Groundwater level prediction in the Doñana wetland area with the Prophet forecasting approach

Aguilera, H.[1], Guardiola-Albert, C.[1], Naranjo-Fernández, N.[1], Ruiz-Bermudo, F.[2]

1 Instituto Geológico y Minero de España (IGME), C/ Ríos Rosas, 23. 28003, Madrid. **h.aguilera@igme.es**

2 Instituto Geológico y Minero de España (IGME),. Oficina de Sevilla

INTRODUCTION

There is an increasing need for accurate groundwater level (GWL) prediction in environmentally sensible and/or pumping areas to support water management.

Some time series models such as ARIMA are not flexible enough (i.e. regular periods, stationarity, uncertainty, etc). Others based on machine learning techniques such as recurrent neural networks are too complex for non-expert analysts.

Here we apply a new simple, flexible and powerful forecasting approach to predict GWL in the surroundings of the tourist area of Matalascañas in the Doñana National Park. (Fig. 1)
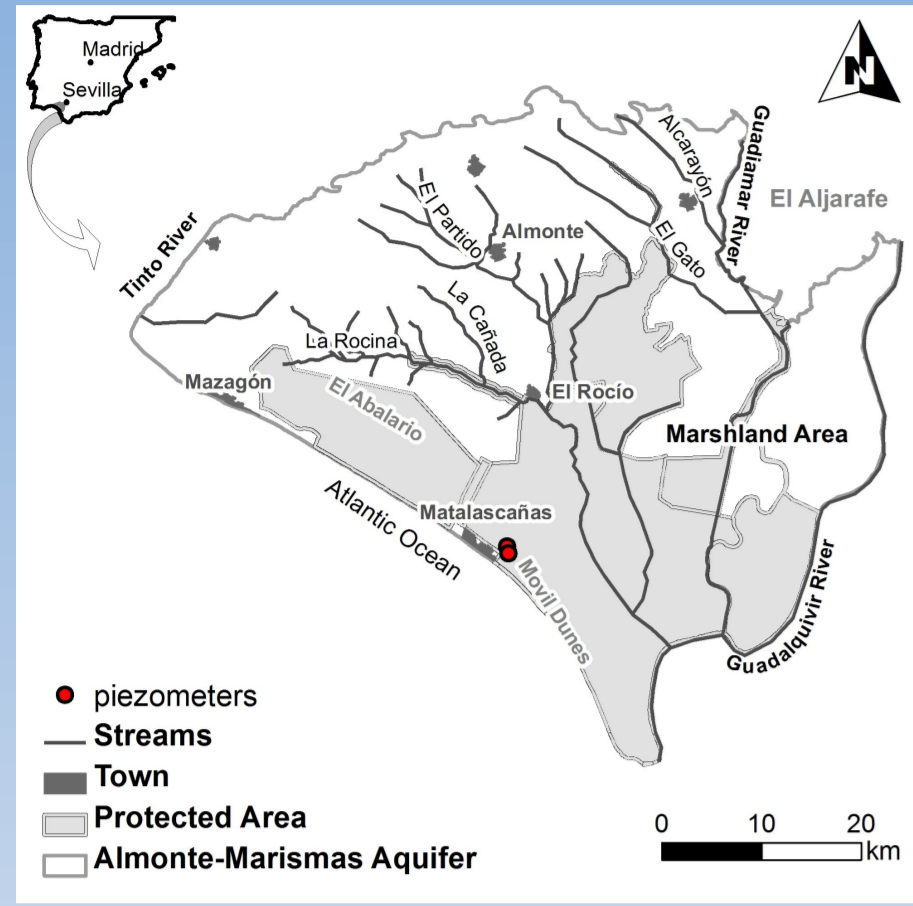

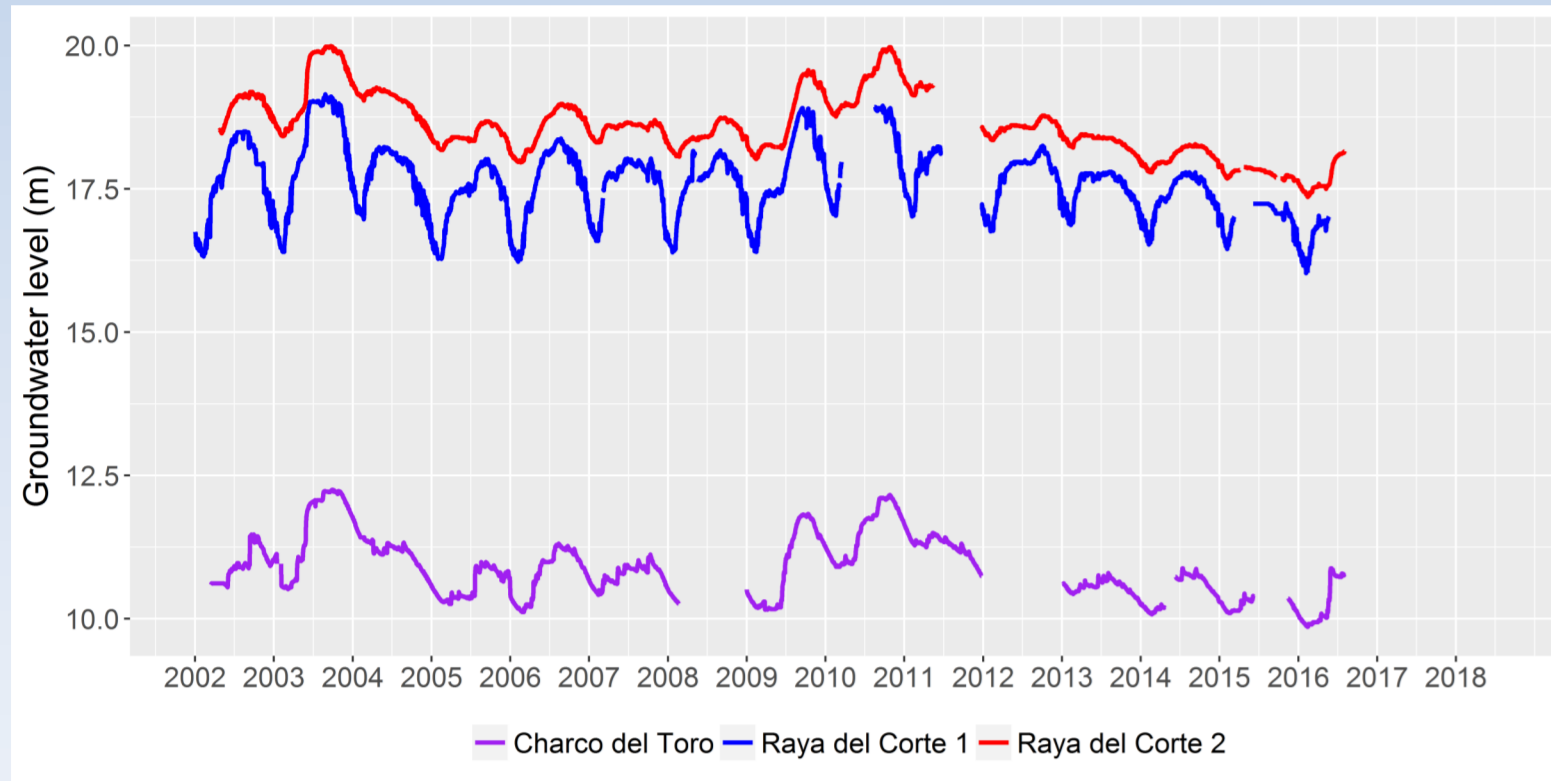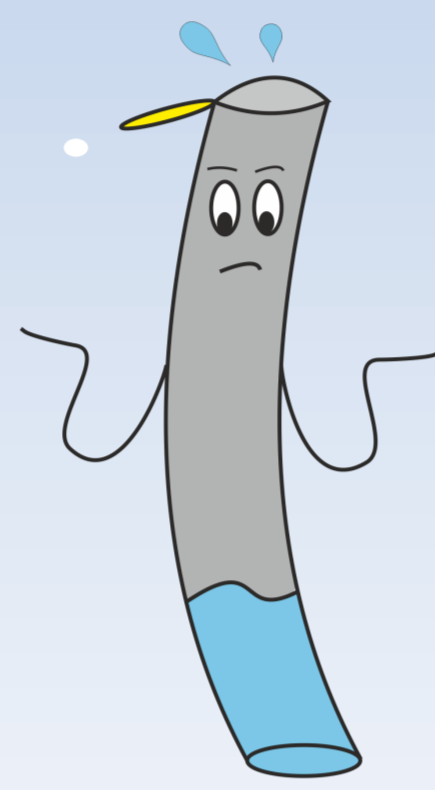**Figure 1.**Location map of piezometers.

Figure 2. Time series of groundwater levels in selected piezometers.

METHODOLOGY

Prophet is a procedure for forecasting time series data released as open source software by Facebook in 2017 (Taylor and Letham, 2017).

It is based on an additive model where linear or non-linear (logistic) trends (g(t)) are fit with yearly and weekly seasonality (st)), plus holidays (h(t)):

$$Y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Prophet uses a Bayesian framework based on prior distributions on the parameters to perform posterior inference including forecast uncertainty: normal distribution for smoothing prior on seasonality and holidays and double exponential for trend changepoints (these can also be directly specified).

Prophet is robust to missing data, shifts in the trend, and large outliers. Packages are available for R and Python.
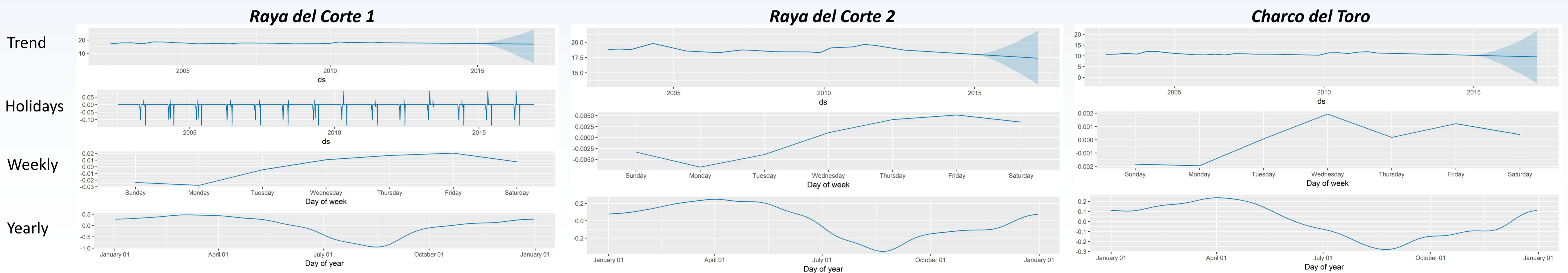
RESULTS

Figure 3. Components of the Prophet forecast for the three piezometers. For the trend and holidays plots the x-axis is labeled 'ds', which stands for 'date stamp'. Only the deeper piezometer Raya del Corte 1 showed a significant holidays effect with an optimum span of days around each holiday of 7 days before and after Eastern Friday, and just the following 7 days for the other two holidays . Notice the negative effect of both holidays as well as weekends on GWLs associated to tourism in Matalascañas. Decreasing trends are observed in shallower aquifer levels (Raya del Corte 2)
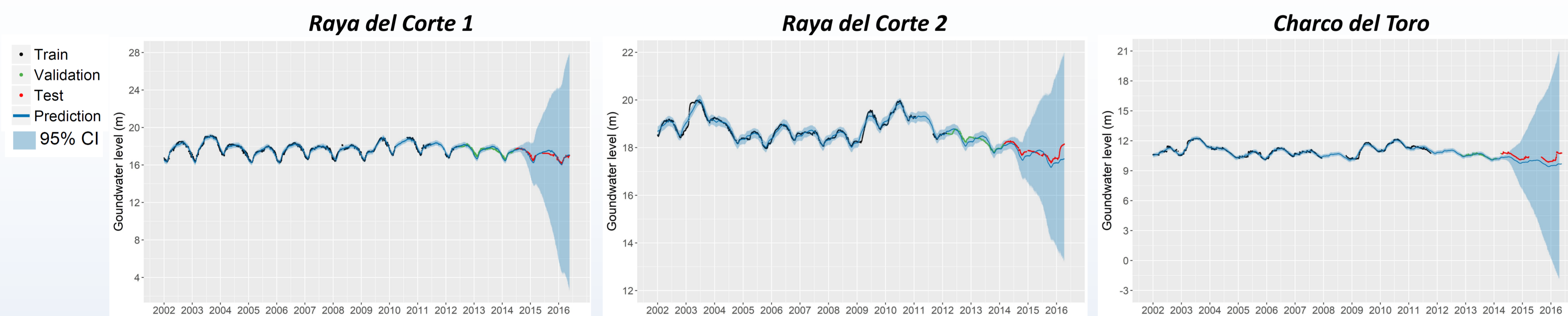
	Raya del Corte 1		Raya del Corte 2		Charco del Toro	
	Train set 2002-2014	Test set 2015-2016	Train set 2002-2014	Test set 2015-2017	Train set 2002-2014	Test set 2015-2017
MAE (m)	0.11	0.15	0.08	0.17	0.08	0.50
MAPE (m)	0.64	0.90	0.42	0.97	0.71	4.80
RMSE (m)	0.14	0.20	0.11	0.22	0.11	0.55

Figure 4. Model fitting plots. The predicted line shows how well the model is capturing the different hydrogeological behaviours. We can see some of the main strenghts of Prophet: the ability to deal with missing data, outliers and dramatic changes in the time series. Uncertainty intervals widen as we move into the future, but they remain quite narrow for the first six months in the test forecast. The errors are low (MAPE < 1%), except for the Charco del Toro where a sharp GWL increase at the end of the forecasting period could not be captured by the model.
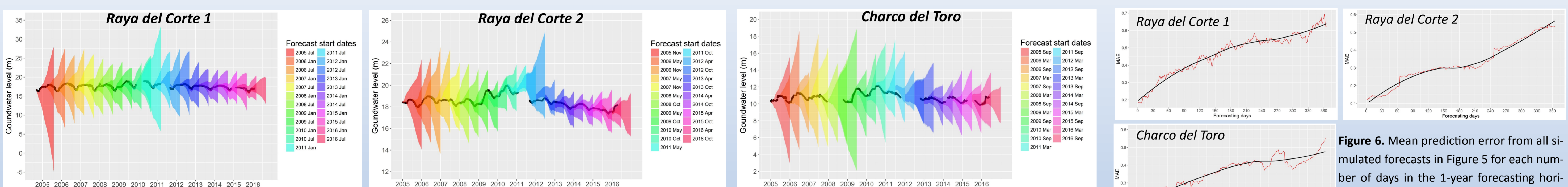

Figure 5. Simulated historical forecasts (SHF) to assess model forecast accuracy. Each color band corresponds to 95% confidence intervals for 1-year forecasts starting at the 6-month interval start dates. Overall, uncertainty increases as we move forward into the future in each simulated horizon and it decreases as we move forward towards the end of the series, thus, improving the ability to predict the future. Also, the widest bands correspond to periods with missing information.
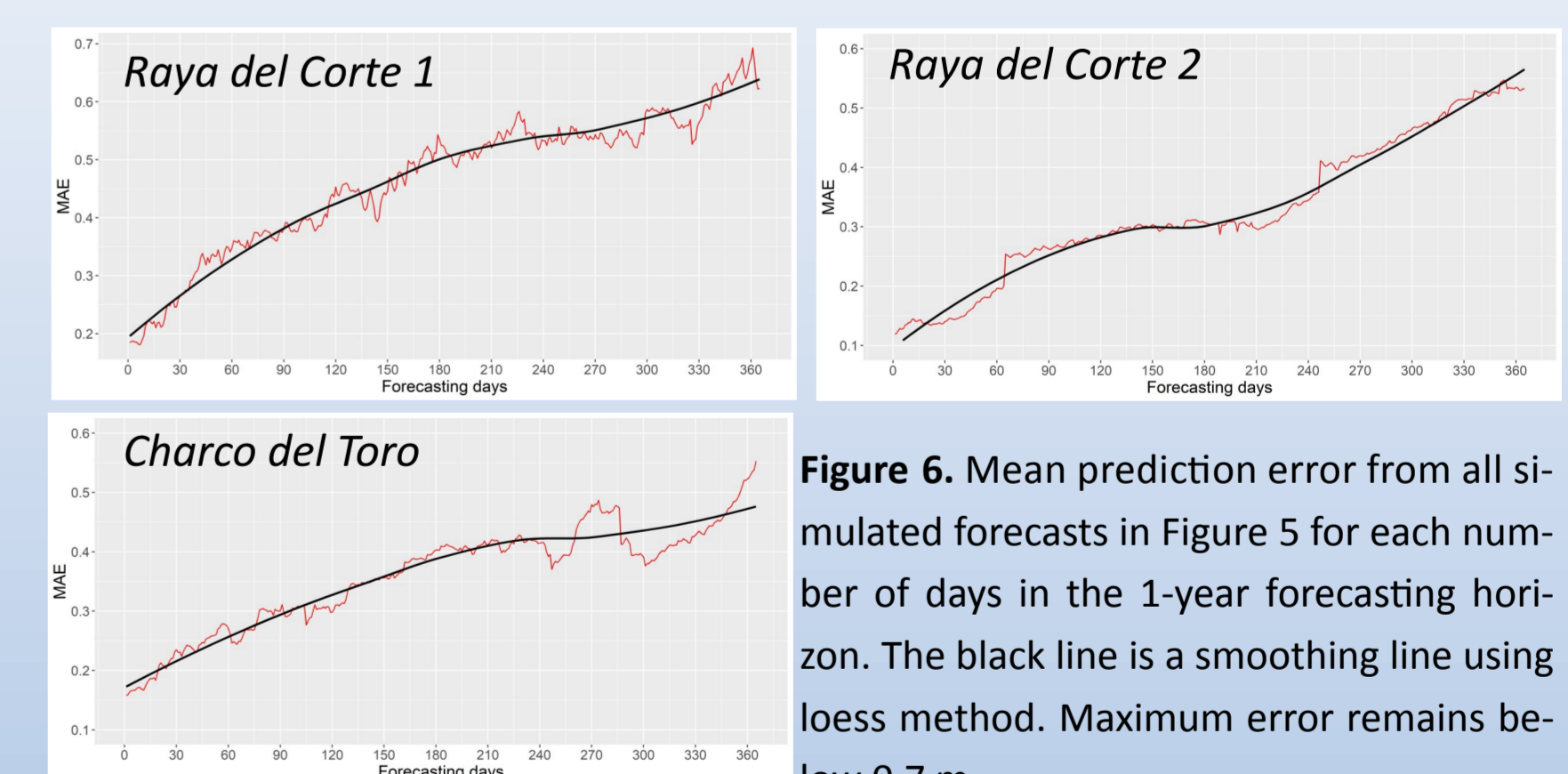
Figure 6. Mean prediction error from all simulated forecasts in Figure 5 for each number of days in the 1-year forecasting horizon. The black line is a smoothing line using loess method. Maximum error remains below 0.7 m.

CONCLUSIONS

- The Prophet model by Facebook has been proven to be a useful tool for GWL prediction in piezometers showing different behaviour.

- The model corroborates previous findings in the Doñana area near the tourist village of Matalascañas, where deeper GWLs (Raya del Corte 1) are strongly affected by pumping as shown by the negative holidays and weekend components of the model.

- The model fitting procedure for maximum forecasting power has yielded very good results, even in the presence of missing data gaps and trend changes. SHF were used to further assess model robustness and forecast accuracy. Prediction uncertainty was calculated at several 1-year forecasting horizons along the series. MAE of these 1-year forecasts remained below 0.7 m.

- Prophet default settings often produce accurate forecasts and an analyst or manager with no training in time series methods can improve or tweak forecasts using a variety of easily-interpretable parameters. For example, smoothing parameters for seasonality allow adjusting how closely to fit historical cycles or holidays, as well as smoothing parameters for trends allow adjusting how aggressively to follow historical trend changes such as pumping influence.

REFERENCES

Taylor SJ, Letham B. (2017) Forecasting at scale. *PeerJ Preprints* 5:e3190v2 https://doi.org/10.7287/peerj.preprints.3190v2

Rebollo AM, Mediavilla C, Ruiz F, Díaz AF (2008). Análisis de la influencia de los bombeos de Matalascañas en el complejo lagunar Charco del Toro-Laguna de Santa Olalla, parque nacional de Doñana (Huelva). In: López-Geta JA, Rubio JC and Matín-Machuca M (eds) *VII Simposio del Agua en Andalucía*. IGME, pp 655-664

https://facebook.github.io/prophet/

Congress on Groundwater and Global Change in the Western Mediterranean 6th - 9th November 2017