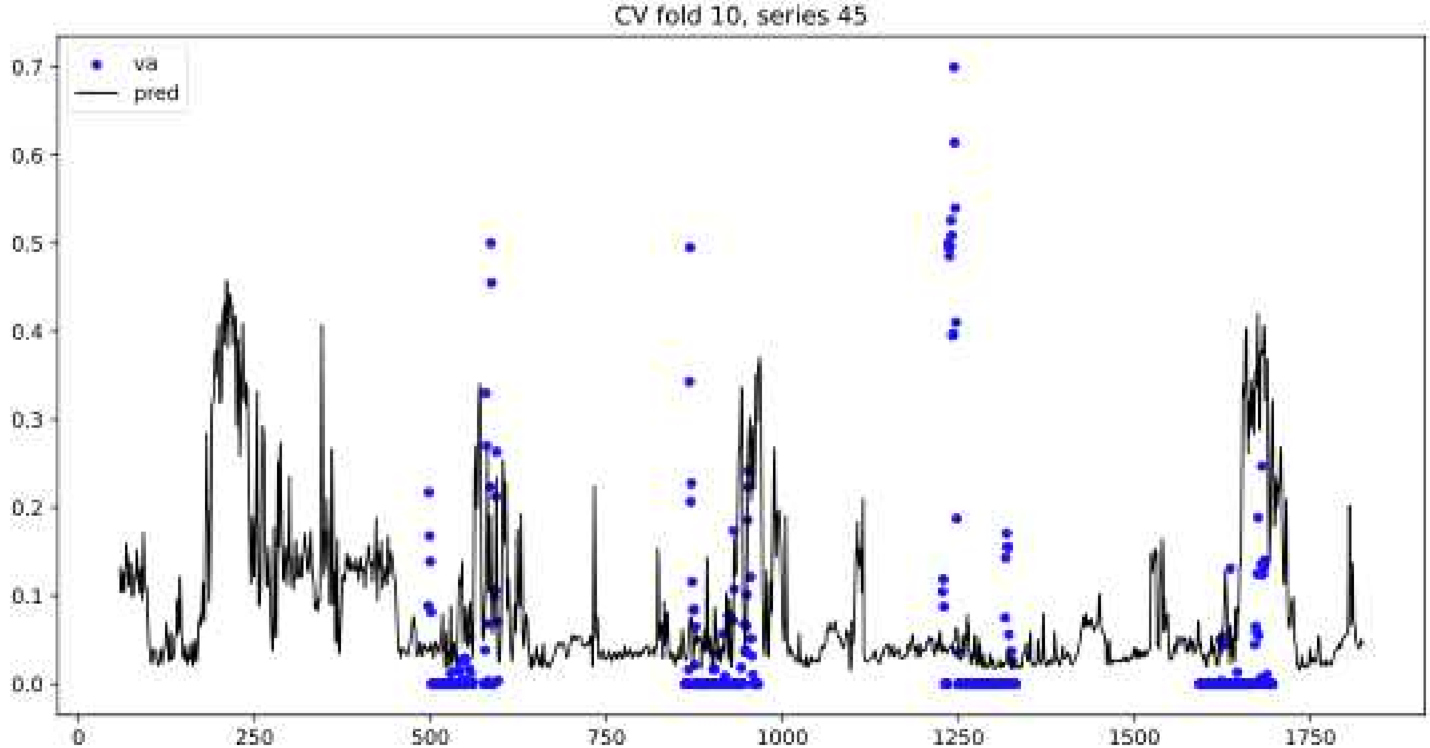
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.0828	0.0970	0.1731	0.2187	0.5926	0.7899

F. sylvatica Prediction Examples I



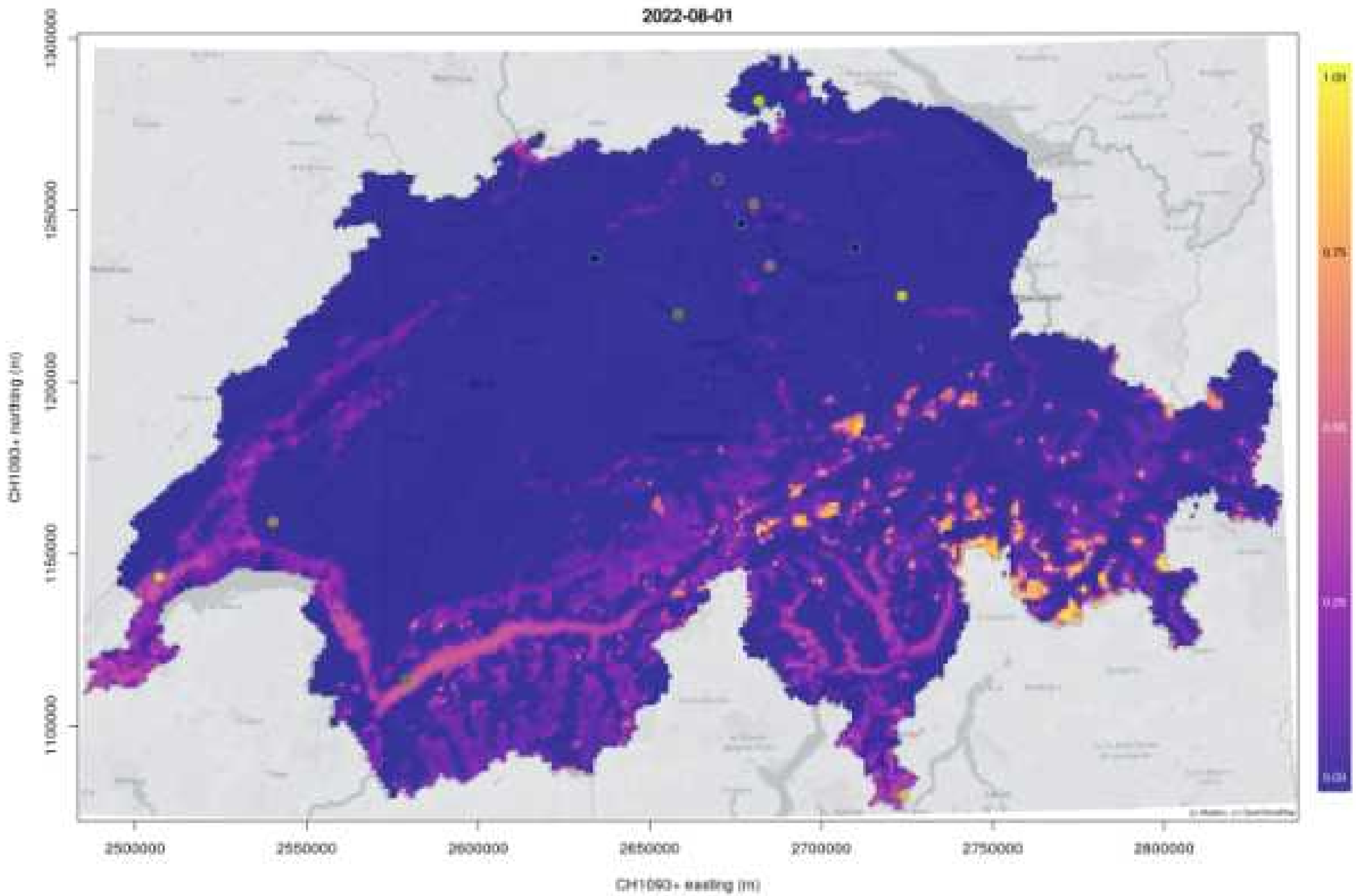
Time series and predictions for series with validation $R^2 = 0.7899$

F. sylvatica Prediction Examples II



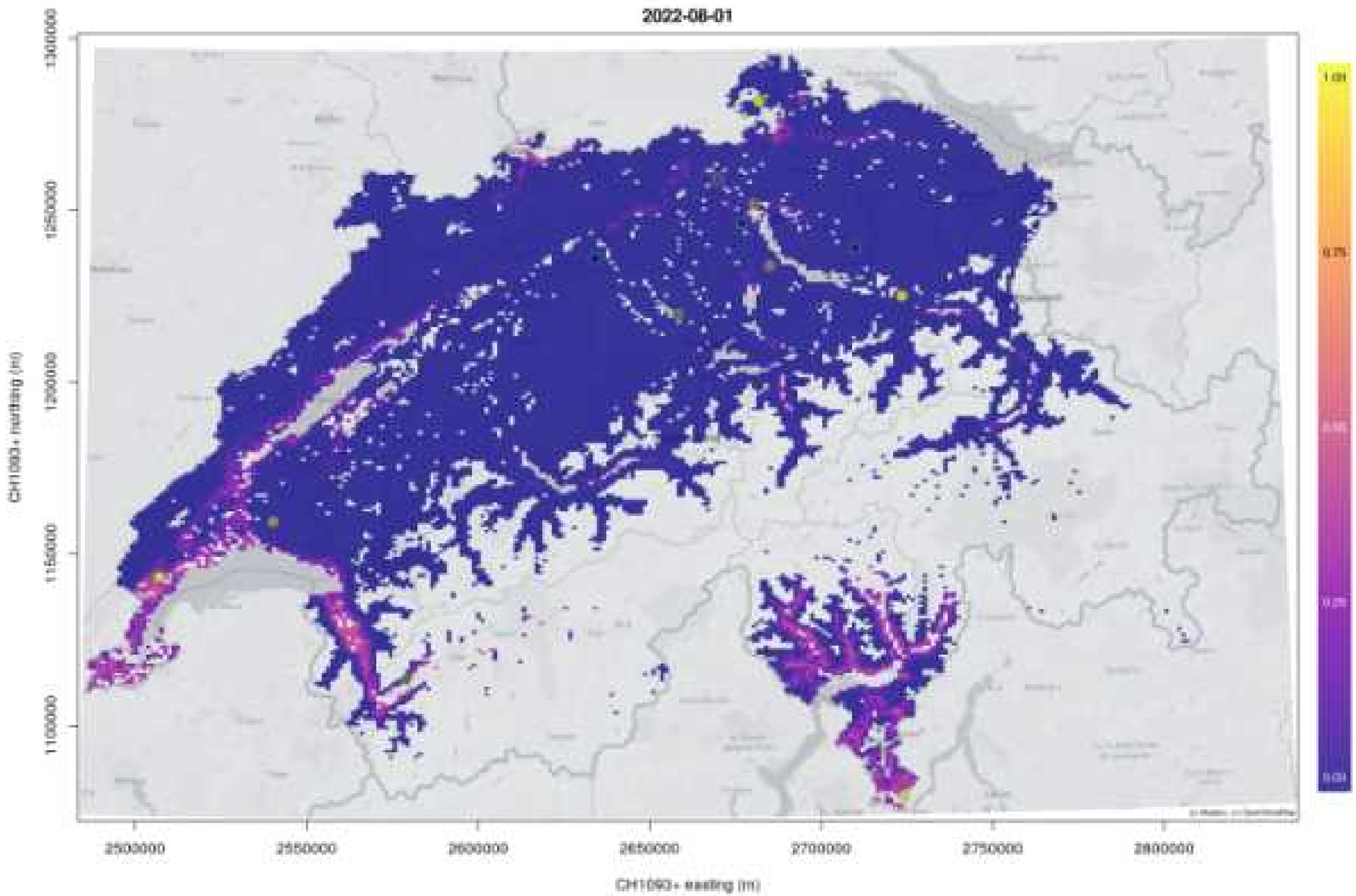
Time series and predictions for series with validation $R^2 = -1.0828$

F. sylvatica Predicted Map I



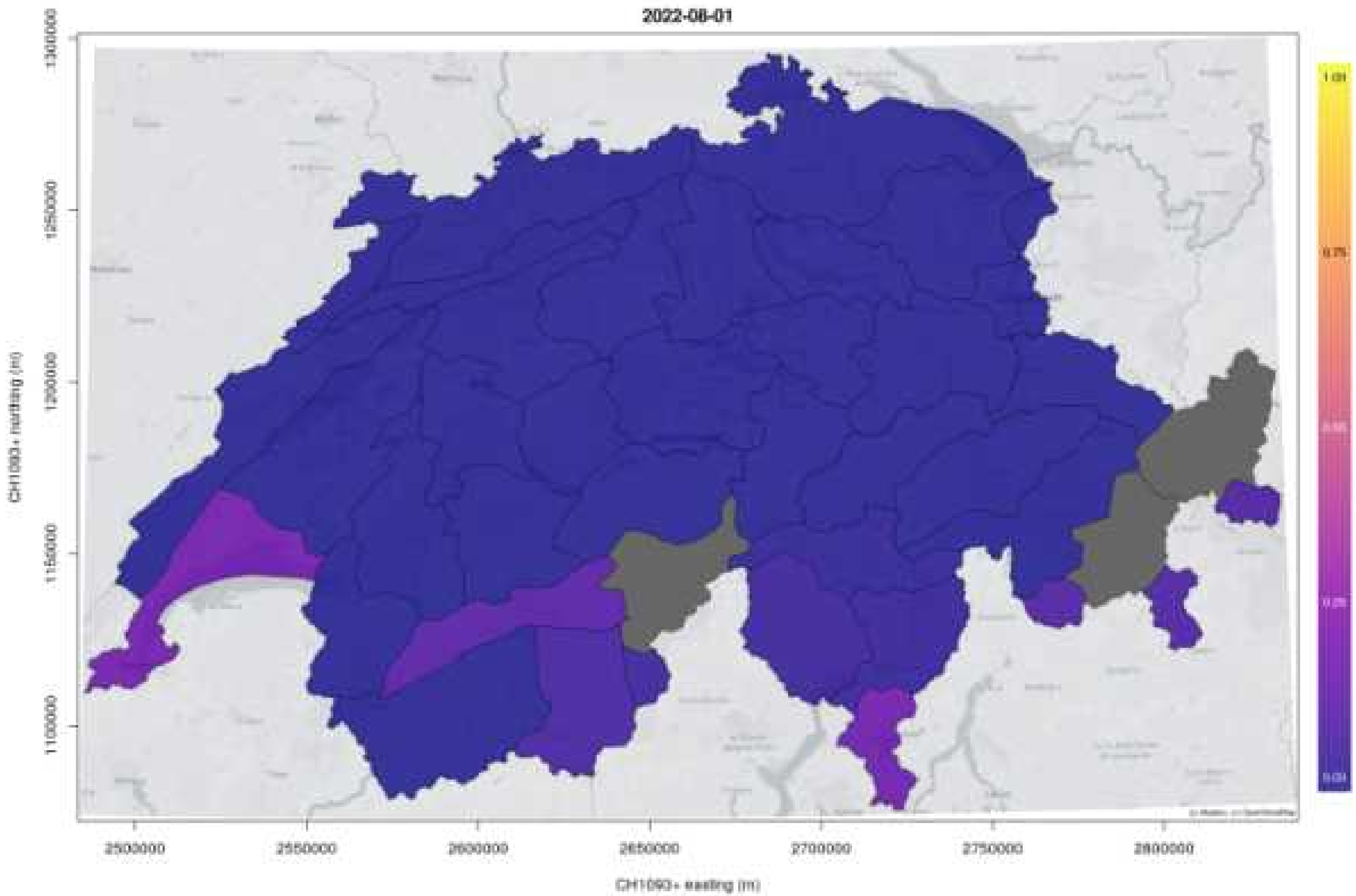
Predicted TWD index map (*F. sylvatica*) for 2022-08-01

F. sylvatica Predicted Map II



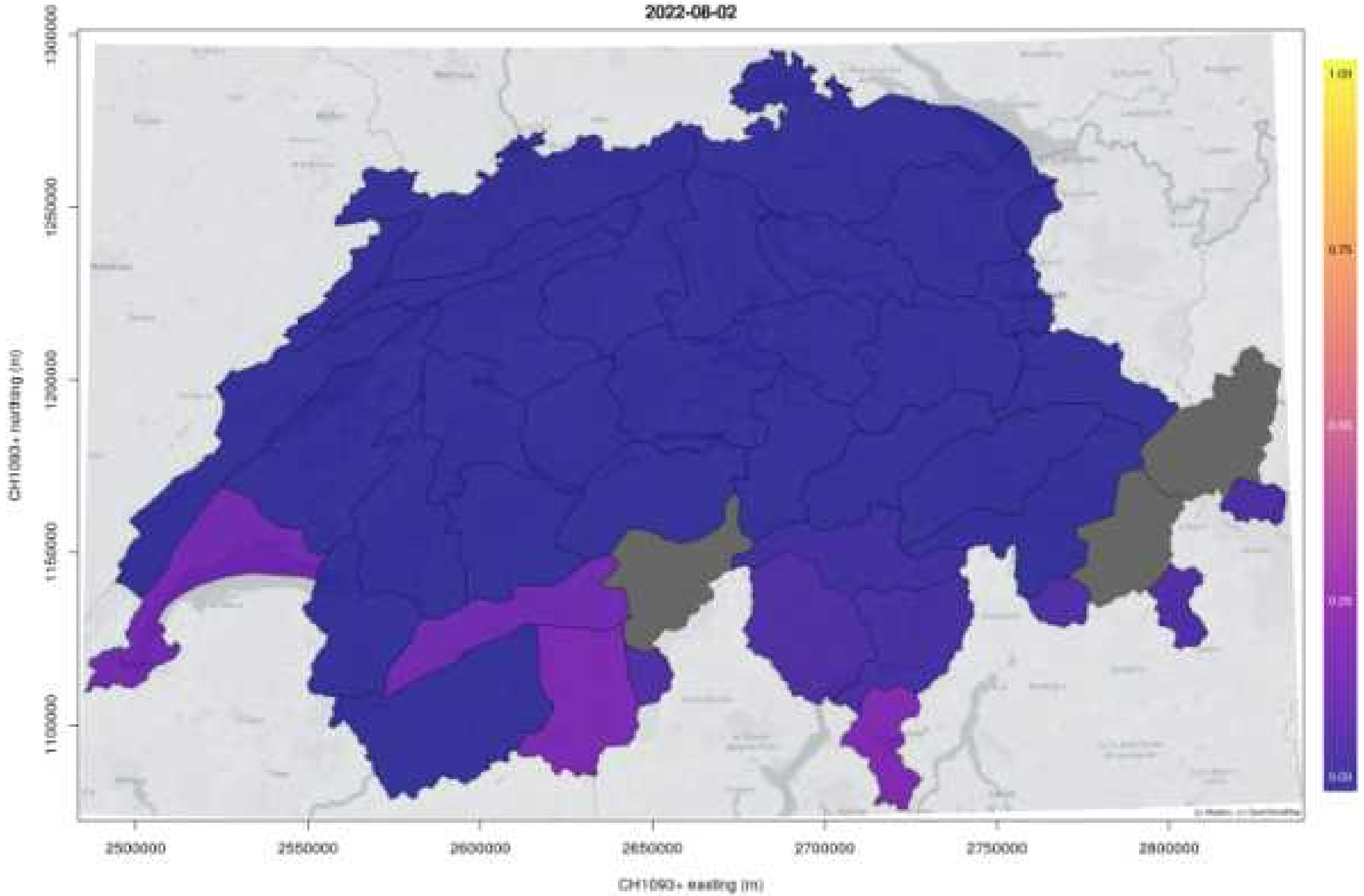
Predicted TWD index map (*F. sylvatica*) for 2022-08-01, with species presence mask

F. sylvatica Predicted Map III



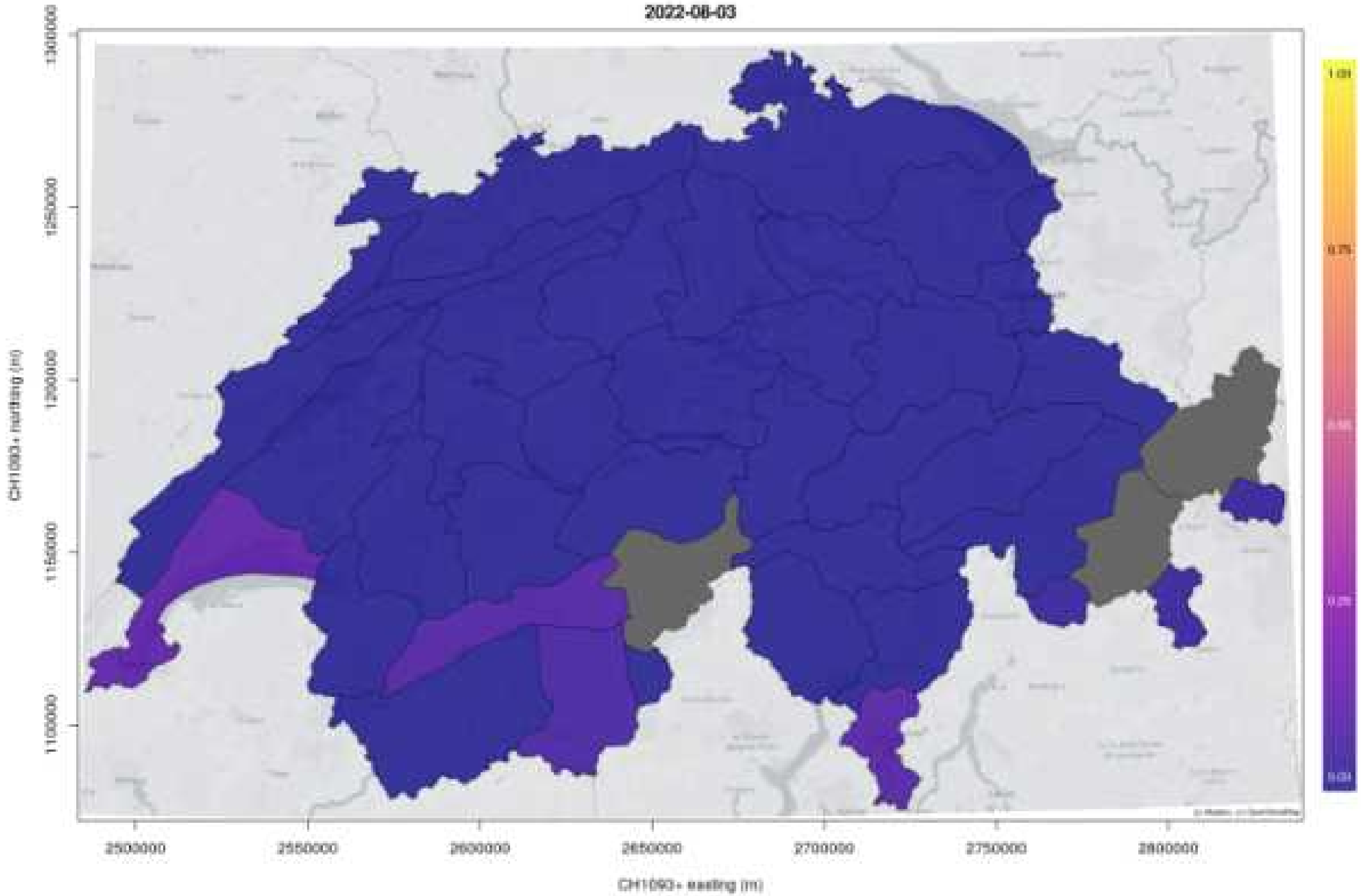
Predicted TWD index map (*F. sylvatica*) for 2022-08-01, median over BAFU large regions

F. sylvatica Predicted Map IV



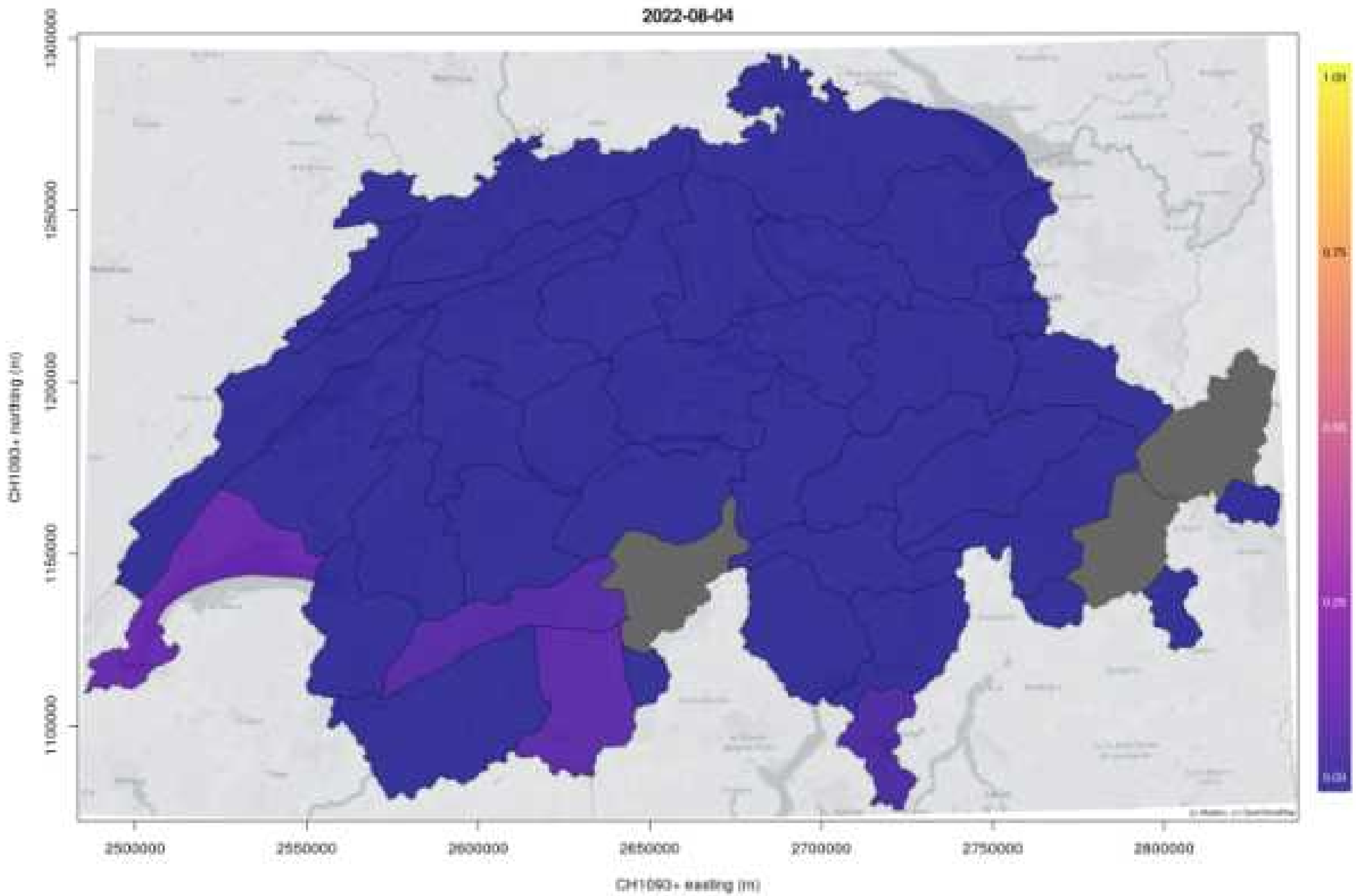
Predicted TWD index map (*F. sylvatica*) for 2022-08-02, median over BAFU large regions

F. sylvatica Predicted Map V



Predicted TWD index map (*F. sylvatica*) for 2022-08-03, median over BAFU large regions

F. sylvatica Predicted Map VI



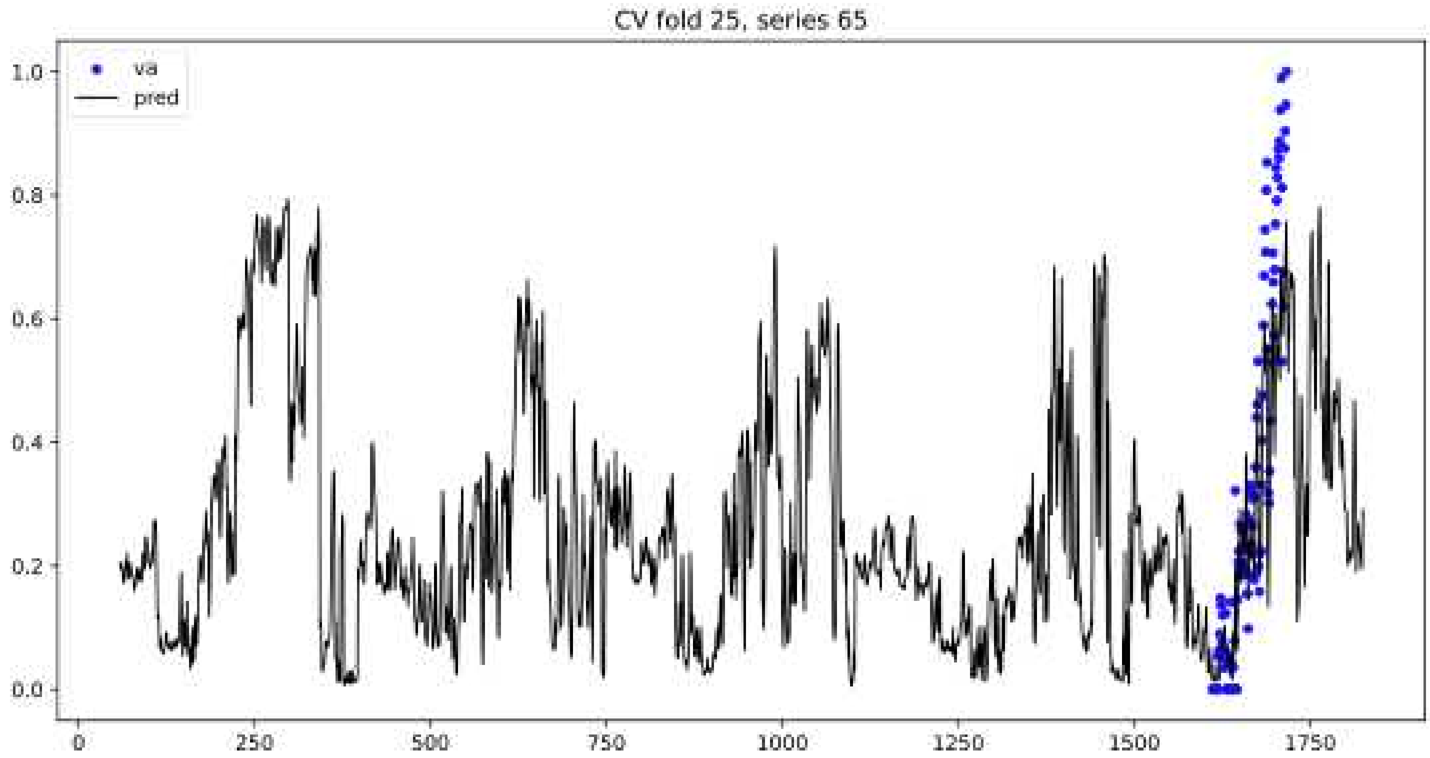
Predicted TWD index map (*F. sylvatica*) for 2022-08-04, median over BAFU large regions

LSTM Joint Fit

- Full fit training $R^2 = 0.4147$
- CV R^2 over $n = 46$ folds:

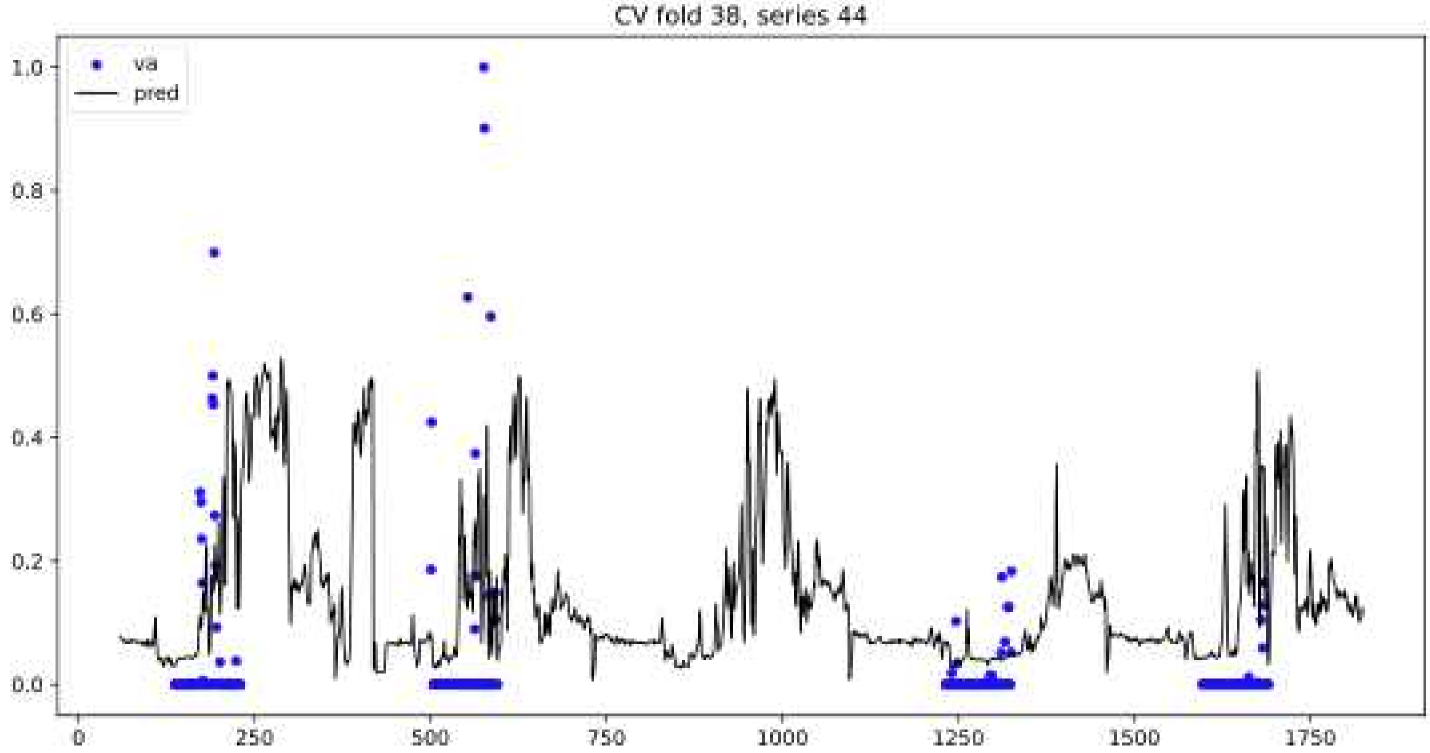
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.61520	0.04005	0.28900	0.22244	0.53185	0.82150

Joint Fit Prediction Examples I



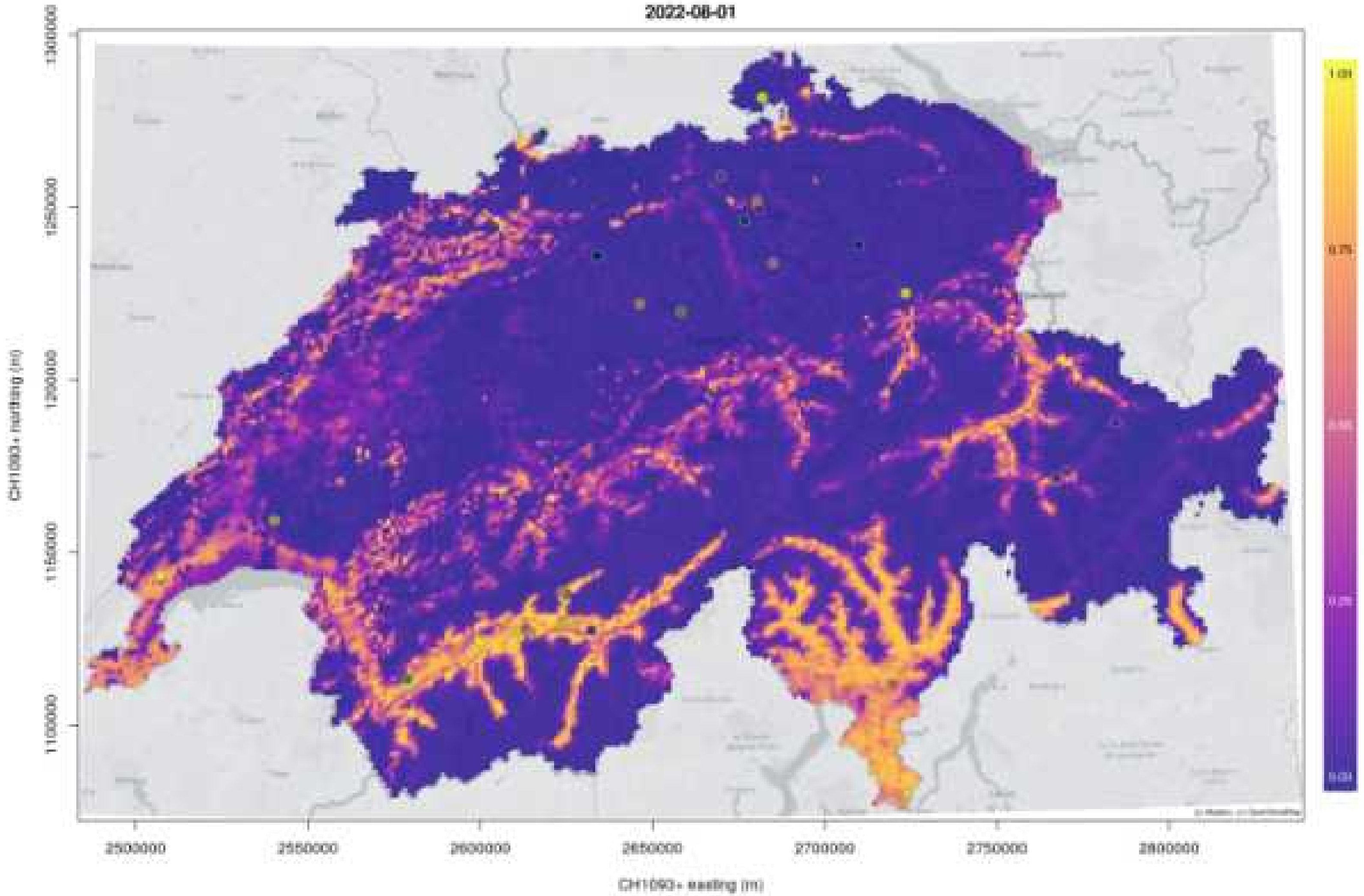
Time series and predictions for series with validation $R^2 = 0.82150$

Joint Fit Prediction Examples II



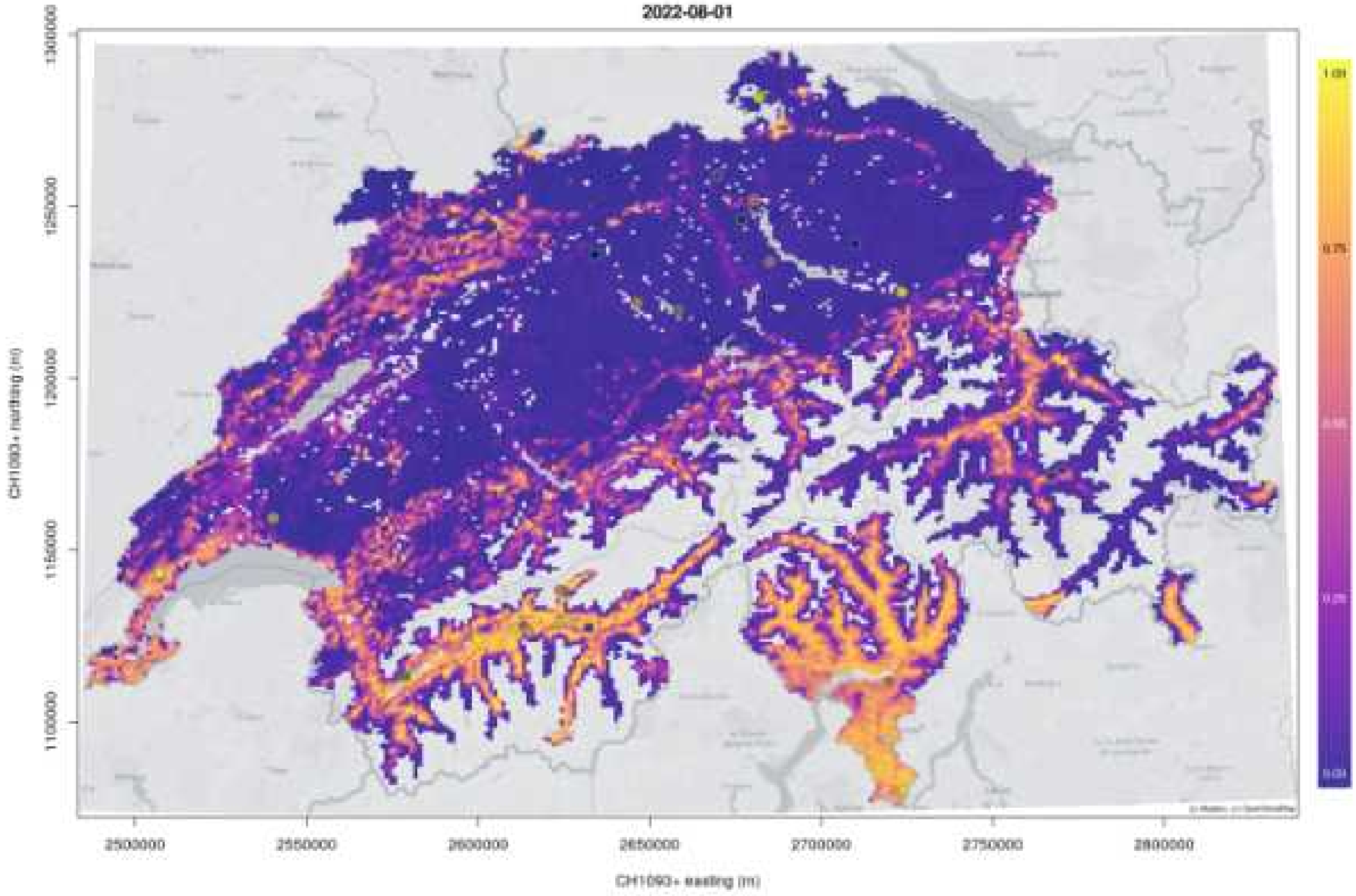
Time series and predictions for series with validation $R^2 = -1.61520$

Joint Fit Predicted Map I



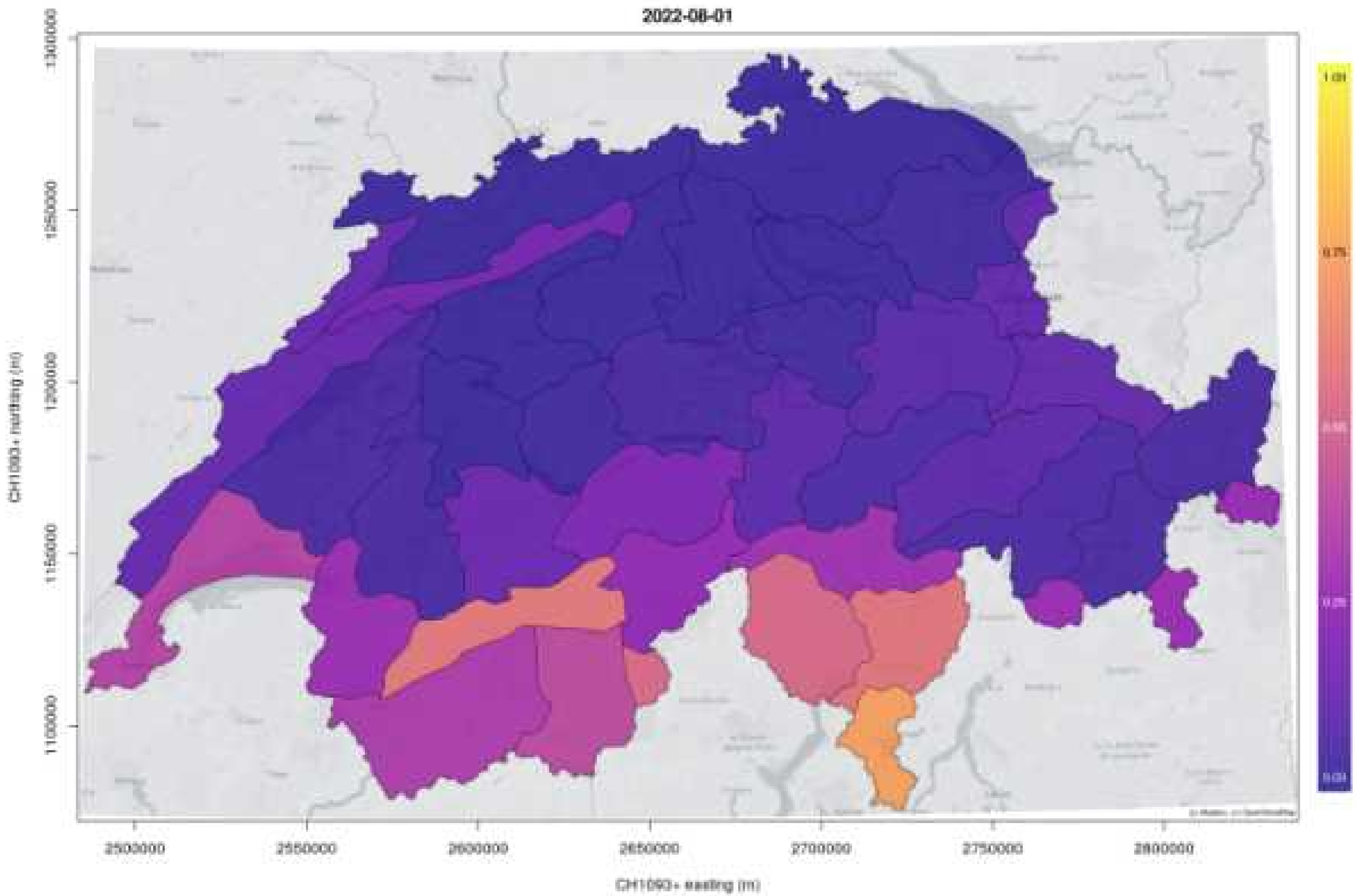
Predicted TWD index map (joint fit) for 2022-08-01

Joint Fit Predicted Map II



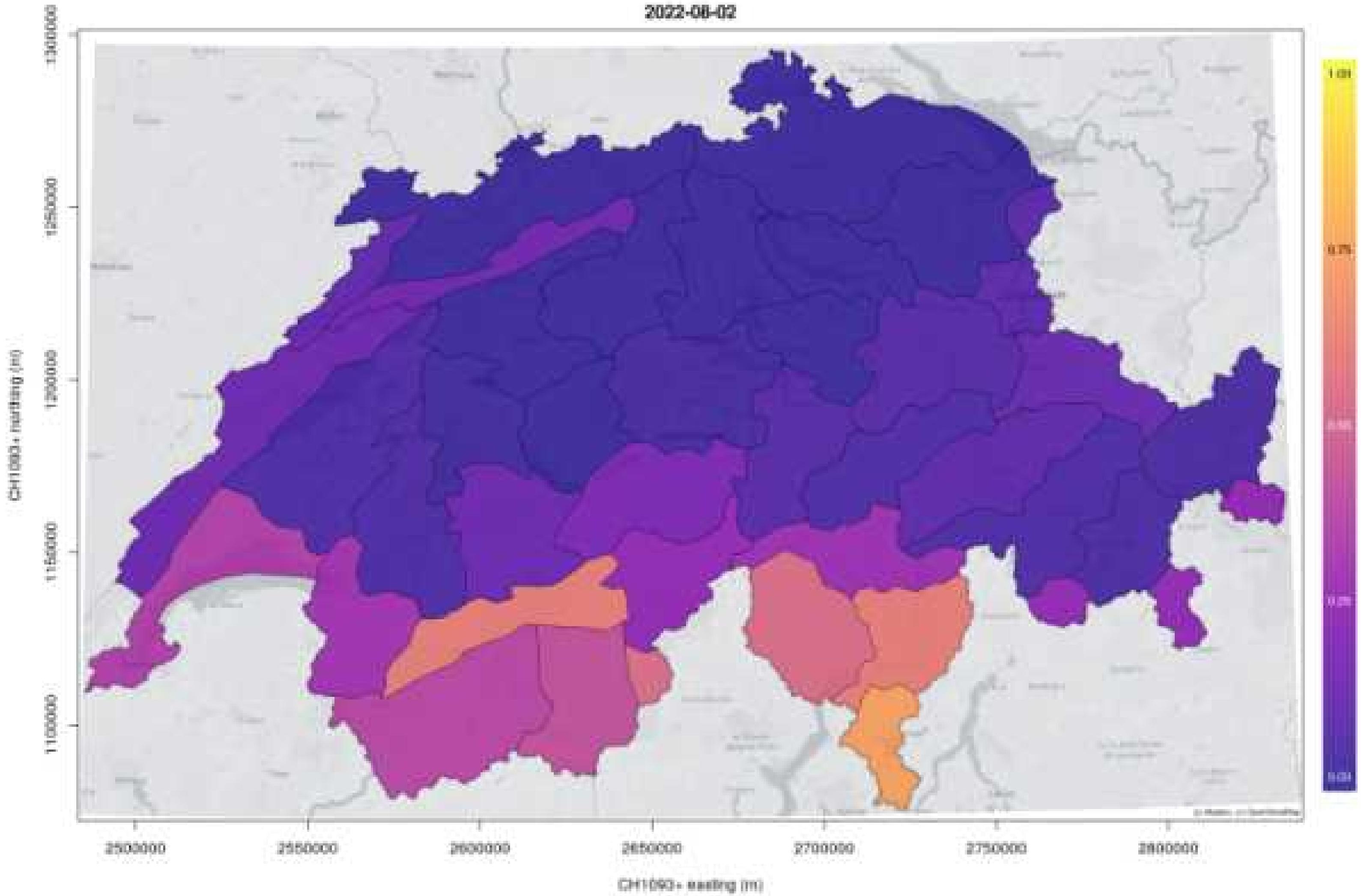
Predicted TWD index map (joint fit) for 2022-08-01, with forest presence mask

Joint Fit Predicted Map III



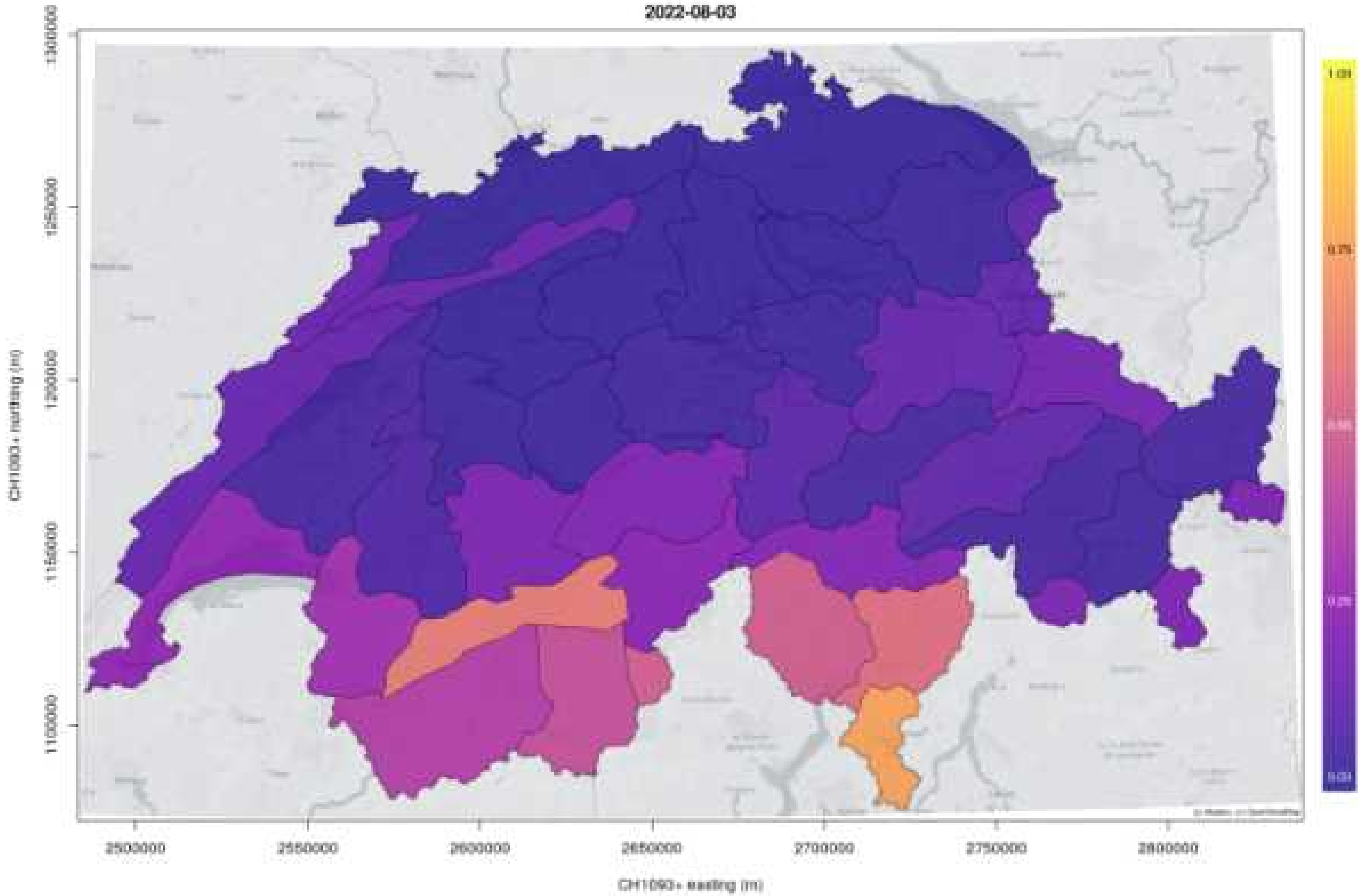
Predicted TWD index map (joint fit) for 2022-08-01, median over BAFU large regions

Joint Fit Predicted Map IV



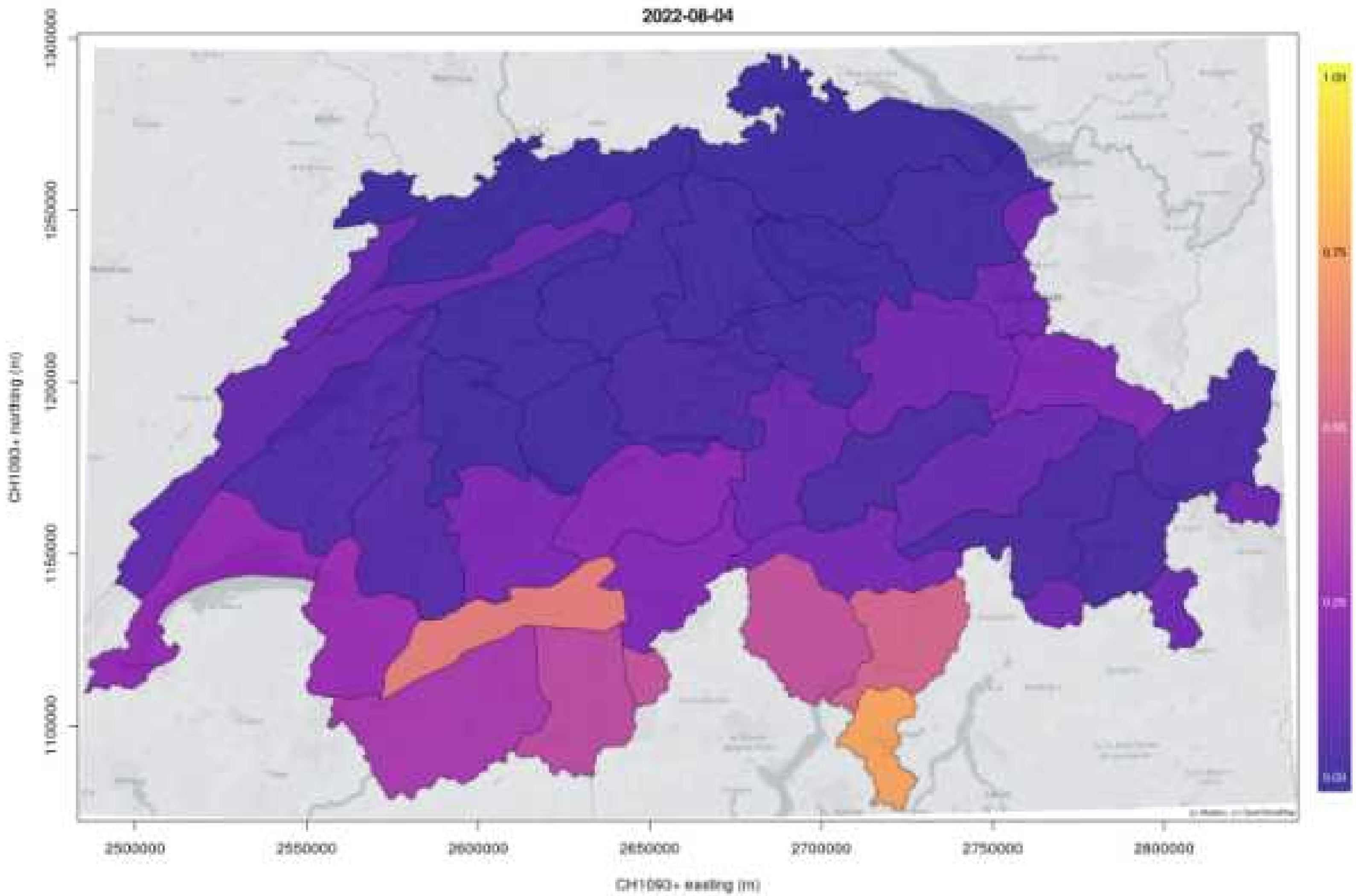
Predicted TWD index map (joint fit) for 2022-08-02, median over BAFU large regions

Joint Fit Predicted Map V



Predicted TWD index map (joint fit) for 2022-08-03, median over BAFU large regions

Joint Fit Predicted Map VI



Predicted TWD index map (joint fit) for 2022-08-04, median over BAFU large regions

Discussion

- Challenging spatial prediction of TWD (index) turned into a feasible temporal modeling task
- Non-autoregressive LSTM model allows to predict at any location/time and accommodates arbitrary missing values in TWD time series.
- Quite fast: on an NVIDIA A100 GPU (using 0.1 of its memory), it takes $\approx$ 30 min for training and 60–90 s to predict one day over 320×224 grid.
- Prediction performance could be improved. Some ideas:
  - Use static variables in addition to dynamic ones, both species- and site-specific
  - Predict multiple days and grid cells at once and enforce some smoothing/regularization in space-time
  - Other recurrent neural network architectures

References I

Federer, C. A., C. Vörösmarty, and B. Fekete (2003). Sensitivity of annual evaporation to soil and root properties in two models of contrasting complexity. *Journal of Hydrometeorology* 4 (6), 1276–1290.

Hochreiter, S. and J. Schmidhuber (1997). Long short-term memory. *Neural computation* 9 (8), 1735–1780.