Supplement of

Regionalization of hydrological model parameters using gradient boost machine

Song et al.

Correspondence to: Jun Xia (xiajun666@whu.edu.cn) and Gangsheng Wang (wanggs@whu.edu.cn)

Supplement

1 DTVGM-PML modelling framework

The modified version of DTVGM incorporates a snow routine derived from HBV model (Seibert and Vis, 2012), an ET routine including PML model (Bai et al., 2018; Leuning et al., 2008; Zhang et al., 2008) and the interception model of Gash (Zhang et al., 2019; Zhang et al., 2016). The overview of model structure is summarized as follows.

1.1 Snow routine

The snowmelt routine which is based on the HBV model simulates daily snowmelt using daily precipitation and temperature as input. Precipitation is separated into snowfall or rainfall depending on whether the temperature is above or below a threshold temperature, T_T, °C, as follows:

$$P_s = \begin{cases} 0, & if\ T \geq T_T \\ P, & if\ T < T_T \end{cases} \tag{S1}$$

$$P_r = P - P_s \tag{S2}$$

where P is daily precipitation (mm day^{-1}); P_s is snowfall (mm day^{-1}); P_r is rainfall (mm day^{-1}); T is daily air temperature (°C). The snowmelt, S_M (mm day^{-1}), is computed with the degree-day method using the degree-day factor C_{MELT} (mm day^{-1} °C^{-1}, typically around 4 mm day^{-1} °C^{-1}, lower values for forested areas compared to open areas), as follows:

$$S_M = \begin{cases} C_{MELT}(T - T_T), & if\ T \geq T_T \\ 0, & if\ T < T_T \end{cases} \tag{S3}$$

Snowpack retains melt water until amount exceeds a certain portion (usually 0.1), of the water equivalent of the snow. When temperatures decrease below T_T, water refreezes again and the amount of refreezing liquid water within snowpack, R_F (mm day^{-1}), is calculated using a refreezing coefficient, C_{FR} (-), as follows:

$$R_F = \begin{cases} 0, & if\ T \geq T_T \\ C_{FR}C_{MELT}(T - T_T), & if\ T < T_T \end{cases} \tag{S4}$$

1.2 ET routine

The PML equation (Zhang et al., 2019; Zhang et al., 2016; Leuning et al., 2008) estimates soil evaporation (Es) and transpiration (Et) using:

$$E_t + E_s = \frac{\varepsilon A + (\rho c_p/\gamma) D_a G_a}{\varepsilon + 1 + G_a/G_s} \tag{S5}$$

$$E_t = \frac{\varepsilon A_c + (\rho c_p/\gamma) D_a G_a}{\varepsilon + 1 + G_a/G_c} \tag{S6}$$

$$E_s = \frac{f \varepsilon A_s}{\varepsilon + 1} \tag{S7}$$

where $\varepsilon = s/\gamma$ (-), in which γ is the psychrometric constant (kPa °C^{-1}) and s is the slope of the curve relating saturation water vapor pressure to temperature (kPa °C^{-1}); ρ is the density of air (kg m^{-3}); c_p is specific heat of air at constant pressure (MJ kg^{-1} K^{-1}); D_a is the water vapor pressure deficit of the air (kPa); G_a is the aerodynamic conductance (m s^{-1}); G_s is the surface conductance (m s^{-1}); G_c is the canopy conductance (m s^{-1}); f is the fraction of evaporation from the soil of the equilibrium rate at the soil surface that determines the water availability for soil evaporation (-); A is the available energy absorbed by the surface (MJ m^{-2} day^{-1}), which is partitioned using leaf area index (LAI) into the canopy absorption, A_c (MJ m^{-2} day^{-1}), and the soil absorption, A_s (MJ m^{-2} day^{-1}). The absorbed fraction of canopy and soil are given respectively by $A_c/A = 1 - \tau$, and $A_s/A = \tau$, where $\tau = \exp(-k_A LAI)$, k_A is an extinction coefficient for A.

The canopy conductance, G_c, is given by

$$G_c = \frac{g_{sx}}{k_Q} ln\left[\frac{Q_h + Q_{50}}{Q_h exp(-k_Q LAI) + Q_{50}}\right]\left[\frac{1}{1 + D_a/D_{50}}\right] \tag{S8}$$

where g_{sx} is the maximum stomatal conductance (m s^{-1}) of leaves at the top of the canopy, k_Q is the extinction coefficient for photosynthetically active radiation, Q_h is the photosynthetically active radiation at the top of canopy (MJ m^{-2} day^{-1}), Q_{50} (MJ m^{-2} day^{-1}) and D_{50} are the values of absorbed photosynthetically active radiation and water vapor deficit at which stomatal conductance is half its maximum value.

The soil evaporation fraction f varies from $f = 1$ when the soil surface is wet to $f = 0$ when dry. It is estimated by the relative soil moisture content $\frac{W}{W_M}$ in Eq. (S12).

The evaporation of precipitation intercepted by the vegetation (E_i) is estimated using a modified version of the widely adopted rainfall interception model of Gash (Zhang et al., 2016), expressed as

$$E_i = f_v P, \qquad if\ P < P_{wet} \tag{S9}$$

$$E_i = f_v P_{wet} + f_{ER}(P - P_{wet}), \quad if\ P \geq P_{wet} \tag{S10}$$

with

$$P_{wet} = -ln\left(1 - \frac{f_{ER}}{f_v}\right)\frac{S_v}{f_{ER}}, \quad S_v = S_l LAI, \quad f_{ER} = f_v F_0, \quad f_v = 1 - exp\left(LAI/LAI_{ref}\right) \tag{S11}$$

where f_{ER} is the ratio of average evaporation rate over average precipitation intensity storms (-); f_v is the fractional area covered by intercepting leaves (-); P is the daily precipitation (mm day^{-1}); P_{wet} is the reference threshold rainfall amount if the canopy is wet (mm day^{-1}), and S_v is the canopy rainfall storage capacity (mm day^{-1}). S_l is the water storage capacity per unit leaf area (mm); F_0 is the specific ratio of average evaporation rate over average rainfall intensity during storms per unit canopy cover (-); LAI_{ref} is the reference leaf area index.

1.3 Runoff routine

As suggested by Xia et al., (2005) who developed a simple relationship between the time-variant runoff coefficient and soil moisture in terms of hydrological data set of more than forty basins in the world, the surface runoff, R_s, is calculated by the following equation,

$$R_s = g_1(\frac{W}{W_M})^{g_2} P_t, \tag{S12}$$

where g_1 is the runoff coefficient when the soil moisture content is equal to the saturated soil moisture, and g_2 is the impact coefficient of the soil moisture content, $\frac{W}{W_M}$ is the relative soil moisture content defined as the ratio of the soil moisture content, W, to the saturated soil moisture content, W_M. P_t is the rainfall passing through the canopy.

The remaining water combining the snowmelt, S_m, become infiltration, I, into the soil moisture store (Eq. (S13)). The subsurface runoff, R_{ss}, is estimated as a linear function of relative soil moisture content (Eq. (S14)). The groundwater recharge, R_r, is also estimated as a linear function of relative soil moisture content (Eq. (S15)). The baseflow, R_g, is simulated as a linear function of the groundwater storage, G (Eq. (S16)).

$$I = P_t - R_s + S_m, \tag{S13}$$

$$R_{ss} = k_s \frac{W}{W_M} I, \tag{S14}$$

$$R_r = k_r \frac{W}{W_M}(I - R_{ss}), \tag{S15}$$

$$R_g = k_g G, \tag{S16}$$

$$\Delta \mathrm{W} = I - R_{ss} - R_r - ET, \tag{S17}$$

$$\Delta \mathrm{G} = R_r - R_g, \tag{S18}$$

where k_s is the subsurface runoff generation coefficient, k_r is the groundwater recharge coefficient, k_g is the groundwater runoff recession parameter, ET is the evapotranspiration generated by PML model.

Table S1: Basic information for the 31 hydrological stations

No.	Station Name	Lon	Lat	River	Area (km^2)	Available data
1	Ayanqian	124.63	48.77	Nenjiang	65439	1982-2004
2	Beibei	106.42	29.85	Jialingjiang	156142	2007-2012
3	Dengyingyan	104.73	29.90	Tuojiang	14484	1982-2008
4	Gaochang	104.42	28.80	Minjiang	136000	2007-2012
5	Gaoqitou	110.35	28.62	Yuanjiang	17698	1999-2005
6	Hongqi	103.57	35.80	Taohe	24973	1982-2009
7	Jian	114.98	27.10	Ganjiang	56223	1982-2009
8	Lanxi	119.47	29.22	Qiantangjiang	18233	1982-2008
9	Liangjiazi	123.00	46.73	Nenjiang	15544	1982-2009
10	Lijiadu	116.16	28.22	Fuhe	15811	1982-2012
11	Linyi	118.40	35.02	Huaihe	10315	1982-2007
12	Meigang	116.82	28.43	Xinjiang	15535	1982-2009
13	Pushi	110.12	28.10	Yuanjiang	54144	1982-2005
14	Sancha	108.95	24.47	Xijiang	16280	1982-2008
15	Shehong	105.40	30.87	Fujiang	23574	1982-2008
16	Shijiao	112.95	23.57	Beijiang	38383	1982-2008
17	Shimen	111.38	29.62	Lishui	15307	1982-2005
18	Tangnaihai	100.15	35.50	Huanghe	121970	1982-2007
19	Taojiang	112.12	28.53	Zishui	26748	1982-2005
20	Tingzikou	105.82	31.85	Jialingjiang	61089	1982-2008
21	Waizhou	115.84	28.63	Ganjiang	83777	1982-2012
22	Wuchang	127.10	44.87	Lalinhe	5642	1982-2002
23	Wulong	107.73	29.33	Wujiang	87920	1982-2012
24	Wuzhou	111.33	23.47	Xijiang	329700	1982-2008
25	Xiangjiaping	109.28	32.87	Xunhe	6448	1982-2008
26	Xiangtan	112.93	27.87	Xiangjiang	94660	1982-2012
27	Xiaoergou	123.72	49.20	Nenjiang	16761	1982-2009
28	Xixian	114.73	32.33	Huaihe	10190	1982-2003
29	Yajiang	101.01	30.03	Yalongjiang	65923	1982-2012
30	Yingluoxia	100.18	38.80	Heihe	10010	1982-2012
31	Zhuqi	119.10	26.15	Minjiang	54500	1982-2008

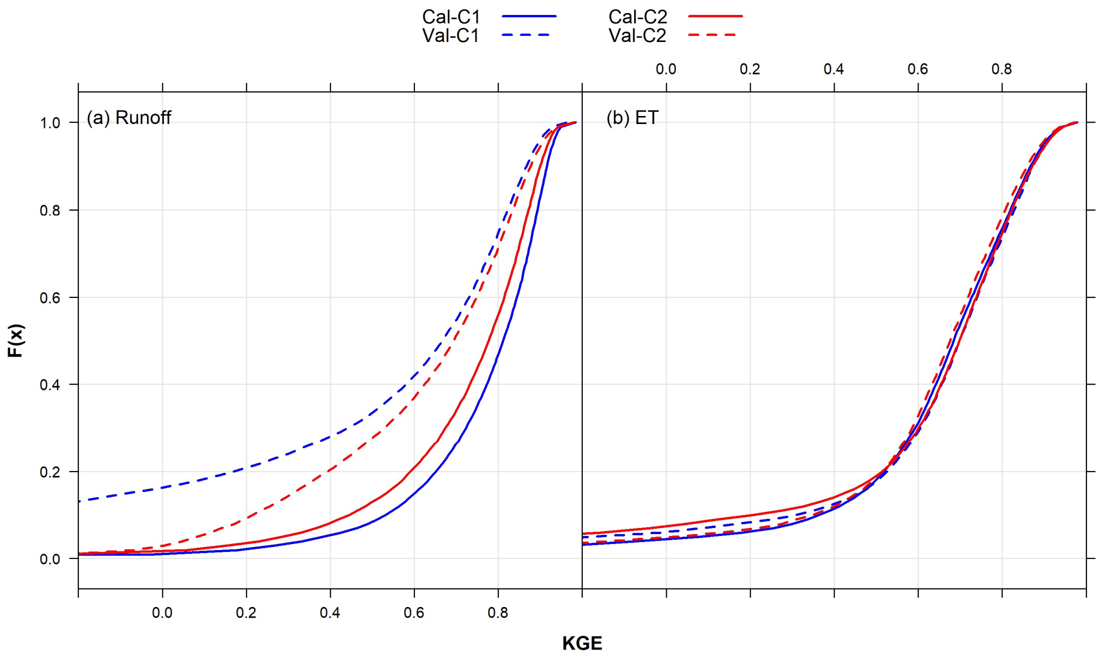


Figure S1: CDFs of KGE for (a) runoff and (b) ET simulation by two split-sample tests (C1: using the former period, 1982-1997, for calibration; C2: using the latter period, 1998-2012, for calibration) in calibration (solid lines) and validation (dashed lines) period.

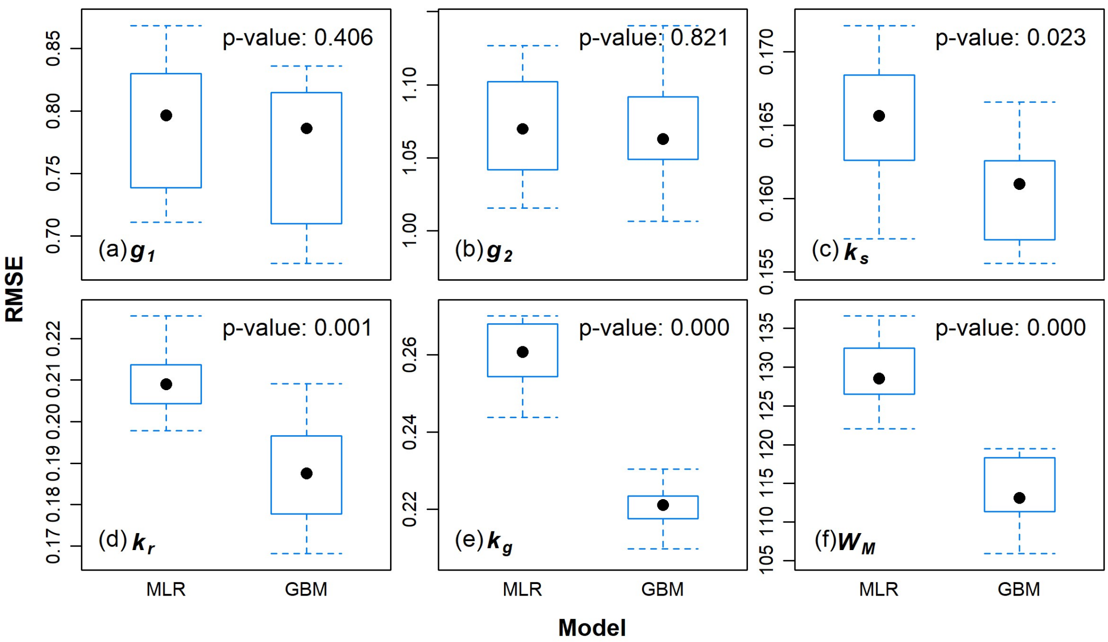


Figure S2: Performance evaluation of MLR and GBM for six parameters, (a) g_1, (b) g_2, (c) k_s, (d) k_r, (e) k_g, (f) W_M, in semi-humid region. MLR and GBM denote the multiple linear regression with stepwise selection and the gradient boosting machine model.

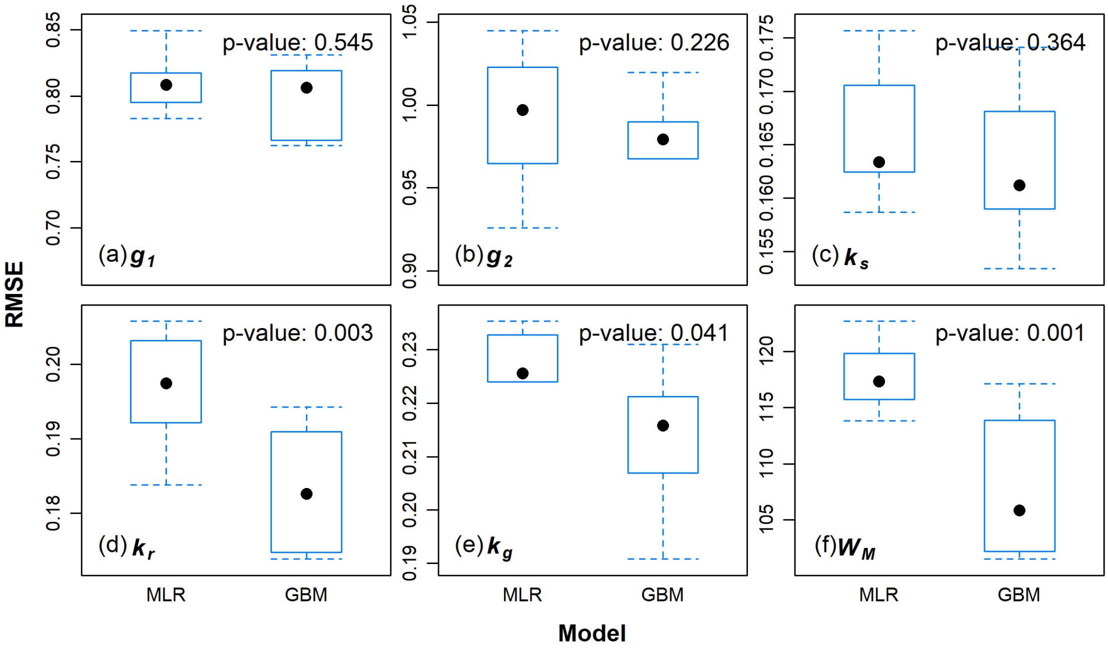


Figure S3: Same as Figure S1 but for semi-arid region.

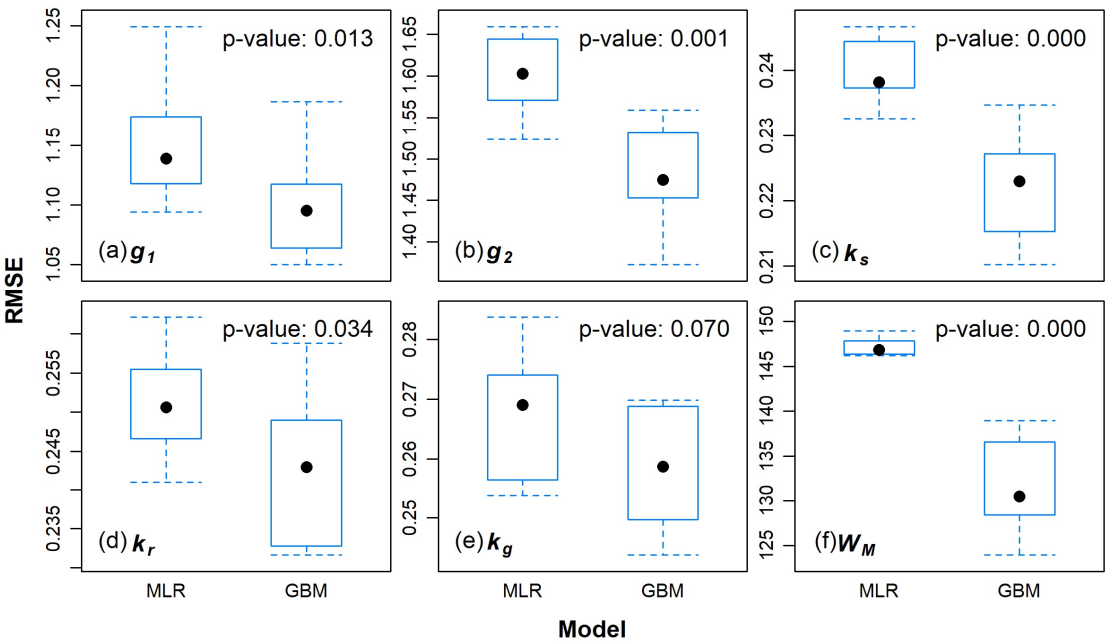


Figure S4: Same as Figure S1 but for arid region.

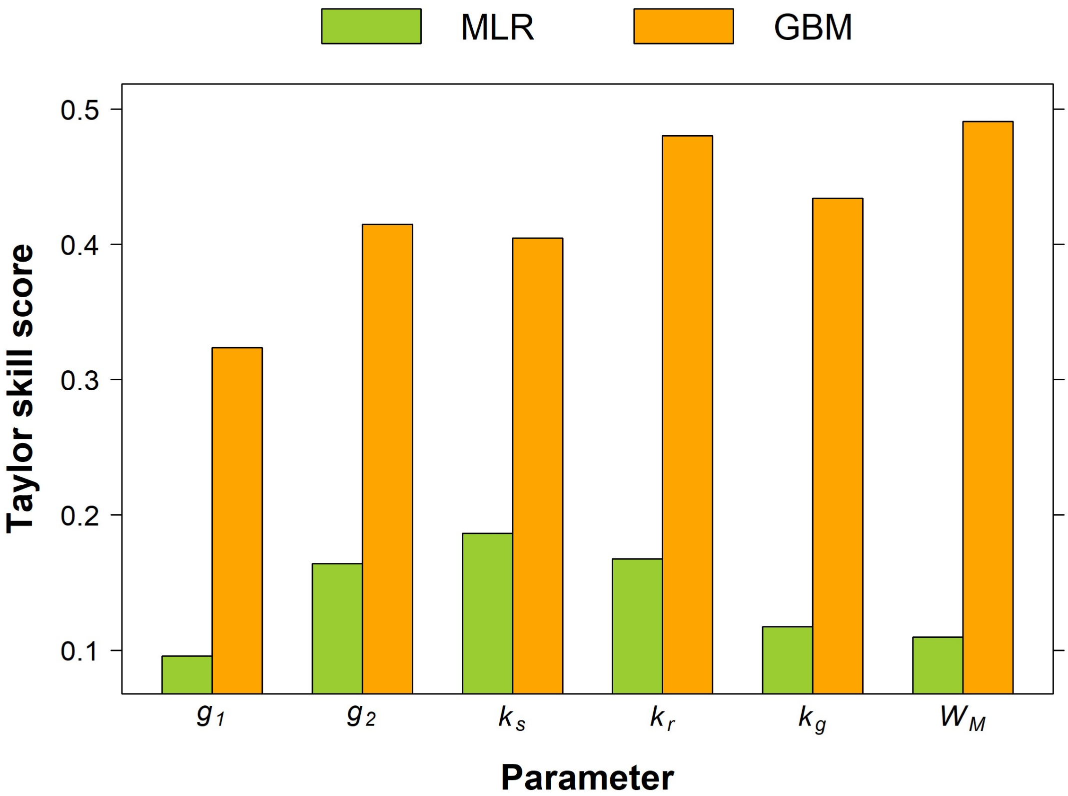
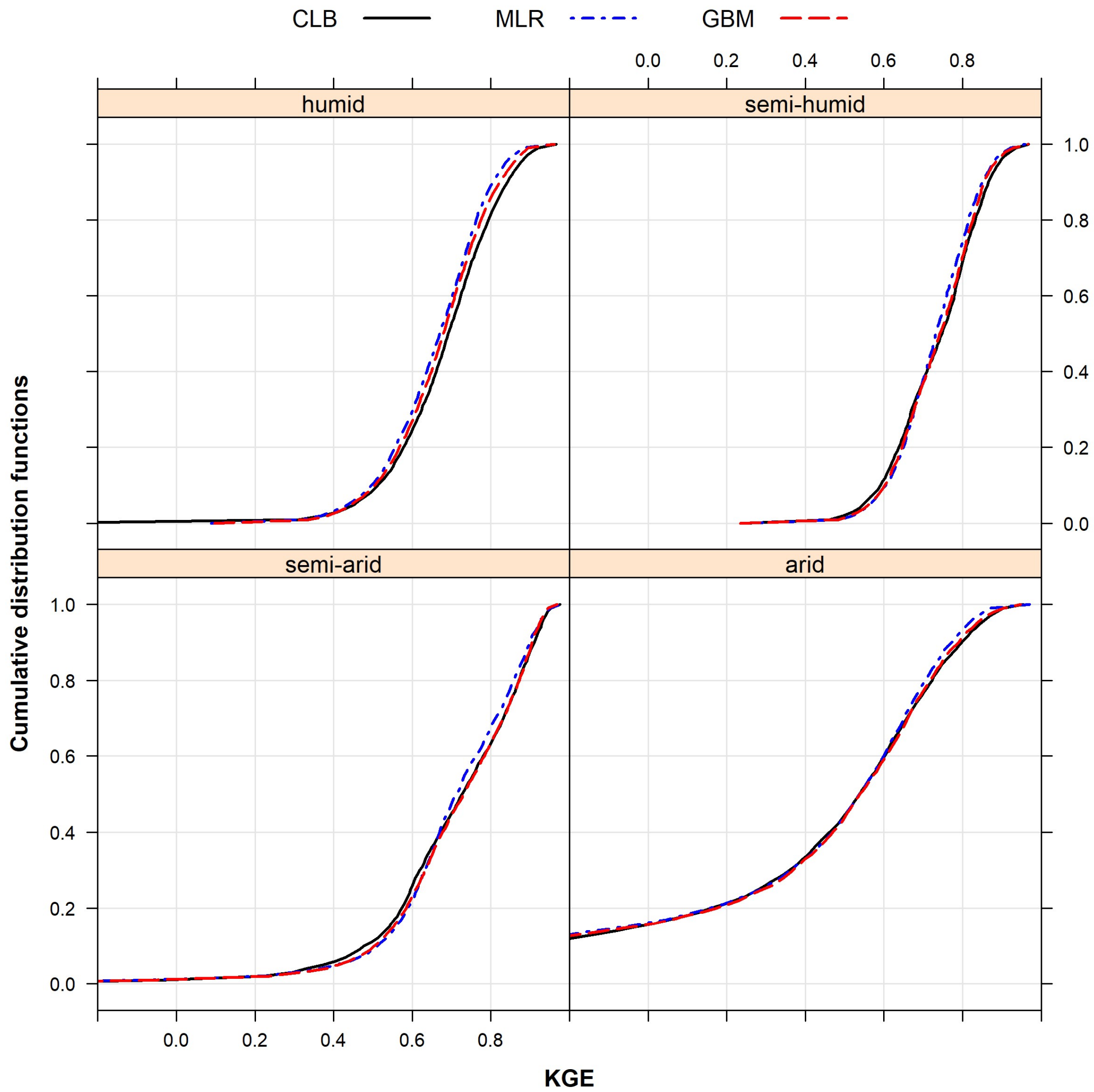


Figure S5: Taylor skill scores of each parameter generated from the MLR and the GBM. The Taylor skill scores are computed using parameters from total grid cells across China.



100 **Figure S6: Cumulative distribution functions (CDFs) of KGE for ET simulation based on three parameter sets (black lines for CLB, blue lines for MLR, red lines for GBM) in the validation period over four climatic zones.**

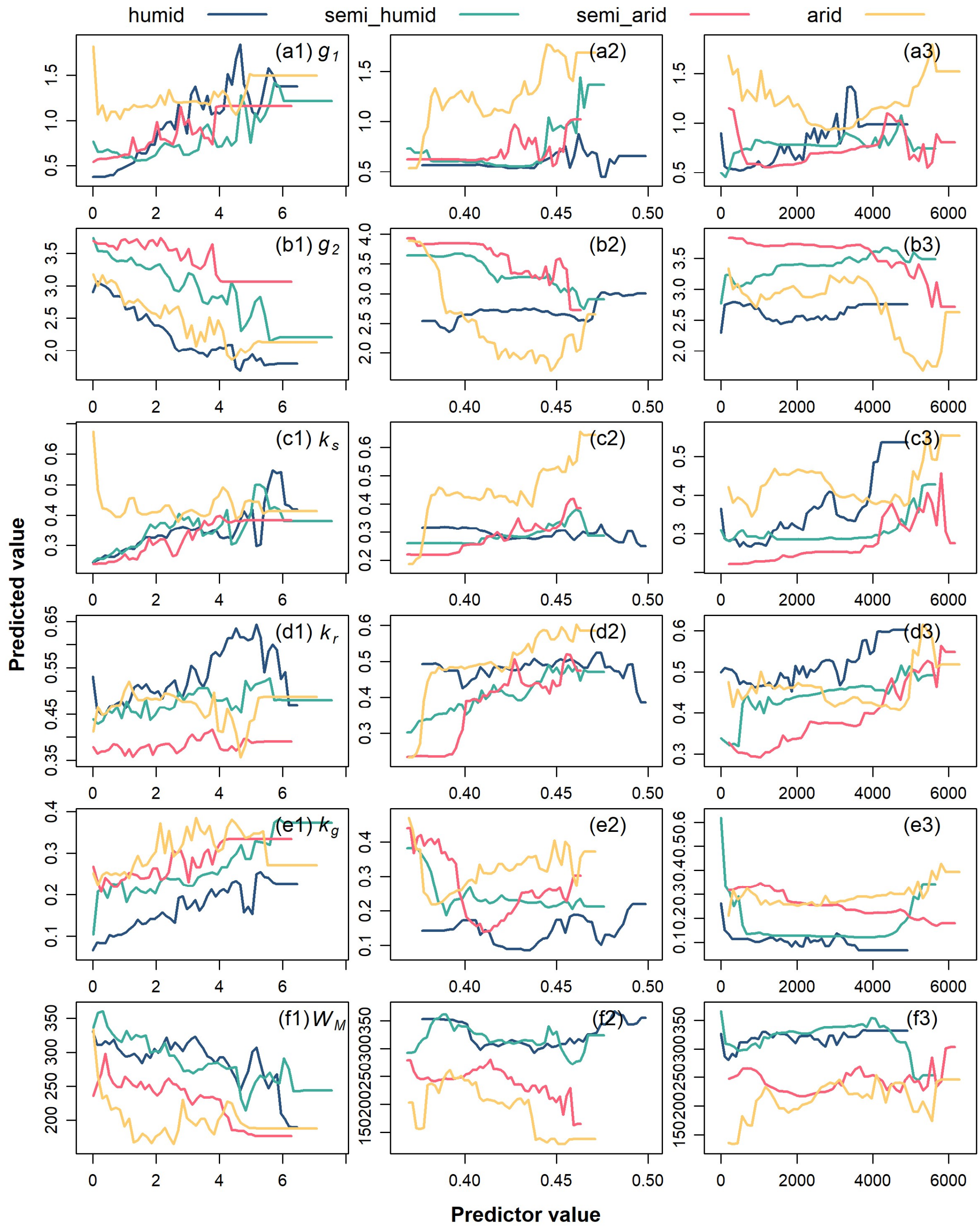


Figure S7: Partial dependence of parameter (a) g_1, (b) g_2, (c) k_s, (d) k_r, (e) k_g, (f) W_M, on (1) slope (slp), (2) saturated moisture content (ths), and (3) elevation in four climatic zones.

105

Reference

Bai, P., Liu, X., Zhang, Y., and Liu, C.: Incorporating vegetation dynamics noticeably improved performance of hydrological model under vegetation greening, Sci. Total Environ., 643, 610-622, 10.1016/j.scitotenv.2018.06.233, 2018

Leuning, R., Zhang, Y., Rajaud, A., Cleugh, H., and Tu, K.: A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation, Water Resour. Res., 44(10), 2008

Seibert, J., and Vis, M.J.P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sc., 16(9), 3315-3325, 10.5194/hess-16-3315-2012, 2012

Xia, J., Wang, G., Tan, G., Ye, A., and Huang, G.H.: Development of distributed time-variant gain model for nonlinear hydrological systems, Science in China Series D: Earth Sciences, 48(6), 2005

Zhang, Y., Chiew, F.H.S., Zhang, L., Leuning, R., and Cleugh, H.A.: Estimating catchment evaporation and runoff using MODIS leaf area index and the Penman-Monteith equation, Water Resour. Res., 44(10), 10.1029/2007WR006563, 2008

Zhang, Y., Kong, D., Gan, R., Chiew, F.H.S., McVicar, T.R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002 – 2017, Remote Sens. Environ., 222, 165-182, 10.1016/j.rse.2018.12.031, 2019

Zhang, Y., Peña-Arancibia, J.L., McVicar, T.R., Chiew, F.H.S., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y.Y., Miralles, D.G., and Pan, M.: Multi-decadal trends in global terrestrial evapotranspiration and its components, Sci Rep-Uk, 6(1), 10.1038/srep19124, 2016