

# ARIK GPP model report

2023-01-30

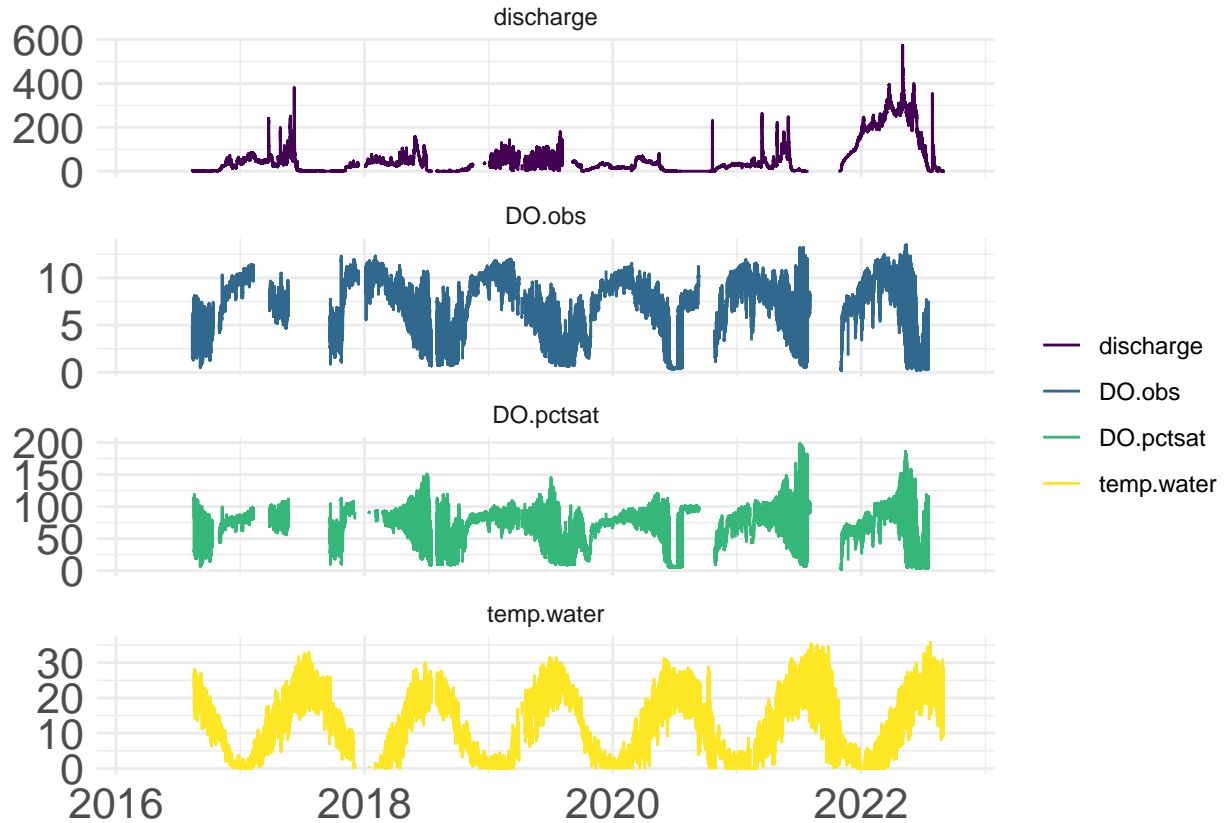


Figure 1: Full metabolism data streams.

The process for building a model suite and selecting a model is iterative. More details for data cleaning will be added here later.

We undergo a process to slim these models down based on a number of criteria. First, we throw out values that have very high estimates of GPP ( $> 1.5\times$  maximum annual values observed in Bernhardt et al. 2022 Figure 1,  $\sim 5000 \text{ g C m}^{-2} \text{ y}^{-1}$ ). Next, any models that have especially poor fits are excluded. In this case I automatically exclude any models that are in the upper 30% percentile of RMSE based on the model set, unless this is fewer than 5 models. This slimmed model set is the group I choose from after one final visual assessment.

I do a visual assessment of the slimmed model set to see how they compare and identify any days where the fit is especially poor across dates. To do this, I first chose the 'best' model of the set:

The 'best' model is: mm1\_satq10

I then compute the model that is most different than the 'best' model and plot the metabolism estimates for each to visually assess.

Table 1: Full model output for all MLE models. 'modelID' represents the model identifier and also information regarding the assumed equations for GPPER. Unmodified models (e.g., 'mm1') represent the default 'mm\_specs' of linear GPPlight relationship. Saturating models (e.g., 'mm1sat') assume a saturating GPPlight relationship. Lastly, Q10 models (e.g., 'mm1satq10') assume a saturating function of GPPlight and an exponential function of GPPTemp. 'modelType' represents how daily K600 values are calculated, whether they are modeled simultaneously with GPP and ER (e.g., 'raw'), modeled as a function of daily discharge with varying relationships (e.g., 'loess', 'lm', 'mean'), modeled from observed night observation (e.g., 'night'), or from empirical gas release estimates (e.g., 'empiricalgam'). The empirical models are further identified by the model used to relate discharge to k, whether GAM or LM. 'RSME' represents the estimated relative root mean square error between observed and modeled dissolved oxygen concentrations standardized to the mean. 'negativeGPP' and 'positiveER' represent the number of days when daily estimates of GPP or ER are negative and positive, respectively. 'meanGPP' is the estimated mean GPP. This value can be scaled to different timeframes or units (i.e., carbon or oxygen) and must be check based on attributes of 'meanGPP'. 'maxK' is the maximum estimated daily k600 value.

modelID	modelType	RMSE	negativeGPP	positiveER	maxK	ER.Kcorr	meanGPP
mm1	raw	26.28	83	105	100400.385	-0.862	10320.356
mm1_sat	raw	24.66	47	85	60701.076	-0.976	5032.246
mm1_satq10	raw	21.92	42	81	14016.096	-0.965	1010.292
mm2	loess	29.03	126	0	17.962	-0.554	1329.131
mm2_sat	loess	27.94	101	0	19.686	-0.620	1142.306
mm2_satq10	loess	23.56	60	27	14016.096	-0.969	656.136
mm3	lm	29.33	127	0	16.173	-0.552	1254.972
mm3_sat	lm	27.05	98	0	25.391	-0.657	1173.342
mm3_satq10	lm	24.39	64	27	14016.096	-0.969	649.366
mm4	mean	31.18	129	0	9.658	NA	1072.305
mm4_sat	mean	27.54	102	0	9.259	NA	909.613
mm4_satq10	mean	24.89	66	27	14016.096	-0.969	527.792
mm5	night	168.23	126	233	10.907	NA	1275.865
mm5_sat	night	573.90	72	146	10.814	NA	1118.463
mm5_satq10	night	31.51	61	154	9.899	NA	1118.220

<sup>a</sup> GPPmean = g C m<sup>-2</sup> y<sup>-1</sup>

The most distant from best is: mm4\_sat

These plots are used to assess any dates that are especially bad (e.g., negative GPP, positive ER, very high GPP, etc.) across all models. This dates identified below for the 'top' model and are excluded from the analysis.

The most distant of the best models. Days between this model and above that are both bad are likely very difficult days to fit. Depending on how many of them there are we need to remove them or take a look at the QC flags or the timeseries data to decide if they should be thrown out.

The number of dates and proportion of bad dates within the best model are:

From here we assess any data removal and rerun until reasonably clean model outputs are achieved and a final model is chosen based on minimizing 'negativeGPP' and 'positiveER' and assessing the reasonableness of 'maxK'.

Lastly, we then take the top model suite and build an average model based on model fit, RMSE. To do so, we first remove models that have very high correlations between ER and K (i.e.,  $r \geq 0.80$ ) as these models are largely driven by variability in the estimates in K. Then we create a model weight based on relative model fit,  $dRMSE$ , the difference in relative root mean squared error from the best fitting model based on  $RMSE$  for a set of models,  $m$ , is:

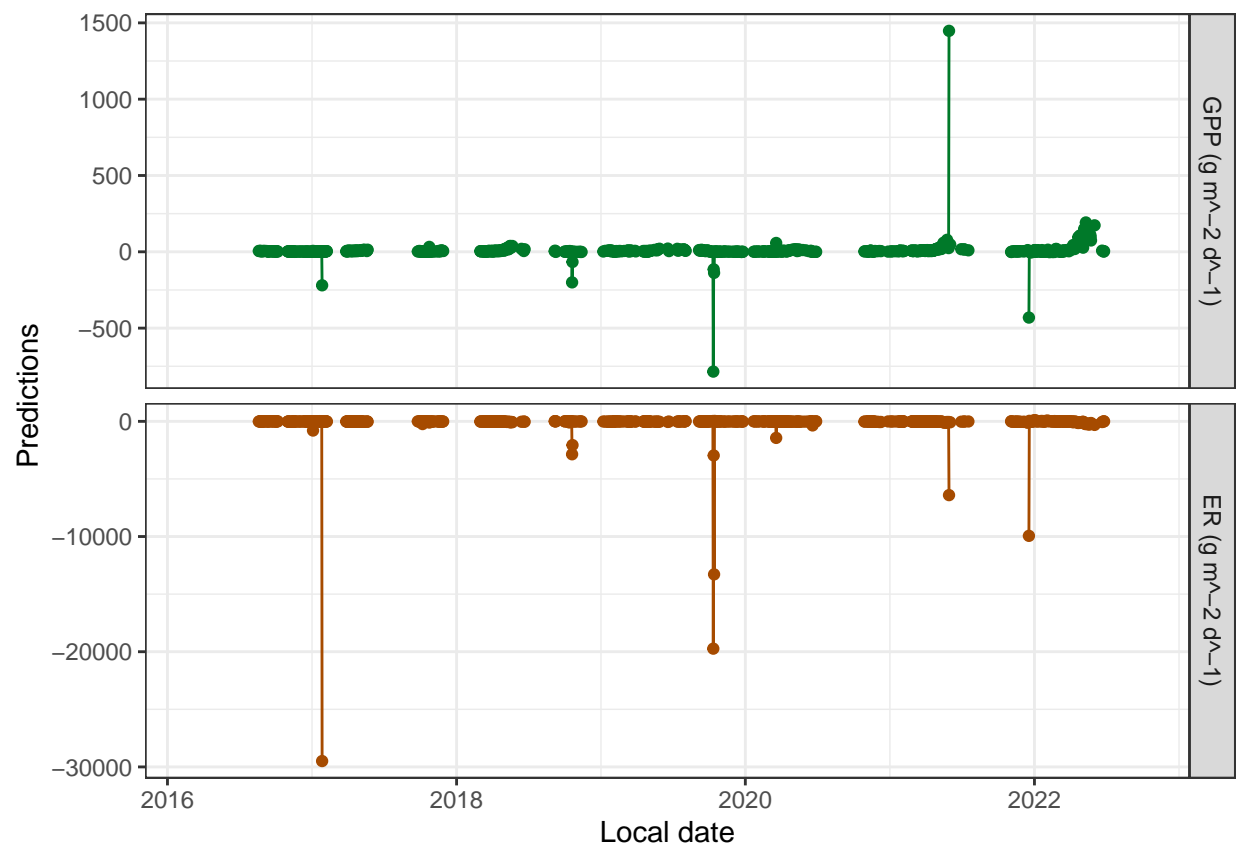


Figure 2: 'Best' model fit.

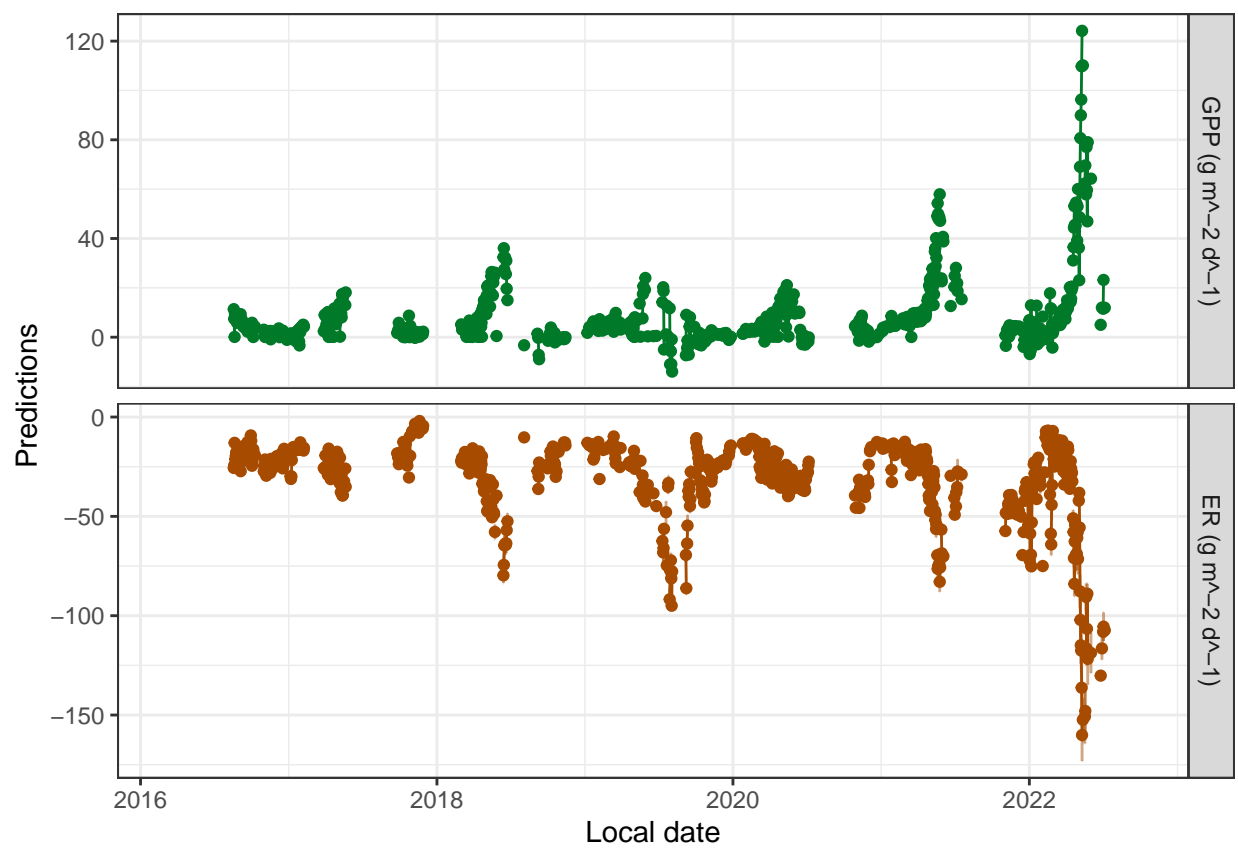


Figure 3: Most distant model.

Table 2: Slimmed model output for all MLE models. 'modelID' represents the model identifier and also information regarding the assumed equations for GPPER. Unmodified models (e.g., 'mm1') represent the default 'mm\_specs' of linear GPPlight relationship. Saturating models (e.g., 'mm1sat') assume a saturating GPPlight relationship. Lastly, Q10 models (e.g., 'mm1satq10') assume a saturating function of GPPlight and an exponential function of GPPTemp. 'modelType' represents how daily K600 values are calculated, whether they are modeled simultaneously with GPP and ER (e.g., 'raw'), modeled as a function of daily discharge with varying relationships (e.g., 'loess', 'lm', 'mean'), modeled from observed night observation (e.g., 'night'), or from empirical gas release estimates (e.g., 'empirical-gam'). The empirical models are further identified by the model used to relate discharge to k, whether GAM or LM. 'RSME' represents the estimated relative root mean square error between observed and modeled dissolved oxygen concentrations standardized to the mean. 'negativeGPP' and 'positiveER' represent the number of days when daily estimates of GPP or ER are negative and positive, respectively. 'meanGPP' is the estimated mean GPP. This value can be scaled to different timeframes or units (i.e., carbon or oxygen) and must be check based on attributes of 'meanGPP'. 'maxK' is the maximum estimated daily k600 value.

modelID	modelType	RMSE	negativeGPP	positiveER	maxK	ER.Kcorr	meanGPP
mm1_satq10	raw	21.92	42	81	14016.096	-0.965	1010.292
mm2	loess	29.03	126	0	17.962	-0.554	1329.131
mm2_sat	loess	27.94	101	0	19.686	-0.620	1142.306
mm2_satq10	loess	23.56	60	27	14016.096	-0.969	656.136
mm3_sat	lm	27.05	98	0	25.391	-0.657	1173.342
mm3_satq10	lm	24.39	64	27	14016.096	-0.969	649.366
mm4_sat	mean	27.54	102	0	9.259	NA	909.613
mm4_satq10	mean	24.89	66	27	14016.096	-0.969	527.792

<sup>a</sup> GPPmean = g C m<sup>-2</sup> y<sup>-1</sup>

lowGPP	lowGPPperc	highGPP	highGPPperc	NAs
68	4.782	74	5.204	3264

$$w_i = \exp(\frac{1}{2}dRMSE_i) / \sum_{j=1}^m \exp(\frac{1}{2}dRMSE_j)$$

From this, we create a weighted average model for GPP estimates, plotted below.

modelID	modelType	RMSE	negativeGPP	positiveER	maxK	ER.Kcorr	meanGPP	dRMSE	RMSEwt
mm2	loess	29.03	126	0	17.962	-0.554	1329.131	-1.98	0.133
mm2_sat	loess	27.94	101	0	19.686	-0.620	1142.306	-0.89	0.229
mm3_sat	lm	27.05	98	0	25.391	-0.657	1173.342	0.00	0.358
mm4_sat	mean	27.54	102	0	9.259	NA	909.613	-0.49	0.280

