

Supporting Information for:
Comparison of chromatographic stationary
phases using a Bayesian-based multilevel model

Paweł Wiczling^{*1}, Agnieszka Kamedulska²

¹Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gen. J. Hallera 107, 80-416 Gdańsk, Poland

²Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gen. J. Hallera 107, 80-416 Gdańsk, Poland

2023-12-26

Table of contents

1	Model	2
2	Table S1. Summary of the MCMC simulations of the marginal posterior distributions of population-level model parameters.	9
3	Figure S1. Raw data for 6 selected analytes.	13
4	Figure S2. Summary of the MCMC simulations of the marginal posterior distributions of population-level model parameters.	19
5	Figure S3. Goodness of fit plots	23
6	Figure S4. Individual gradient predictions.	27
7	Figure S5. Population gradient predictions.	33
8	Figure S6. Limited data gradient predictions.	39
9	Figure S7. Uncertainty chromatograms.	45

10 Figure S8. Individual izocratic predictions.	48
11 Figure S9. Population isocratic predictions.	54
12 Figure S10. The expected utility maps based on individual predictions	60
13 Figure S11. The expected utility maps based on limited data predictions	65
Licenses	70
Original Computing Environment	70

1 Model

In this work $z = 1, \dots, 51530$ denotes observation, $i=1, \dots, 141$ - analyte, $col=1, \dots, 5$ - column, $m=1, \dots, 2$ - organic modifier and $r=1, \dots, R[i]$ - dissociation step for i -th analyte. The observed retention times ($t_{Robs,z}$) were described using the following model:

$$t_{Robs,z} \sim student_t(\nu, t_{R,z}, \sigma_{col[z],i[z]}),$$

where z denotes z -th observation and $student_t$ denotes the Student's t-distribution with the mean given by the predicted retention time $t_{R,z}$, scale $\sigma_{col,i}$ (analyte and column specific), and normality parameter ν (set to 3).

Gradient retention time $t_{R,z}$ was calculated utilizing the well-known integral equation:

$$\int_0^{t_{R,z} - t_{0,z} - t_e} \frac{dt}{t_{0,z} \cdot ki_z(t)} = 1,$$

where $ki_z(t)$ denotes instantaneous isocratic retention factor corresponding to the mobile phase composition at time t at column inlet for analyte and chromatographic conditions corresponding to the z -th observation, $t_{0,z}$ denotes column hold-up (dead) time and t_e denotes extra column-time. The numerical solution of this integral equation was carried out using method of steps with 4 and 10 steps for methanol and acetonitrile gradients. The following function described the relationship between the isocratic retention factor and pH for an i -th analyte with $R[i]$ dissociation steps and $R[i] + 1$ forms.

$$ki_z(t) = \frac{k_{z,i[z],1}(t) + \sum_{r=1}^{R[i[z]]} k_{z,i[z],r+1}(t) \cdot 10^{r \cdot pH_z(t) - \sum_{r=1}^{R[i]} pKa_{z,i[z],r}(t)}}{1 + \sum_{r=1}^R 10^{r \cdot pH_z(t) - \sum_{r=1}^{R[i[z]]} pKa_{z,i[z],r}(t)}}$$

$\log(k_{z,i,r})$ was assumed to depend on the organic modifier content based on the Neue equation, on temperature assuming linear equation, and on the pH of the mobiles phase (for ionized forms of analytes).

$$\begin{aligned} \log(k_{z,i,r}(t)) = & \log kw_{col[z],i,r} - \frac{S1_{m[z],col[z],i,r} \cdot (1 + S2_{m[z]}) \cdot \varphi_z(t)}{1 + S2_{m[z]} \cdot \varphi_z(t)} \\ & + apH_{col[z],m[z],i,r} \cdot (pH_z(t) - 7) + dlogkT_{col[z],i} \cdot (T_z - 25)/10, \end{aligned}$$

where $\log kw_{col,i,r}$ represents logarithm of retention factors extrapolated to 0% of organic modifier content for column col , i -th analyte, and r -th analyte form; $S1_{i,m,col,r}$ and $S2_m$ are the slopes in the Neue equation for column col , modifier m , i -th analyte, and r -th analyte form. In this parametrization of the Neue equation, $S1$ reflects the difference between logarithm of retention factors corresponding to water (0% of organic modifier content) and MeOH or ACN (100% of organic modifier content) as eluents. $dlogkT_{col,i}$ denotes the change in $\log kw$ due to the increase in temperature by $10^\circ C$. $apH_{col,m,i,r}$ denotes the pH effects on $\log kw$ for ionized forms of analyte.

Further we assume a linear relationship between pKa values and the organic modifier content:

$$pKa_{z,i,r}(t) = pKaw_{i,r} + \alpha_{m[z],i,r} \cdot \varphi_z(t),$$

where $pKa_{z,i,r}(t)$ denotes dissociation constant of an i -th analyte, and r -th dissociation step form and chromatographic conditions corresponding the z -th observation, $pKaw_{i,r}$ denotes aqueous pKa , and $\alpha_{m,i,r}$ denotes the slope for m -th modifier, i -th analyte and r -th form. The linear relationships is generally valid for $\varphi < 0.8$.

The relationship between pH and the organic modifier content for various combinations of organic modifier and buffer was experimentally determined prior to the chromatographic analysis. The obtained relationships was then described using quadratic equations for each nominal pH, temperature and organic modifier:

$$pH_z(t) = pH_{o_z} + \alpha1_z \cdot \varphi_z(t) + \alpha2_z \cdot \varphi_z(t)^2,$$

The individual values of $\log kw$, $S1$ were first defined for the neutral form of an analyte in MeOH for the Xbridge Shield RP18 column (denoted as $\log kw_{N_i}$ and $S1_{mN_i}$). The effect of ACN was described as $dS1_{N_i}$, thus $S1$ in ACN equals

$S1aN_i = S1mN_i + dS1N_i$, the effect of column ($c = 1, \dots, 4$) was described by $clogkwN_{c,i}$, $cS1mN_{c,i}$, and $cdS1N_{c,i}$. The individual parameters for the neutral forms were described using the following equations:

$$\begin{aligned}
 \begin{bmatrix} logkwN_i \\ S1mN_i \end{bmatrix} &\sim MVN \left(\begin{bmatrix} \widehat{logkwN} + \beta_{logkwN} \cdot (logP_i - 2.2) \\ \widehat{S1mN} + \beta_{S1mN} \cdot (logP_i - 2.2) \end{bmatrix}, \Omega \right) \\
 dS1N_i &\sim N(\widehat{dS1N}, \omega_{dS1N}) \\
 \Omega &= diag([\omega_{logkwN}, \omega_{S1mN}]) \cdot \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \cdot diag([\omega_{logkwN}, \omega_{S1mN}]) \\
 \begin{bmatrix} clogkwN_{1,i} \\ clogkwN_{2,i} \\ clogkwN_{3,i} \\ clogkwN_{4,i} \end{bmatrix} &\sim MVN \left(\begin{bmatrix} \widehat{clogkwN}_1 + c\beta_{clogkwN,1} \cdot (logP_i - 2.2) \\ \widehat{clogkwN}_2 + c\beta_{clogkwN,2} \cdot (logP_i - 2.2) \\ \widehat{clogkwN}_3 + c\beta_{clogkwN,3} \cdot (logP_i - 2.2) \\ \widehat{clogkwN}_4 + c\beta_{clogkwN,4} \cdot (logP_i - 2.2) \end{bmatrix}, c\Omega \right) \\
 c\Omega &= diag(c\omega_{clogkwN}) \cdot \begin{bmatrix} 1 & c\rho_{12} & c\rho_{13} & c\rho_{14} \\ c\rho_{21} & 1 & c\rho_{23} & c\rho_{24} \\ c\rho_{31} & c\rho_{32} & 1 & c\rho_{34} \\ c\rho_{41} & c\rho_{42} & c\rho_{43} & 1 \end{bmatrix} \cdot diag(c\omega_{clogkwN}) \\
 cS1mN_{c,i} &\sim N(c\widehat{S1mN}_c + c\beta_{cS1mN,c} \cdot (logP_i - 2.2), c\omega_{cS1mN,c}) \text{ for } c=1, \dots, 4 \\
 cdS1N_{c,i} &\sim N(c\widehat{dS1N}_c, c\omega_{cdS1N,c}) \text{ for } c=1, \dots, 4 \\
 dlogkT_{c,i} &\sim N(\widehat{dlogkT}_c, \omega_{T,c}) \text{ for } c=1, \dots, 4
 \end{aligned}$$

were MVN denotes the multivariate normal distribution; \widehat{logkwN} , $\widehat{S1mN}$, $\widehat{dS1N}$ are the mean values of individual chromatographic parameters that correspond to a typical neutral analyte with $logP = 2.2$ at $25^\circ C$ for Xbridge Shield RP18 stationary phase. β s are regression coefficients between the individual chromatographic parameters and the $logP_i$. ω denotes the standard deviation for between analyte variability (BAV). $dlogkT$ denotes the effect of temperature for a typical analyte and ω_T the standard deviation for between analyte variability for temperature effects. Similar set of equations was used for column effects. Here c denoted the column effect (4 differences with respect to the reference Xbridge Shield RP18 column).

The difference in retention between the ionized and the neutral form of an analyte was separately estimated for acids and bases for $dlogkwA$, $dlogkwB$, $dS1mA$, $dS1mB$, $ddS1A$ and $ddS1B$ parameters. Similar set of equations was used for column effects.

$$\begin{aligned}
 d\logkw{A}_a &\sim N(\widehat{d\logkw{A}}, \kappa_{d\logkw{A}}), \\
 d\logkw{B}_b &\sim N(\widehat{d\logkw{B}}, \kappa_{d\logkw{B}}), \\
 dS1mA_a &\sim N(\widehat{dS1mA}, \kappa_{dS1mA}), \\
 dS1mB_b &\sim N(\widehat{dS1mB}, \kappa_{dS1mB}), \\
 ddS1A_a &\sim N(\widehat{ddS1A}, \kappa_{ddS1A}), \\
 ddS1B_b &\sim N(\widehat{ddS1B}, \kappa_{ddS1B}), \\
 cd\logkw{A}_{c,a} &\sim N(\widehat{cd\logkw{A}_c}, c\kappa_{cd\logkw{c}}) \text{ for } c=1,\dots,4, \\
 cd\logkw{B}_{c,b} &\sim N(\widehat{cd\logkw{B}_c}, c\kappa_{cd\logkw{c}}) \text{ for } c=1,\dots,4, \\
 cdS1mA_{c,a} &\sim N(\widehat{cdS1mA_c}, c\kappa_{cdS1m,c}) \text{ for } c=1,\dots,4, \\
 cdS1mB_{c,b} &\sim N(\widehat{cdS1mB_c}, c\kappa_{cdS1m,c}) \text{ for } c=1,\dots,4, \\
 cddS1A_{c,a} &\sim N(\widehat{cddS1A_c}, c\kappa_{cddS1,c}) \text{ for } c=1,\dots,4, \\
 cddS1B_{c,b} &\sim N(\widehat{cddS1B_c}, c\kappa_{cddS1,c}) \text{ for } c=1,\dots,4.
 \end{aligned}$$

where $a=1\dots46$ and $b=1\dots120$ denote the indexes of acidic and basic groups.

Similarly pKa and α parameters were described separately for acids and bases:

$$\begin{aligned}
 pKaw{A}_a &\sim N(\widehat{pKaw{A}_a}, \tau_{pKaw{A}}), \\
 pKaw{B}_b &\sim N(\widehat{pKaw{B}_b}, \tau_{pKaw{B}}), \\
 \alpha{m}{A}_a &\sim N(\widehat{\alpha{m}{A}}, \tau_{\alpha{m}{A}}), \\
 \alpha{m}{B}_b &\sim N(\widehat{\alpha{m}{B}}, \tau_{\alpha{m}{B}}), \\
 d\alpha{A}_a &\sim N(\widehat{d\alpha{A}}, \tau_{d\alpha{A}}), \\
 d\alpha{B}_b &\sim N(\widehat{d\alpha{B}}, \tau_{d\alpha{B}}).
 \end{aligned}$$

Further, we created the matrices containing the value of $\logkw{S1}, \logkw{S2}, pKaw$, and α for i -th analyte, col -th column, m -th modifier, and r -th dissociation step based on the value of neutral form and effects of column, organic modifier, and dissociation. This transformation was denoted as $f(\cdot)$. The exact procedure can be found in the stan code.

$$\begin{aligned}
logkw_{col,i,r} &= f(logkw_i, clogkw_{c,i}, dlogkw_A_a, cdlogkw_{c,a}, dlogkw_B_b, cdlogkw_{c,b}, \dots) \\
S1_{m,col,i,r} &= f(S1mN_i, cS1mN_{c,i}, dS1mA_a, cdS1mA_{c,a}, S1mB_b, cdS1mB_{c,b}, \dots) \\
dS1N_i, cdS1N_{c,i}, ddS1A_a, cddS1A_{c,a}, ddS1B_b, cddS1B_{c,b}, \dots) \\
apH_{m,col,i,r} &= f(\widehat{apH}_1, \widehat{capH}_{c,1}, \widehat{apH}_2, \widehat{capH}_{c,2}, \dots) \\
S2_m &= 10^{f(\log S2_m, d\log S2, \dots)} \\
pKaw_{i,r} &= f(pKawA_a, pKawB_b, \dots) \\
\alpha_{m,i,r} &= f(\alpha mA_a, d\alpha A_a, \alpha mB_b, d\alpha B_b, \dots)
\end{aligned}$$

Three dots represent additional arguments. Residual error model assumes different parameters for each column and analyte:

$$\begin{aligned}
log(\sigma_{col,i}) &= f(log\sigma_i, clog\sigma_{c,i}, \dots) \\
log\sigma_i &\sim N(log(m\sigma), s\sigma) \\
clog\sigma_{c,i} &\sim N(clogm\sigma_c, cs\sigma_c) \text{ for } c=1,\dots,4,
\end{aligned}$$

The detailed description of parameters and used priors is provided in the following table (BAV denotes between analyte variability):

Table 1: Description of model parameters

Name	Namecode	Description	Priors
XBridge Shield RP18 parameters			
\widehat{logkwN}	logkwHat	typical logkw [Neutral]	N(2.2,2)
$\widehat{S1mN}$	S1mHat	effect of MeOH on logkw [Neutral]	N(4, 1)
$\widehat{dS1N}$	dS1Hat	effect of ACN on S1m [Neutral]	N(1, 1)
β_{logkwN}	beta[1]	effect of logP on logkw [Neutral]	N(1, 0.125)
β_{S1mN}	beta[2]	effect of logP on S1m [Neutral]	N(0.5, 0.5)
$\widehat{dlogkwA}$	dlogkwHat[1]	effect of dissociation on logkw [Acids]	N(-1, 0.125)
$\widehat{dlogkwB}$	dlogkwHat[2]	effect of dissociation on logkw [Bases]	N(-1, 0.125)
$\widehat{dS1mA}$	dS1mHat[1]	effect of dissociation on S1m [Acids]	N(0, 0.5)
$\widehat{dS1mB}$	dS1mHat[2]	effect of dissociation on S1m [Bases]	N(0, 0.5)
$\widehat{ddS1A}$	ddS1Hat[1]	effect of dissociation on dS1 [Acids]	N(0, 0.25)
$\widehat{ddS1B}$	ddS1Hat[2]	effect of dissociation on dS1 [Bases]	N(0, 0.25)
$\widehat{apH_A}$	apH[1]	effect of pH on logkw [Acids]	N(0, 0.1)
$\widehat{apH_B}$	apH[2]	effect of pH on logkw [Bases]	N(0, 0.1)
\widehat{dlogkT}	dlogkTHat	effect of temperature on logkw	N(-0.087, 0.022)
ω_{logkwN}	omega[1]	sd of BAV for logkw [Neutral]	N+(0, 2)
ω_{S1mN}	omega[2]	sd of BAV for S1 [Neutral]	N+(0, 2)

Table 1: Description of model parameters (*continued*)

Name	Namecode	Description	Priors
ω_{dS1N}	omega[3]	sd of BAV for dS1 [Neutral]	N+(0, 2)
ρ	rho	correlation logkw vs S1 [Neutral]	LKJCORRN(0.75, 0.125)
ω_T	omegaT	sd of BAV for dlogkT [Neutral]	N+(0, 0.022)
κ_{dlogkw}	kappa[1]	sd of BAV for dlogkw [Acids and Bases]	N+(0, 0.25)
κ_{dS1m}	kappa[2]	sd of BAV for dS1m [Acids and Bases]	N+(0, 0.25)
κ_{ddS1}	kappa[3]	sd of BAV for ddS1 [Acids and Bases]	N+(0, 0.25)
between column differences			
$clogkwN_c$	clogkwHat[c]	effect of column c on logkw [Neutral]	N(0, 1)
$cS1mN_c$	cS1mHat[c]	effect of column c on S1m [Neutral]	N(0, 0.5)
$cdS1N_c$	cdS1Hat[c]	effect of column c on dS1 [Neutral]	N(0, 0.25)
$c\beta_{logkwN,c}$	cbeta[c,1]	effect of column c on beta[1] [Neutral]	N(0, 0.25)
$c\beta_{dS1mN,c}$	cbeta[c,2]	effect of column c on beta[2] [Neutral]	N(0, 0.25)
$cdlogkwA_c$	cdlogkwHat[c,1]	effect of column c on dlogkw [Acids]	N(0, 0.0625)
$cdlogkwB_c$	cdlogkwHat[c,2]	effect of column c on dlogkw [Bases]	N(0, 0.0625)
$cdS1mB_c$	cdS1mHat[c,1]	effect of column c on dS1m [Acids]	N(0, 0.25)
$cdS1mB_c$	cdS1mHat[c,2]	effect of column c on dS1m [Bases]	N(0, 0.25)
$cddS1A_c$	cddS1Hat[c,1]	effect of column c on ddS1 [Acids]	N(0, 0.125)
$cddS1B_c$	cddS1Hat[c,2]	effect of column c on ddS1 [Bases]	N(0, 0.125)
$cdlogkT_c$	cdlogkTHat[c]	effect of column c on dlogkT	N(0, 0.011)
$capHA_c$	capH[c,1]	effect of column c on apH [Acids]	N(0, 0.05)
$capHB_c$	capH[c,2]	effect of column c on apH [Bases]	N(0, 0.05)
$c\omega_{logkwN,c}$	comega[c,1]	sd of BAV for clogkw [Neutral]	N+(0, 1)
$c\rho$	crho	corr between clogkw [Neutral]	LKJ(2)
$c\omega_{S1mN,c}$	comega[c,2]	sd of BAV for cS1 [Neutral]	N+(0, 1)
$c\omega_{dS1N,c}$	comega[c,3]	sd of BAV for cdS1 [Neutral]	N+(0, 1)
$c\kappa_{dlogkw,c}$	ckappa[c,1]	sd of BAV for cdlogkw [Acids and Bases]	N+(0, 0.125)
$c\kappa_{dS1m,c}$	ckappa[c,2]	sd of BAV for cdS1m [Acids and Bases]	N+(0, 0.125)
$c\kappa_{ddS1,c}$	ckappa[c,3]	sd of BAV for cddS1 [Acids and Bases]	N+(0, 0.125)
$c\omega_{T,c}$	comegaT[c]	sd of BAV for dlogkT	N+(0, 0.011)
S2			
$\logS2m$	logS2mHat	typical value of S2m (log10 scale)	N(-0.7, 0.125);
$dlogS2$	dlogS2Hat	effect of ACN on logS2m	N(1, 0.125);
pKa			
$\alpha m\bar{A}$	alphamHat[1]	effect of MeOH on pKa [Acids]	N(2, 0.25)
$\alpha m\bar{B}$	alphamHat[2]	effect of MeOH on pKa [Bases]	N(-1, 0.25)
$d\alpha\bar{A}$	dalphahat[1]	effect of ACN on alpham [Acids]	N(0, 0.125)
$d\alpha\bar{B}$	dalphahat[2]	effect of ACN on alpham [Bases]	N(0, 0.125)
τ_{pKa}	tau[1]	sd of BAV for pKalit	N+(0, 0.25)
$\tau_{\alpha m}$	tau[2]	sd of BAV for alpham	N+(0, 0.125)

Table 1: Description of model parameters (*continued*)

Name	Namecode	Description	Priors
$\tau_{d\alpha}$	tau[3]	sd of BAV for dalpha	N+(0, 0.125)
Residuals			
$m\sigma$	msigma	typical sd of residuals for XBridge	N+(0,1)
$s\sigma$	ssigma	sd of BAV of residuals for XBridge	N(0,1)
$clogmsigma_c$	clogmsigma[c]	effect of column c on msigma (log scale)	N+(0,0.125)
$c\sigma_c$	cssigma[c]	sd of BAV of residuals for column c	N+(0,0.125)

2 TABLE S1. SUMMARY OF THE MCMC SIMULATIONS OF THE MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL MODEL PARAMETERS.

2 Table S1. Summary of the MCMC simulations of the marginal posterior distributions of population-level model parameters.

Mean denotes sample mean, MCSE denotes Monte Carlo Standard Error, Std-Dev denotes sample standard deviation, 5%, 50%, 95% denote corresponding quantiles, N_Eff denotes effective sample size, R_Hat denotes a measure of chain equilibrium must be within 0.05 of 1.0.

variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
logkwHat	3.60	3.60	0.08	0.08	3.47	3.72	1.00	5947	3030
S1mHat	4.92	4.92	0.08	0.08	4.79	5.05	1.00	5251	3063
dS1Hat	0.61	0.61	0.05	0.05	0.53	0.69	1.00	3259	2365
dlogkwHat[1]	-0.79	-0.79	0.07	0.07	-0.91	-0.67	1.00	7672	3143
dlogkwHat[2]	-0.97	-0.97	0.05	0.05	-1.05	-0.88	1.00	5300	2998
dS1mHat[1]	0.17	0.17	0.12	0.12	-0.03	0.36	1.00	4231	3375
dS1mHat[2]	0.12	0.11	0.07	0.07	0.00	0.24	1.00	2482	2783
ddS1Hat[1]	0.28	0.28	0.08	0.08	0.15	0.41	1.00	6200	3217
ddS1Hat[2]	-0.67	-0.67	0.05	0.06	-0.76	-0.58	1.00	6993	3444
logS2mHat	-0.31	-0.31	0.01	0.01	-0.33	-0.28	1.02	432	986
dlogS2Hat	0.42	0.42	0.01	0.01	0.41	0.43	1.01	552	1093
beta[1]	0.84	0.84	0.04	0.04	0.77	0.91	1.00	7278	3170
beta[2]	0.51	0.51	0.05	0.05	0.43	0.58	1.00	3976	2814
dlogkTHat	-0.09	-0.09	0.00	0.00	-0.09	-0.08	1.00	4047	2966
apH[1]	-0.03	-0.03	0.00	0.00	-0.03	-0.03	1.00	4583	3724
apH[2]	0.08	0.08	0.00	0.00	0.08	0.08	1.00	3114	3288
omega[1]	0.92	0.92	0.06	0.06	0.83	1.02	1.00	5132	2669
omega[2]	0.93	0.93	0.06	0.06	0.84	1.03	1.00	4993	3096
omega[3]	0.55	0.55	0.03	0.03	0.50	0.61	1.00	8536	3214
omegaT	0.03	0.03	0.00	0.00	0.03	0.04	1.00	7750	3273
kappa[1]	0.59	0.58	0.03	0.03	0.53	0.65	1.00	4404	3301
kappa[2]	0.69	0.69	0.05	0.05	0.62	0.77	1.00	3416	3502
kappa[3]	0.55	0.55	0.04	0.04	0.49	0.61	1.00	3604	3242
rho[2,1]	0.87	0.87	0.02	0.02	0.83	0.91	1.00	3500	3223
msigma	0.39	0.39	0.03	0.03	0.35	0.44	1.00	9144	3009
ssigma	0.81	0.81	0.05	0.05	0.74	0.90	1.00	12125	2859

variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
alphamHat[1]	2.22	2.22	0.15	0.15	1.96	2.47	1.00	1721	2559
alphamHat[2]	-1.35	-1.35	0.10	0.10	-1.51	-1.18	1.01	1282	2346
dalphahat[1]	0.22	0.22	0.10	0.10	0.06	0.38	1.00	2171	2544
dalphahat[2]	-0.20	-0.20	0.07	0.07	-0.32	-0.08	1.01	1514	2412
tau[1]	0.88	0.88	0.05	0.05	0.81	0.97	1.00	6372	3337
tau[2]	0.96	0.96	0.06	0.06	0.87	1.06	1.00	2193	2683

2 TABLE S1. SUMMARY OF THE MCMC SIMULATIONS OF THE MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL MODEL PARAMETERS.

tau[3]	0.79	0.79	0.05	0.05	0.71	0.88	1.00	1967	2870
variable	mean	median	sd	mad	q5	q95	rhats	ess_bulk	ess_tail
clogkwHat[1]	0.43	0.43	0.01	0.01	0.41	0.45	1.01	554	1048
clogkwHat[2]	0.18	0.18	0.01	0.01	0.16	0.21	1.01	734	1437
clogkwHat[3]	0.11	0.11	0.01	0.01	0.09	0.13	1.01	612	1192
clogkwHat[4]	0.18	0.18	0.01	0.01	0.16	0.19	1.01	726	1350
cS1mHat[1]	0.59	0.59	0.02	0.02	0.56	0.62	1.00	2127	2781
cS1mHat[2]	-0.10	-0.10	0.02	0.02	-0.14	-0.06	1.00	2010	2704
cS1mHat[3]	0.37	0.37	0.02	0.02	0.34	0.40	1.00	1308	3004
cS1mHat[4]	0.50	0.50	0.02	0.02	0.47	0.52	1.00	1837	3007
cdS1Hat[1]	0.15	0.15	0.01	0.01	0.13	0.17	1.01	928	1643
cdS1Hat[2]	0.81	0.81	0.03	0.02	0.77	0.85	1.01	379	872
cdS1Hat[3]	0.51	0.51	0.04	0.04	0.44	0.58	1.05	153	462
cdS1Hat[4]	0.04	0.04	0.02	0.02	0.01	0.06	1.00	607	1094
cdlogkwHat[1,1]	0.04	0.04	0.02	0.02	0.02	0.07	1.00	2224	2870
cdlogkwHat[2,1]	-0.05	-0.05	0.02	0.02	-0.08	-0.02	1.00	1549	2160
cdlogkwHat[3,1]	-0.03	-0.03	0.02	0.02	-0.06	0.00	1.00	1668	2740
cdlogkwHat[4,1]	0.04	0.04	0.02	0.02	0.02	0.07	1.00	1319	2124
cdlogkwHat[1,2]	0.00	0.00	0.01	0.01	-0.02	0.02	1.00	1667	2364
cdlogkwHat[2,2]	-0.04	-0.04	0.02	0.02	-0.07	-0.01	1.01	1538	2256
cdlogkwHat[3,2]	-0.06	-0.06	0.01	0.01	-0.08	-0.04	1.01	1175	1869
cdlogkwHat[4,2]	-0.03	-0.03	0.01	0.01	-0.05	-0.01	1.00	1137	2304
cdS1mHat[1,1]	-0.01	-0.01	0.04	0.04	-0.09	0.06	1.00	3059	3209
cdS1mHat[2,1]	-0.19	-0.19	0.06	0.06	-0.29	-0.10	1.00	2963	2774
cdS1mHat[3,1]	-0.49	-0.49	0.05	0.05	-0.56	-0.41	1.00	2231	2921
cdS1mHat[4,1]	-0.23	-0.23	0.04	0.04	-0.30	-0.16	1.00	2450	3229
cdS1mHat[1,2]	0.36	0.36	0.03	0.03	0.31	0.41	1.00	2707	3167
cdS1mHat[2,2]	-0.41	-0.41	0.04	0.04	-0.47	-0.35	1.00	2232	3005
cdS1mHat[3,2]	-0.09	-0.09	0.03	0.03	-0.14	-0.04	1.00	1905	2758
cdS1mHat[4,2]	0.29	0.29	0.03	0.03	0.24	0.34	1.00	2106	2691
cddS1Hat[1,1]	0.05	0.05	0.03	0.03	0.01	0.09	1.00	4704	3379
cddS1Hat[2,1]	0.13	0.13	0.05	0.05	0.06	0.21	1.00	2515	2663
cddS1Hat[3,1]	0.46	0.46	0.09	0.09	0.32	0.61	1.00	847	1433
cddS1Hat[4,1]	0.00	0.00	0.02	0.02	-0.04	0.04	1.00	4056	3302
cddS1Hat[1,2]	0.17	0.17	0.02	0.02	0.14	0.19	1.00	2144	3206
cddS1Hat[2,2]	0.57	0.57	0.03	0.03	0.52	0.62	1.00	1653	2285
cddS1Hat[3,2]	0.50	0.51	0.06	0.06	0.41	0.60	1.03	359	938
cddS1Hat[4,2]	-0.04	-0.04	0.02	0.01	-0.06	-0.01	1.00	3877	3310
cbeta[1,1]	0.00	0.00	0.01	0.01	-0.01	0.01	1.01	886	1478
cbeta[2,1]	-0.02	-0.02	0.01	0.01	-0.03	0.00	1.01	1160	1953
cbeta[3,1]	-0.03	-0.03	0.01	0.01	-0.04	-0.02	1.01	945	1688
cbeta[4,1]	-0.02	-0.02	0.01	0.01	-0.03	-0.01	1.01	1001	1807
cbeta[1,2]	-0.07	-0.07	0.01	0.01	-0.09	-0.05	1.00	2265	2809
cbeta[2,2]	0.10	0.10	0.01	0.01	0.08	0.12	1.00	2647	2838
cbeta[3,2]	-0.03	-0.03	0.01	0.01	-0.04	-0.01	1.00	2162	2837

2 TABLE S1. SUMMARY OF THE MCMC SIMULATIONS OF THE MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL MODEL PARAMETERS.

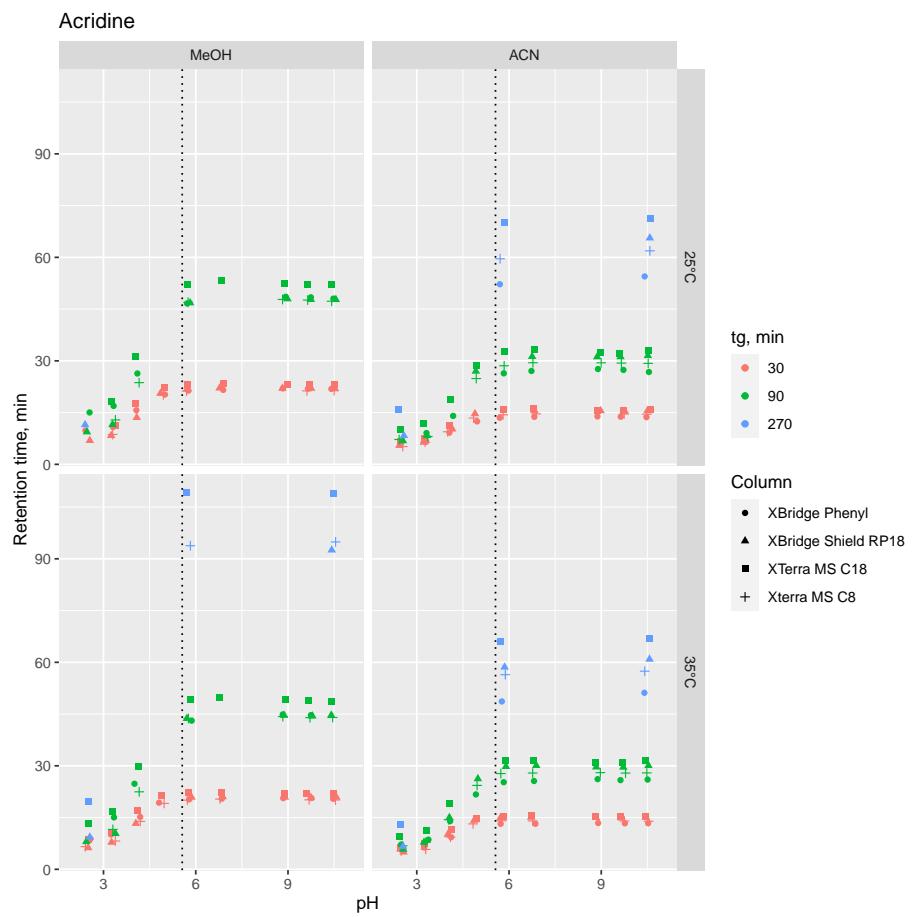
cbeta[4,2]	-0.02	-0.02	0.01	0.01	-0.03	0.00	1.00	2452	2849
cdlogkTHat[1]	-0.01	-0.01	0.00	0.00	-0.01	0.00	1.00	3880	3363
cdlogkTHat[2]	-0.02	-0.02	0.00	0.00	-0.02	-0.02	1.00	4513	3484
cdlogkTHat[3]	-0.01	-0.01	0.00	0.00	-0.01	0.00	1.00	3368	3541
cdlogkTHat[4]	0.00	0.00	0.00	0.00	-0.01	0.00	1.00	4096	3513
capH[1,1]	-0.01	-0.01	0.00	0.00	-0.02	-0.01	1.00	4697	3790
capH[2,1]	-0.02	-0.02	0.00	0.00	-0.02	-0.02	1.00	5050	3925
capH[3,1]	0.00	0.00	0.00	0.00	0.00	0.01	1.00	4877	3408
capH[4,1]	-0.02	-0.02	0.00	0.00	-0.02	-0.02	1.00	4981	3932
capH[1,2]	-0.03	-0.03	0.00	0.00	-0.04	-0.03	1.00	3850	3894
capH[2,2]	-0.04	-0.04	0.00	0.00	-0.05	-0.04	1.00	3629	3707
capH[3,2]	-0.05	-0.05	0.00	0.00	-0.05	-0.05	1.00	3402	3568
capH[4,2]	-0.01	-0.01	0.00	0.00	-0.02	-0.01	1.00	3709	3544
comega[1,1]	0.12	0.12	0.01	0.01	0.10	0.13	1.01	862	1591
comega[2,1]	0.13	0.13	0.01	0.01	0.12	0.15	1.00	1315	2135
comega[3,1]	0.12	0.12	0.01	0.01	0.11	0.13	1.00	1157	2056
comega[4,1]	0.10	0.10	0.01	0.01	0.09	0.11	1.00	1212	1998
comega[1,2]	0.06	0.06	0.02	0.02	0.02	0.09	1.03	315	289
comega[2,2]	0.15	0.15	0.02	0.02	0.12	0.18	1.00	2256	2851
comega[3,2]	0.05	0.05	0.01	0.01	0.03	0.07	1.01	426	343
comega[4,2]	0.02	0.02	0.01	0.01	0.00	0.04	1.00	760	1516
comega[1,3]	0.14	0.14	0.01	0.01	0.13	0.16	1.01	1387	2315
comega[2,3]	0.29	0.29	0.02	0.02	0.26	0.32	1.00	864	1617
comega[3,3]	0.47	0.47	0.03	0.03	0.42	0.52	1.02	701	1289
comega[4,3]	0.16	0.16	0.01	0.01	0.15	0.19	1.00	1393	1907
ckappa[1,1]	0.07	0.07	0.01	0.01	0.06	0.08	1.00	1256	2247
ckappa[2,1]	0.11	0.11	0.01	0.01	0.10	0.13	1.00	1718	2961
ckappa[3,1]	0.08	0.08	0.01	0.01	0.07	0.09	1.01	1764	2733
ckappa[4,1]	0.08	0.08	0.01	0.01	0.07	0.09	1.00	1491	2505
ckappa[1,2]	0.07	0.07	0.03	0.04	0.02	0.13	1.01	417	852
ckappa[2,2]	0.20	0.20	0.03	0.03	0.15	0.25	1.00	1664	2343
ckappa[3,2]	0.11	0.11	0.03	0.03	0.06	0.15	1.01	862	830
ckappa[4,2]	0.04	0.03	0.02	0.03	0.00	0.08	1.03	479	1426
ckappa[1,3]	0.07	0.07	0.02	0.02	0.04	0.11	1.00	865	1522
ckappa[2,3]	0.27	0.27	0.02	0.02	0.24	0.31	1.00	1741	2782
ckappa[3,3]	0.72	0.72	0.04	0.04	0.65	0.78	1.01	1281	2450
ckappa[4,3]	0.09	0.09	0.02	0.02	0.06	0.11	1.01	745	676
comegat[1]	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1656	2300
comegat[2]	0.01	0.01	0.00	0.00	0.01	0.01	1.00	2194	2711
comegat[3]	0.00	0.00	0.00	0.00	0.00	0.00	1.01	1408	2289
comegat[4]	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2595	2298
clogmsigma[1]	0.20	0.20	0.02	0.02	0.17	0.22	1.00	2901	3087
clogmsigma[2]	-0.01	-0.01	0.02	0.02	-0.05	0.02	1.00	3536	3376
clogmsigma[3]	-0.26	-0.26	0.02	0.02	-0.29	-0.22	1.00	3010	3356
clogmsigma[4]	-0.10	-0.10	0.02	0.02	-0.13	-0.07	1.00	3325	3444
cssigma[1]	0.07	0.07	0.03	0.03	0.02	0.11	1.01	574	1063

2 TABLE S1. SUMMARY OF THE MCMC SIMULATIONS OF THE
MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL
MODEL PARAMETERS.

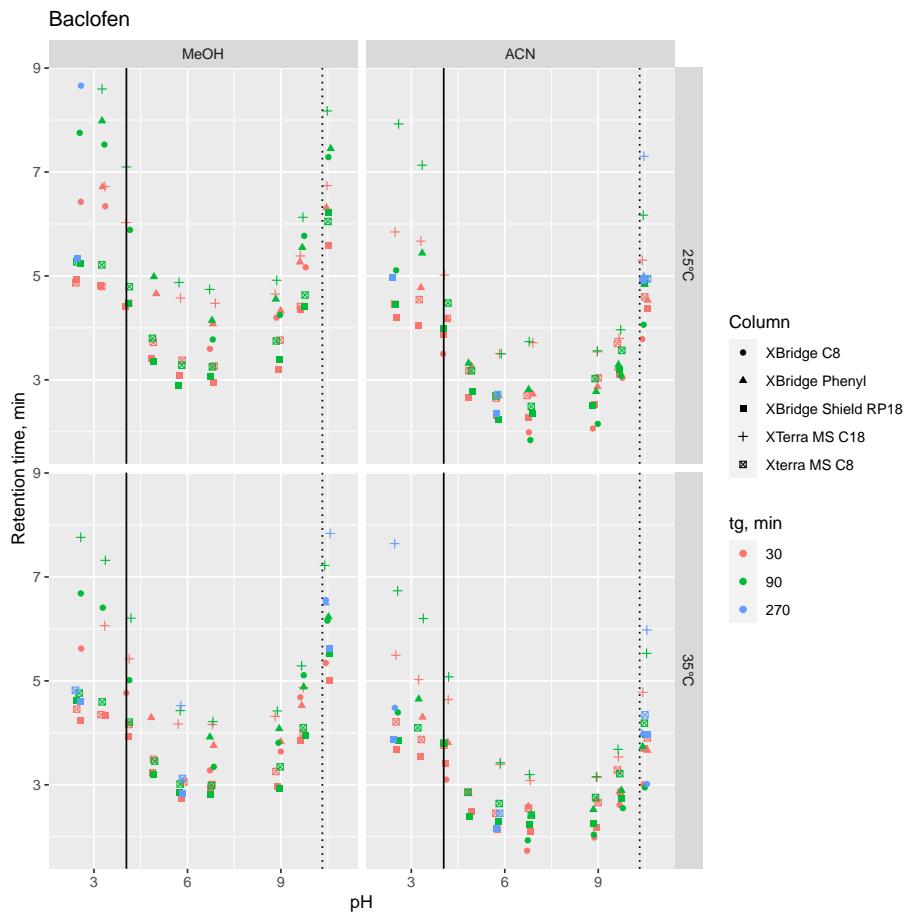
cssigma[2]	0.17	0.17	0.02	0.02	0.14	0.20	1.00	1761	2633
cssigma[3]	0.16	0.16	0.02	0.02	0.13	0.19	1.00	1919	2850
cssigma[4]	0.08	0.08	0.03	0.02	0.03	0.12	1.01	560	769

3 FIGURE S1. RAW DATA FOR 6 SELECTED ANALYTES.

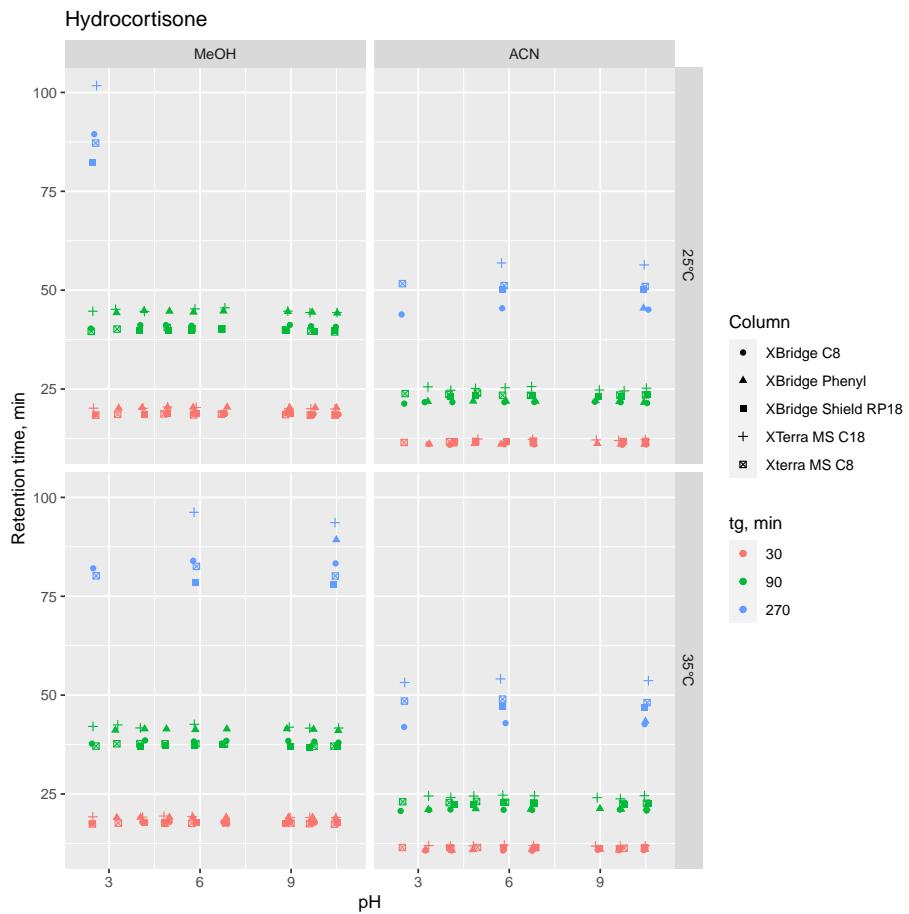
3 Figure S1. Raw data for 6 selected analytes.



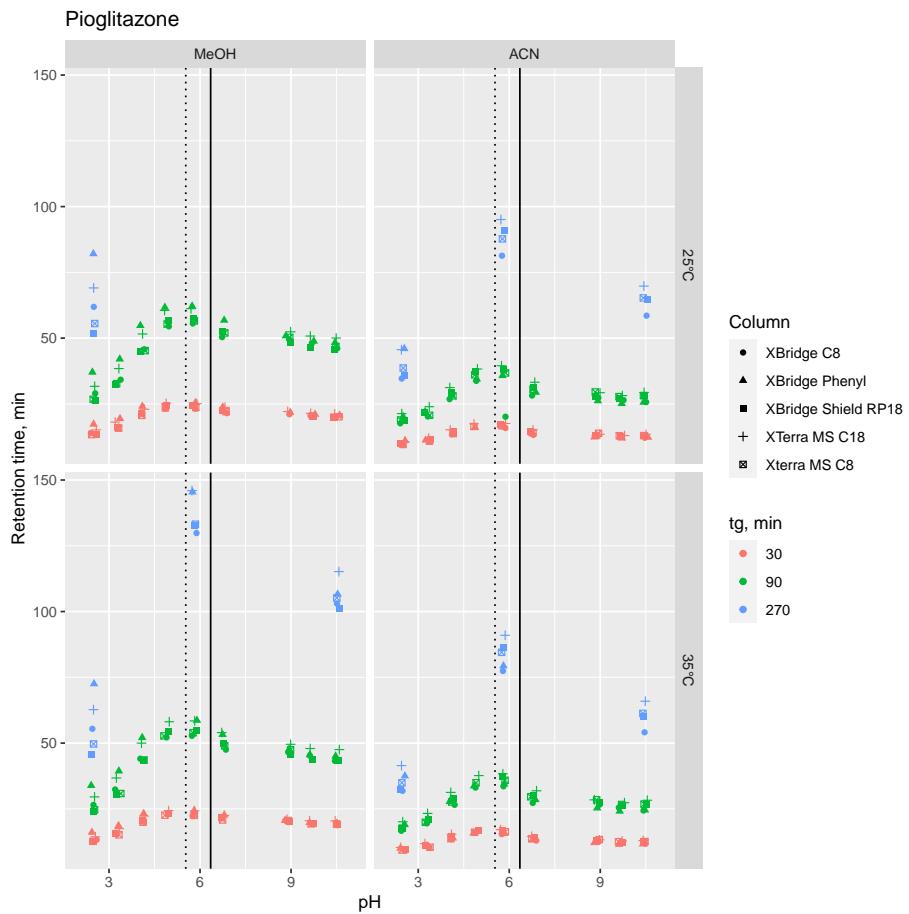
3 FIGURE S1. RAW DATA FOR 6 SELECTED ANALYTES.



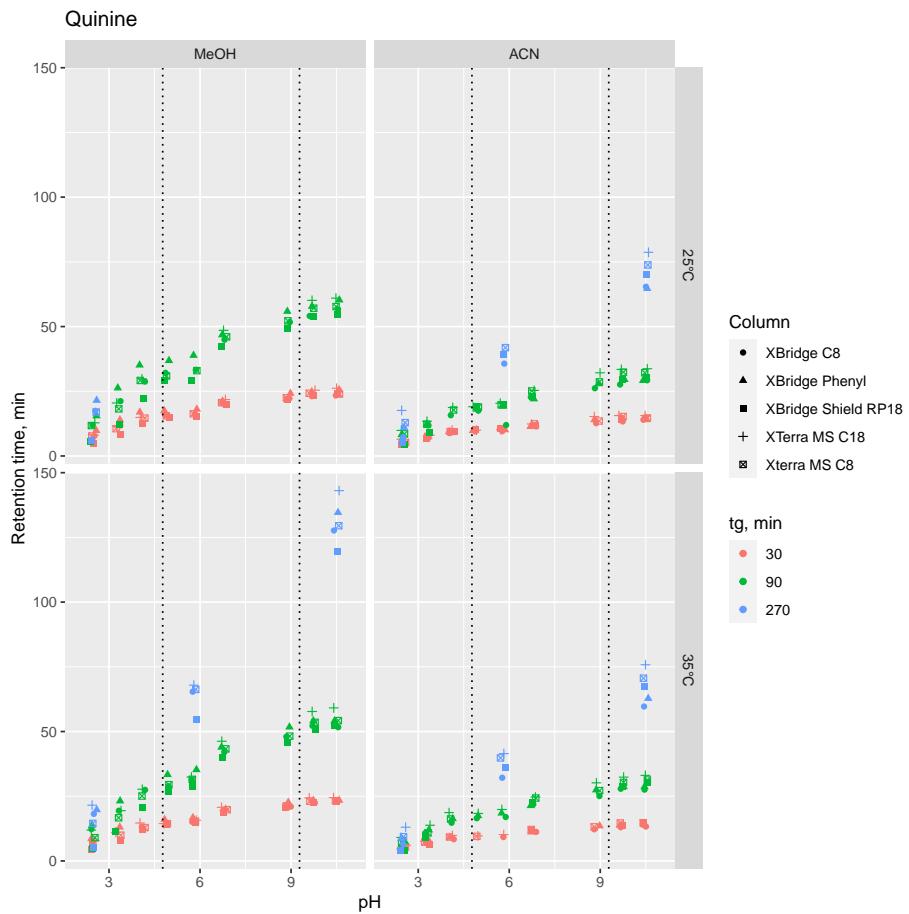
3 FIGURE S1. RAW DATA FOR 6 SELECTED ANALYTES.



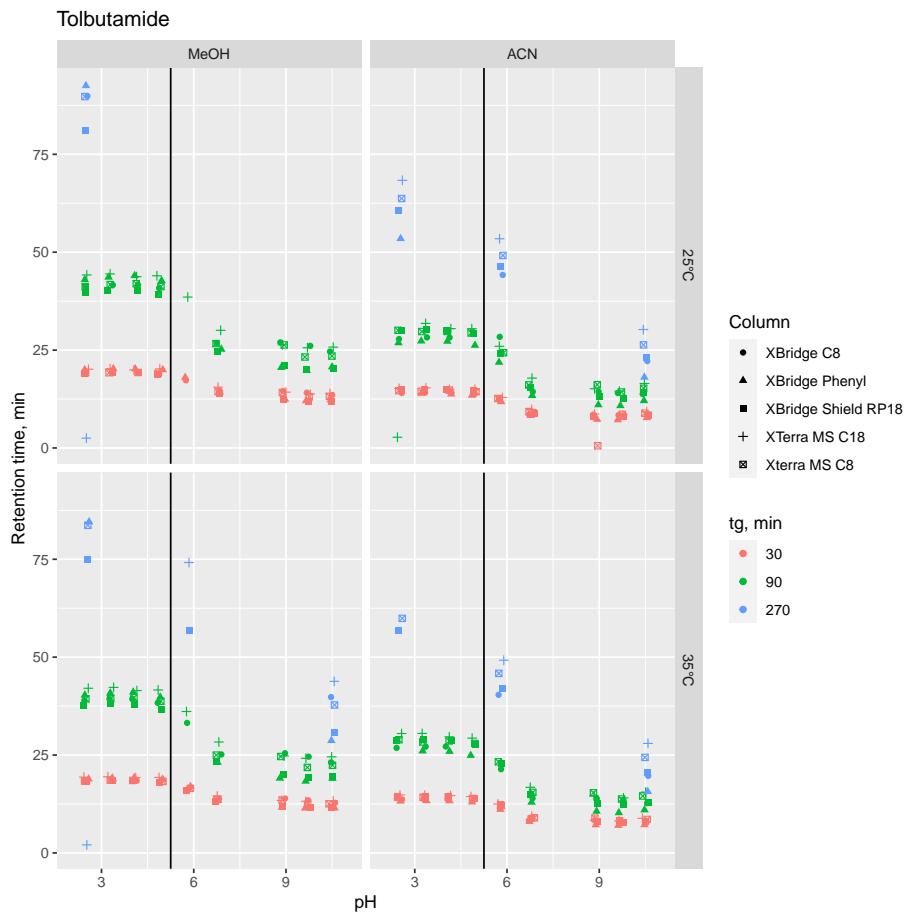
3 FIGURE S1. RAW DATA FOR 6 SELECTED ANALYTES.



3 FIGURE S1. RAW DATA FOR 6 SELECTED ANALYTES.

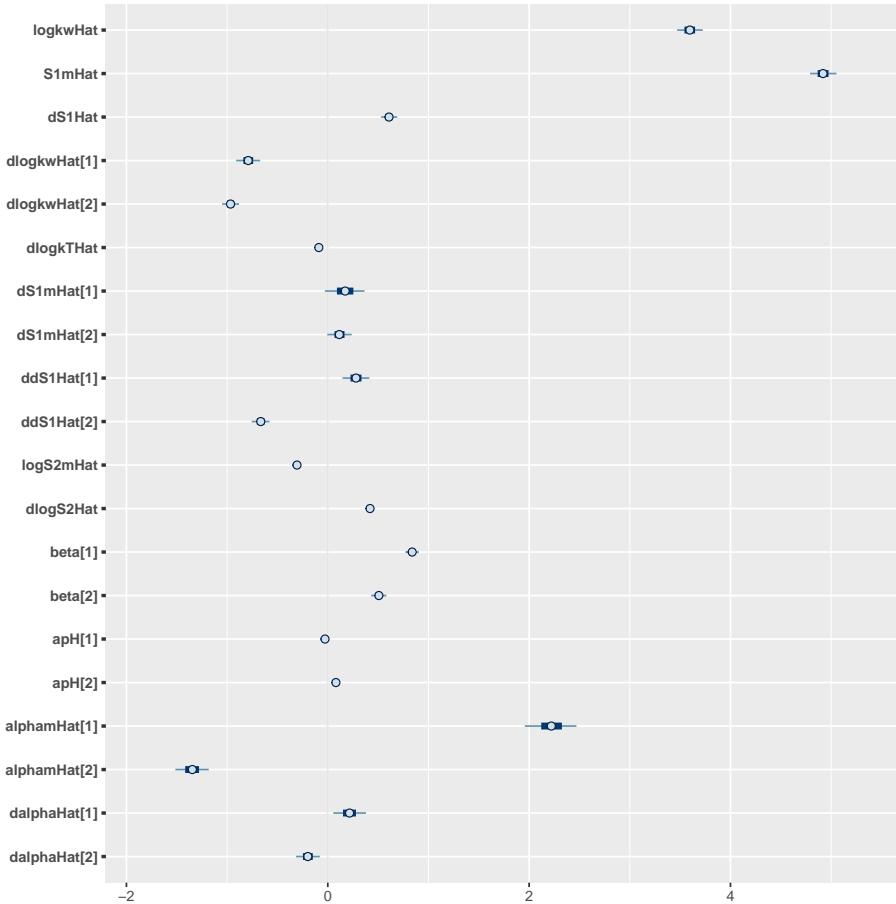


3 FIGURE S1. RAW DATA FOR 6 SELECTED ANALYTES.

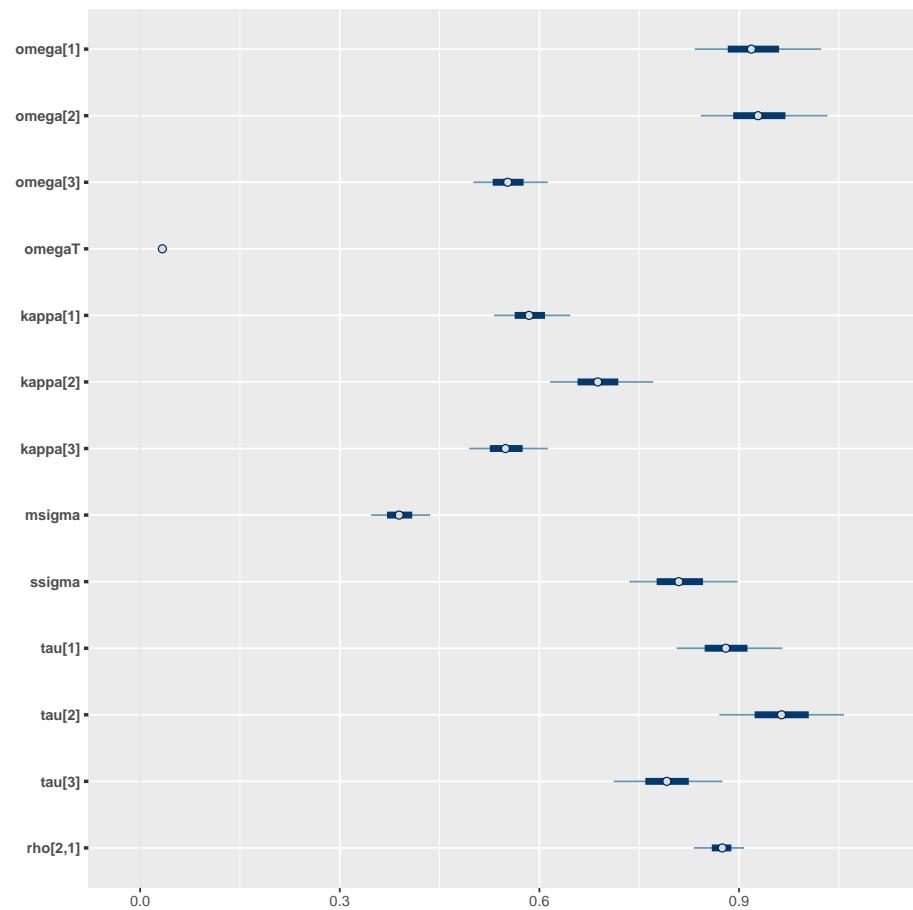


4 FIGURE S2. SUMMARY OF THE MCMC SIMULATIONS OF THE MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL MODEL PARAMETERS.

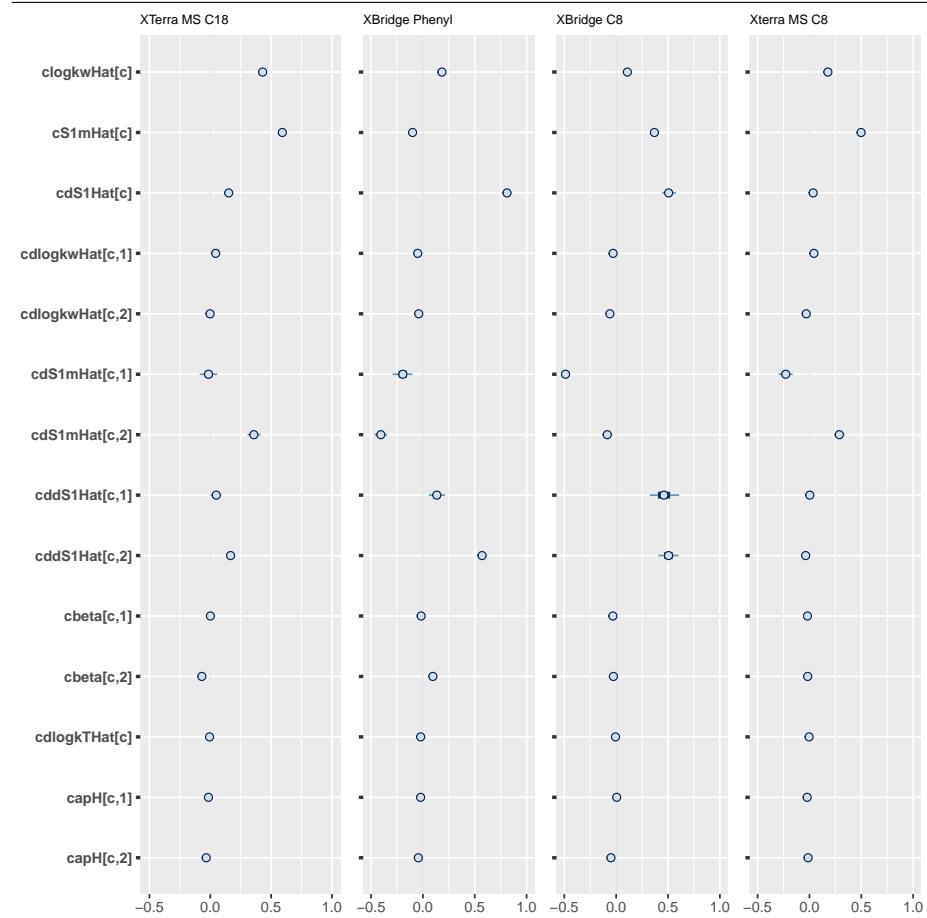
4 Figure S2. Summary of the MCMC simulations of the marginal posterior distributions of population-level model parameters.



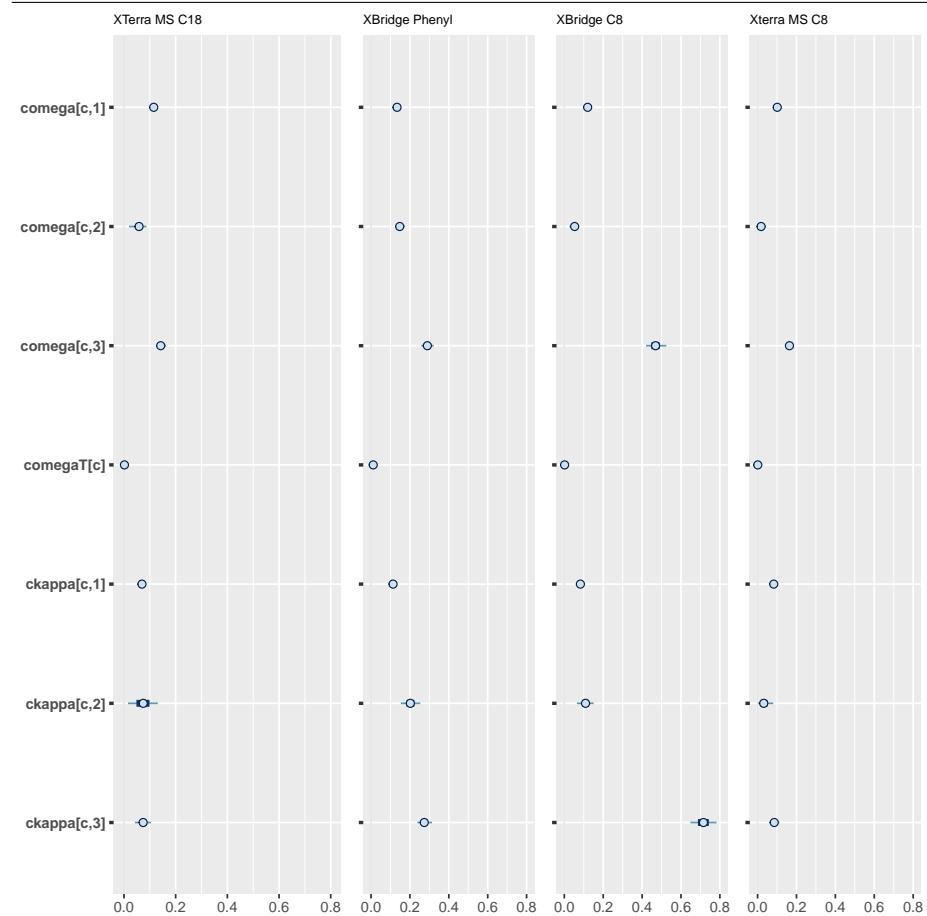
4 FIGURE S2. SUMMARY OF THE MCMC SIMULATIONS OF THE MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL MODEL PARAMETERS.



4 FIGURE S2. SUMMARY OF THE MCMC SIMULATIONS OF THE MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL MODEL PARAMETERS.

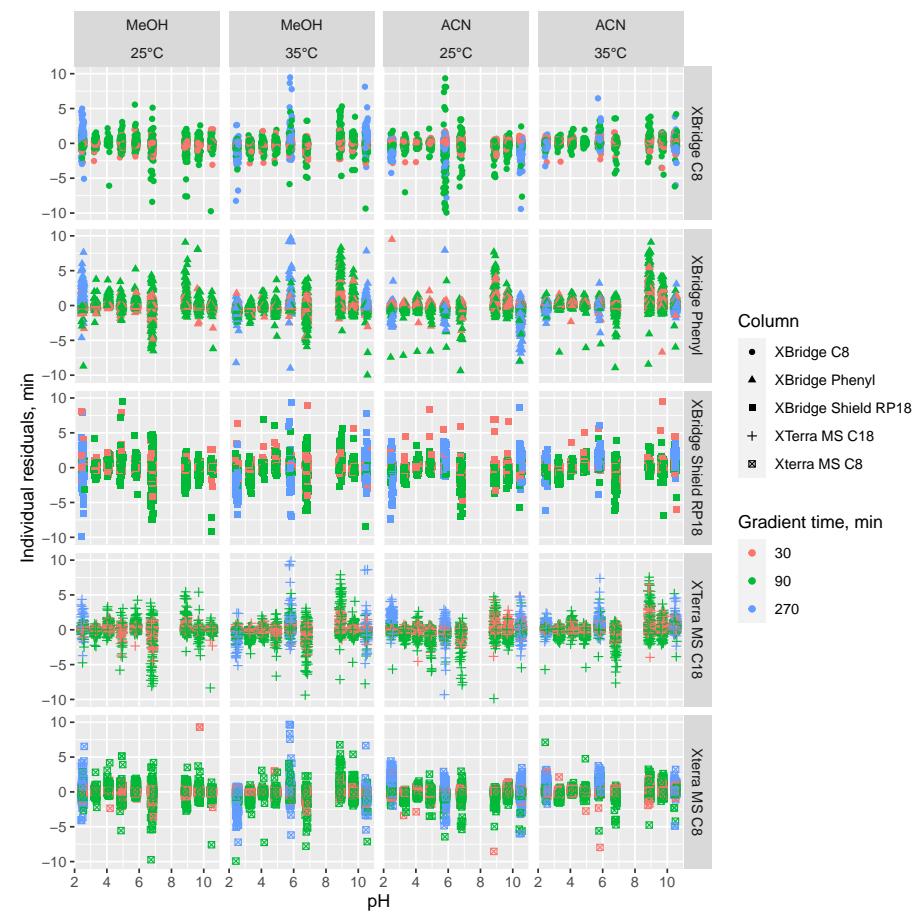


4 FIGURE S2. SUMMARY OF THE MCMC SIMULATIONS OF THE MARGINAL POSTERIOR DISTRIBUTIONS OF POPULATION-LEVEL MODEL PARAMETERS.



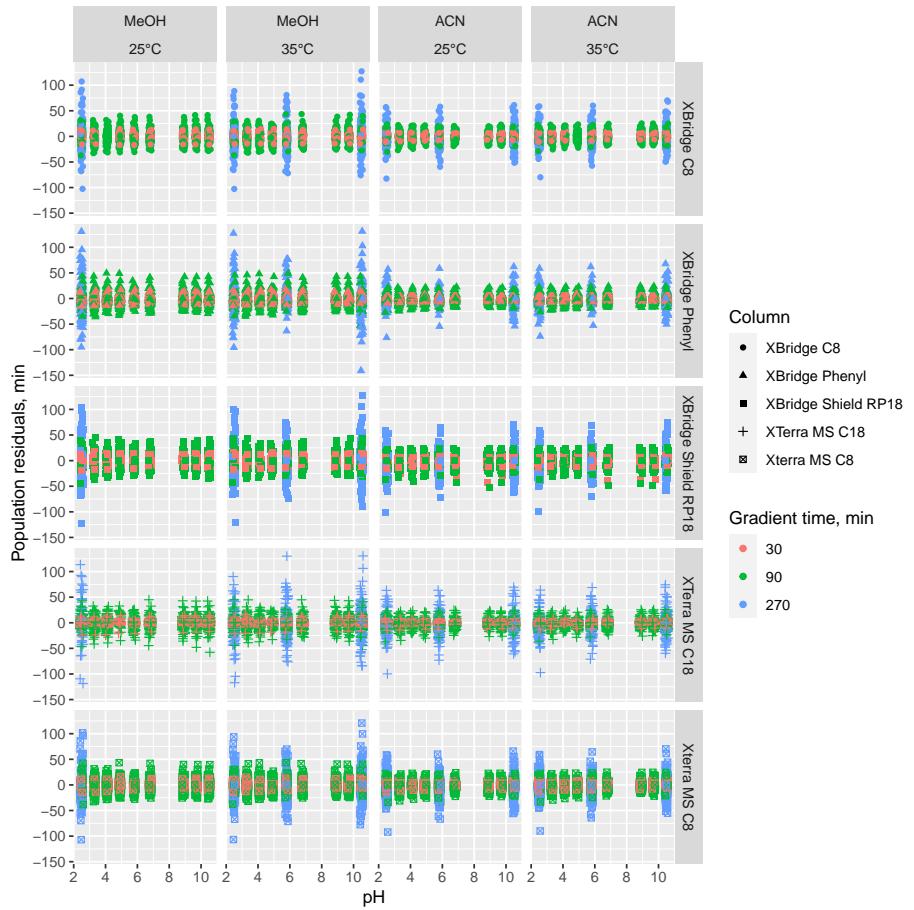
5 Figure S3. Goodness of fit plots

Individual residuals:



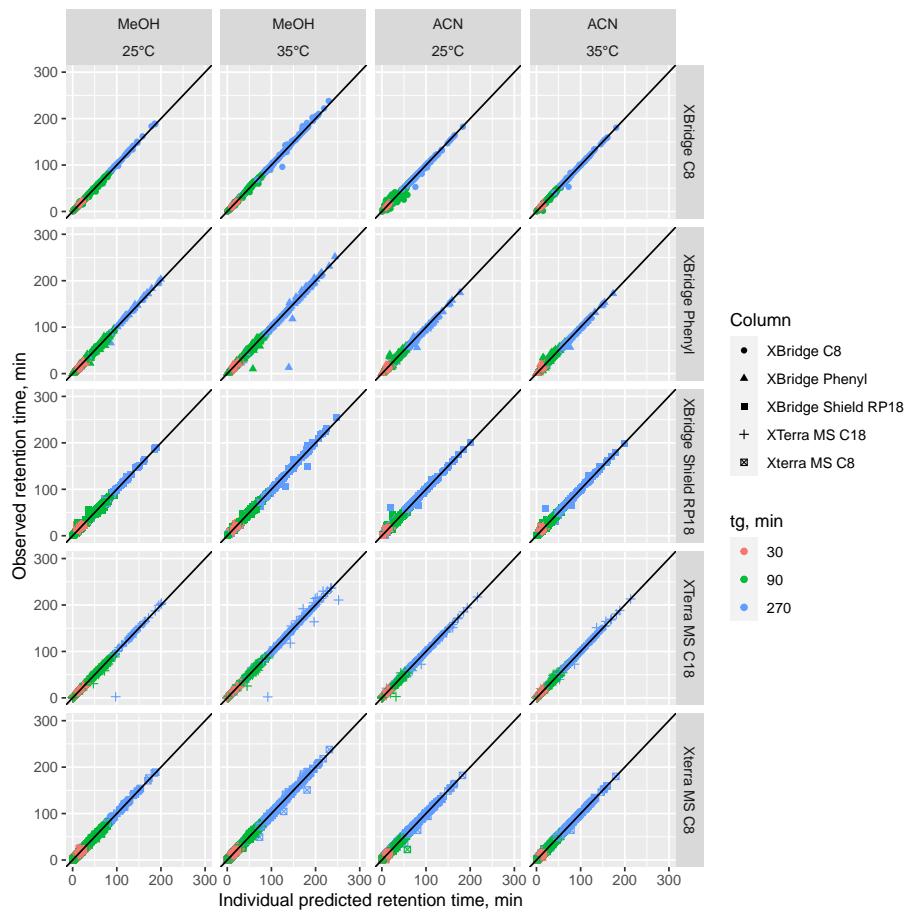
5 FIGURE S3. GOODNESS OF FIT PLOTS

Population residuals:



5 FIGURE S3. GOODNESS OF FIT PLOTS

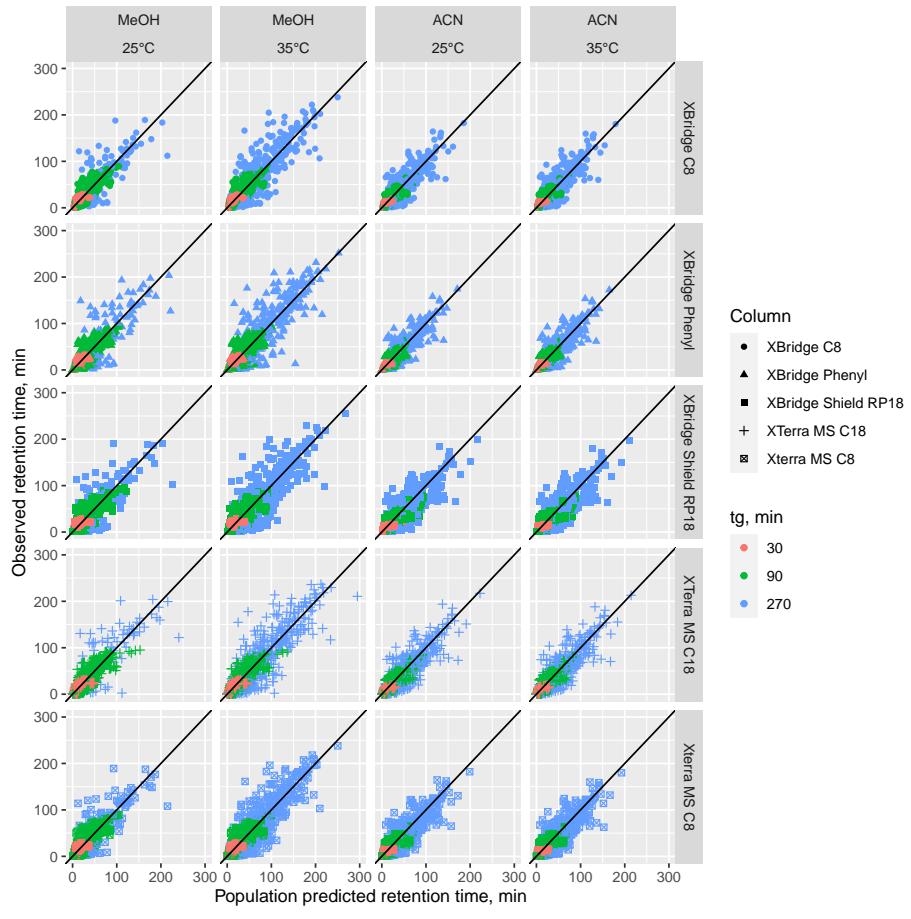
The observed vs. the mean individual-predicted retention times (i.e., a posteriori mean of a predictive distribution conditioned on the observed data from the



same analyte).

5 FIGURE S3. GOODNESS OF FIT PLOTS

The observed vs. the mean population-predicted retention times (i.e., a posteriori means of predictive distributions corresponding to the future observations of

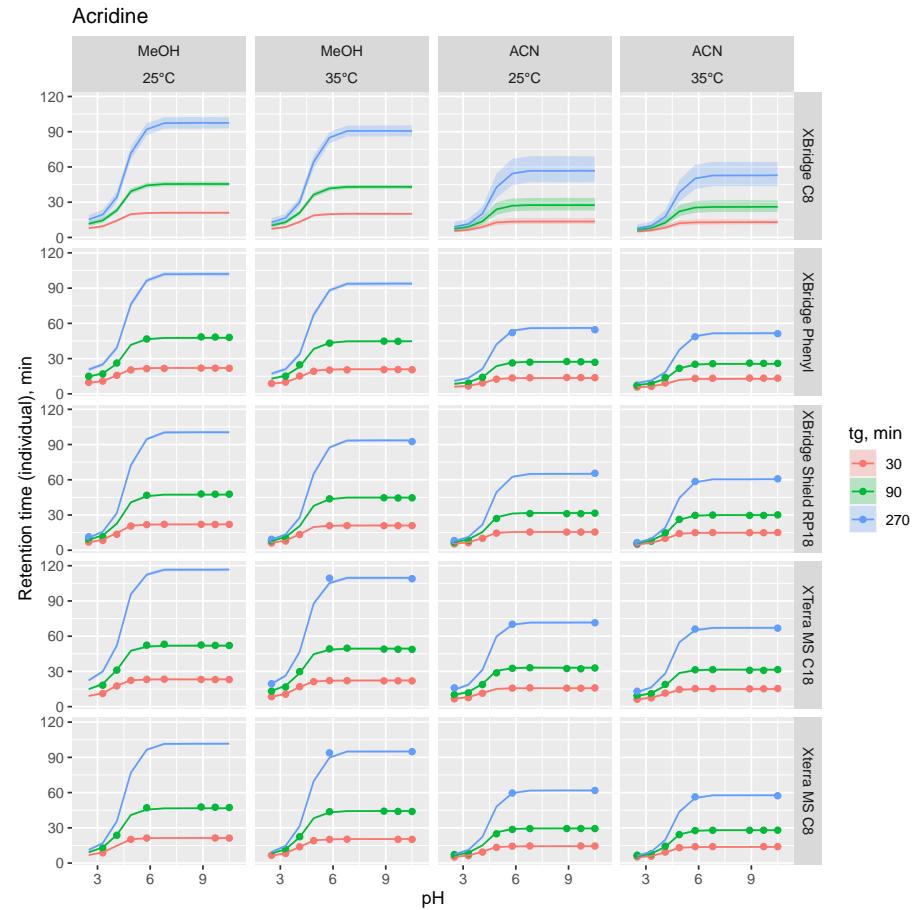


a new analyte)

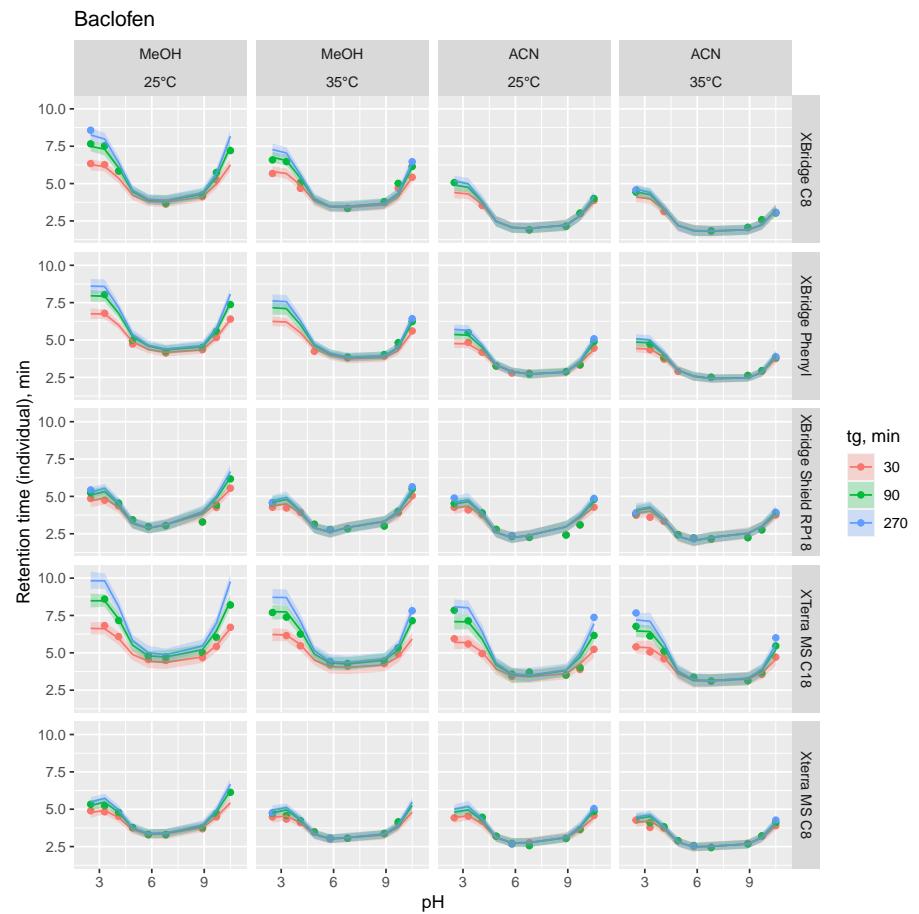
6 FIGURE S4. INDIVIDUAL GRADIENT PREDICTIONS.

6 Figure S4. Individual gradient predictions.

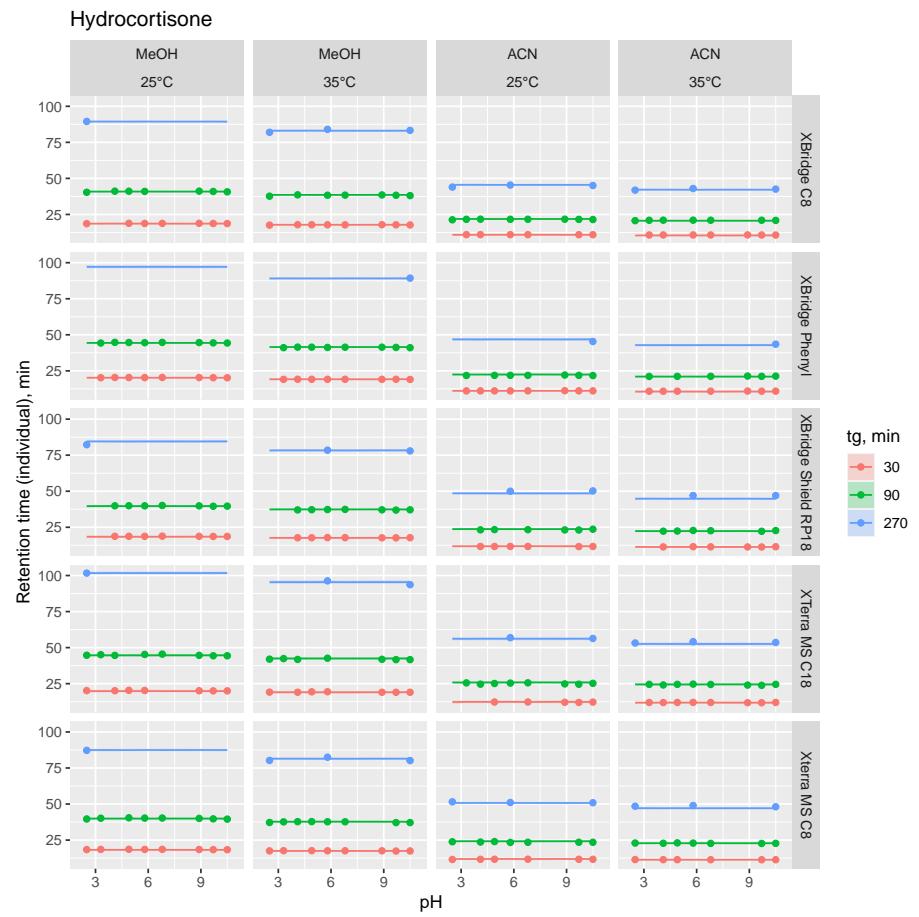
Predictions represented as posterior median (line) and 5th-95th percentiles (areas) for a 6 exemplary analytes. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.



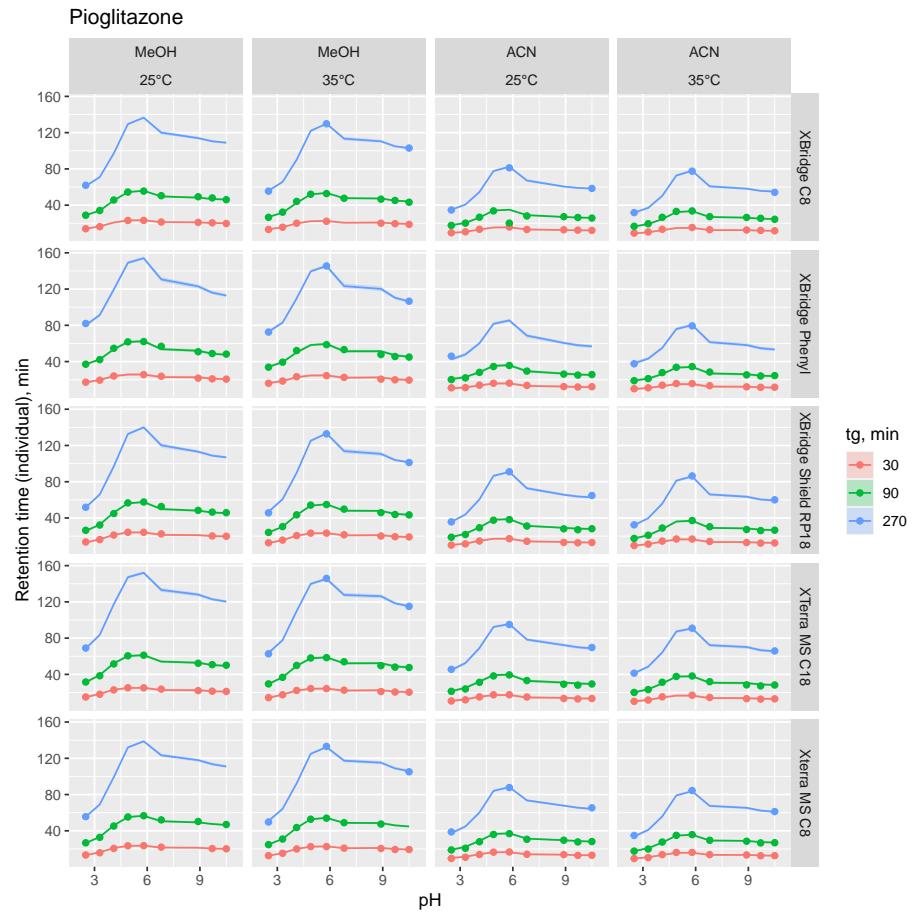
6 FIGURE S4. INDIVIDUAL GRADIENT PREDICTIONS.



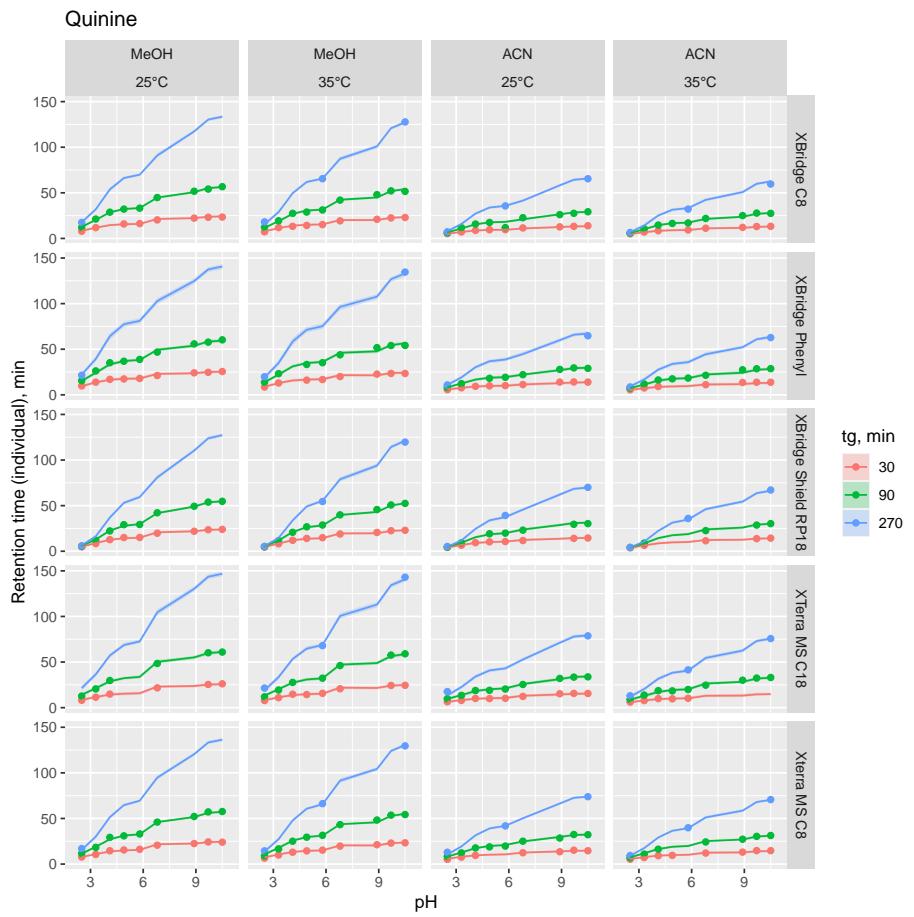
6 FIGURE S4. INDIVIDUAL GRADIENT PREDICTIONS.



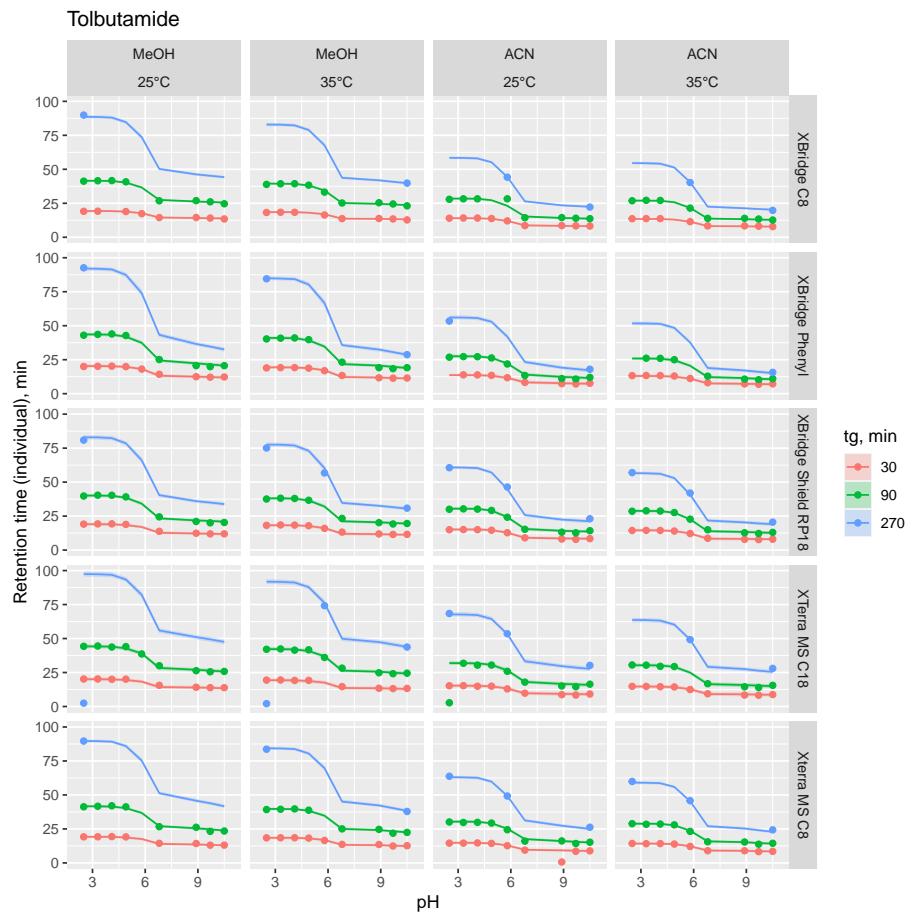
6 FIGURE S4. INDIVIDUAL GRADIENT PREDICTIONS.



6 FIGURE S4. INDIVIDUAL GRADIENT PREDICTIONS.



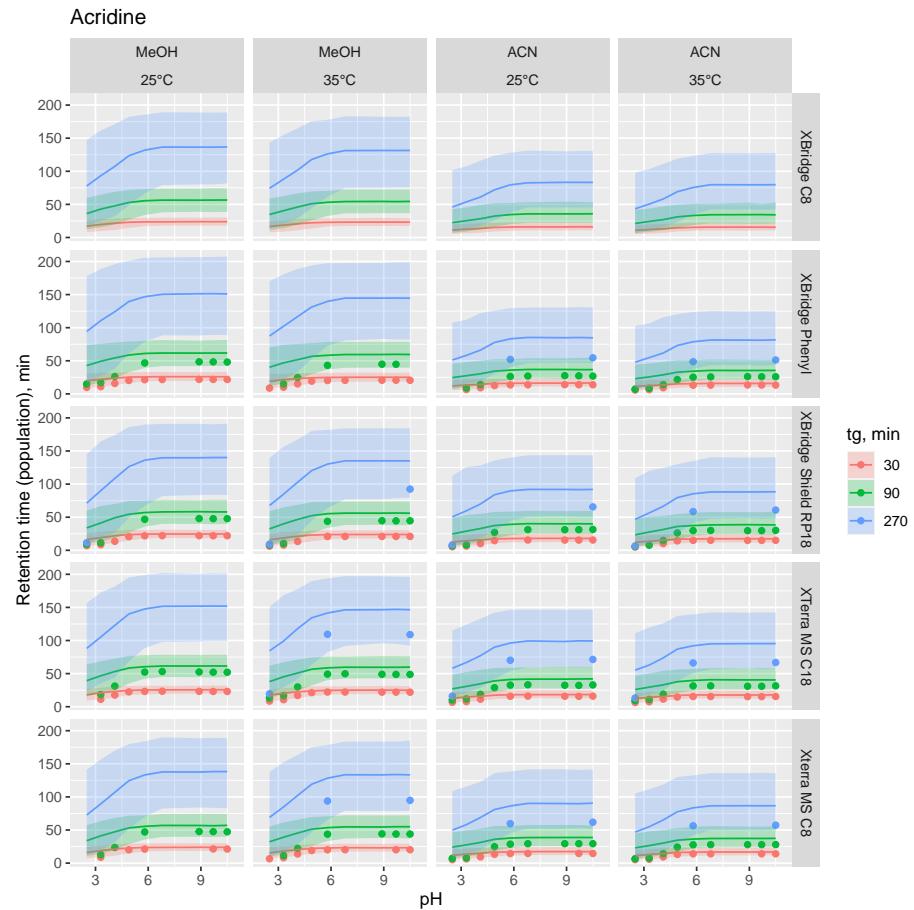
6 FIGURE S4. INDIVIDUAL GRADIENT PREDICTIONS.



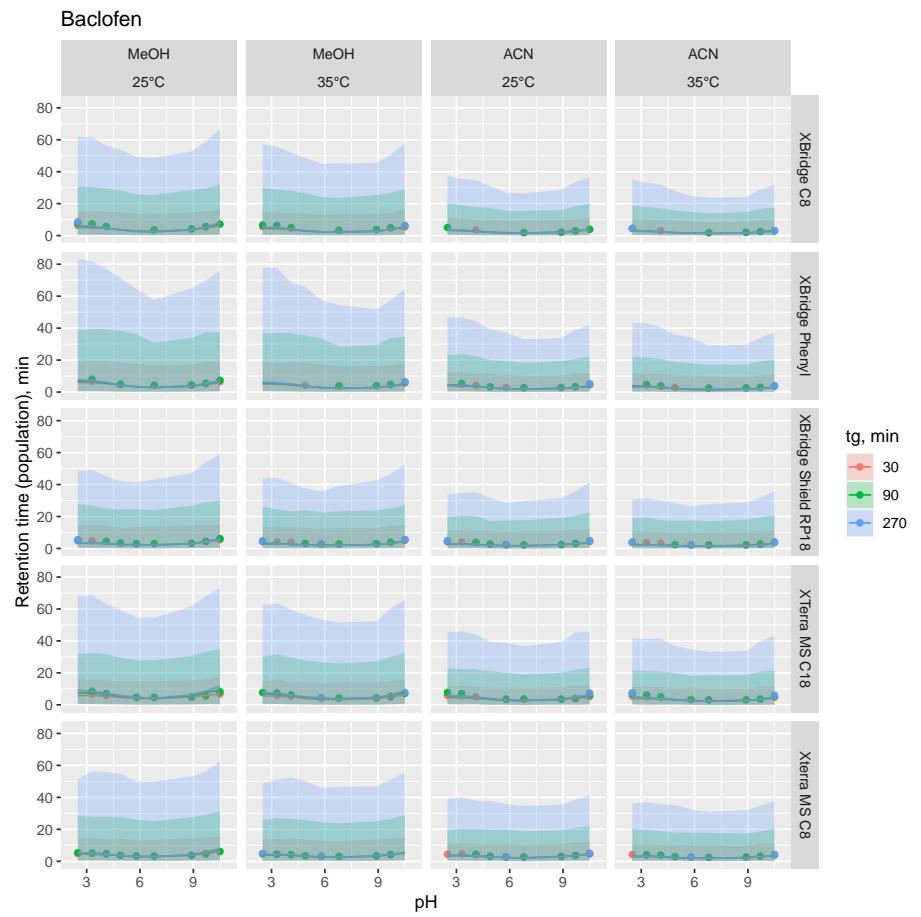
7 FIGURE S5. POPULATION GRADIENT PREDICTIONS.

7 Figure S5. Population gradient predictions.

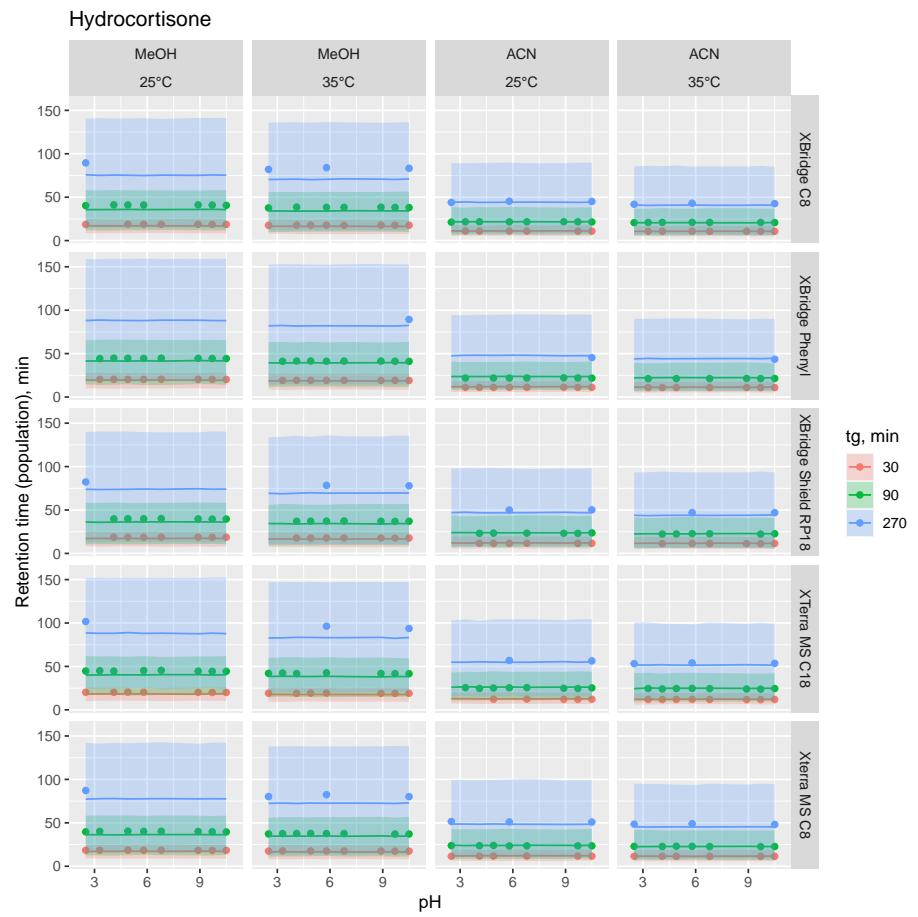
Predictions represented as posterior median (line) and 5th-95th percentiles (areas) for a 6 exemplary analytes. Predictions corresponding to future observations given only population-level parameters and predictors (logP and pKa).



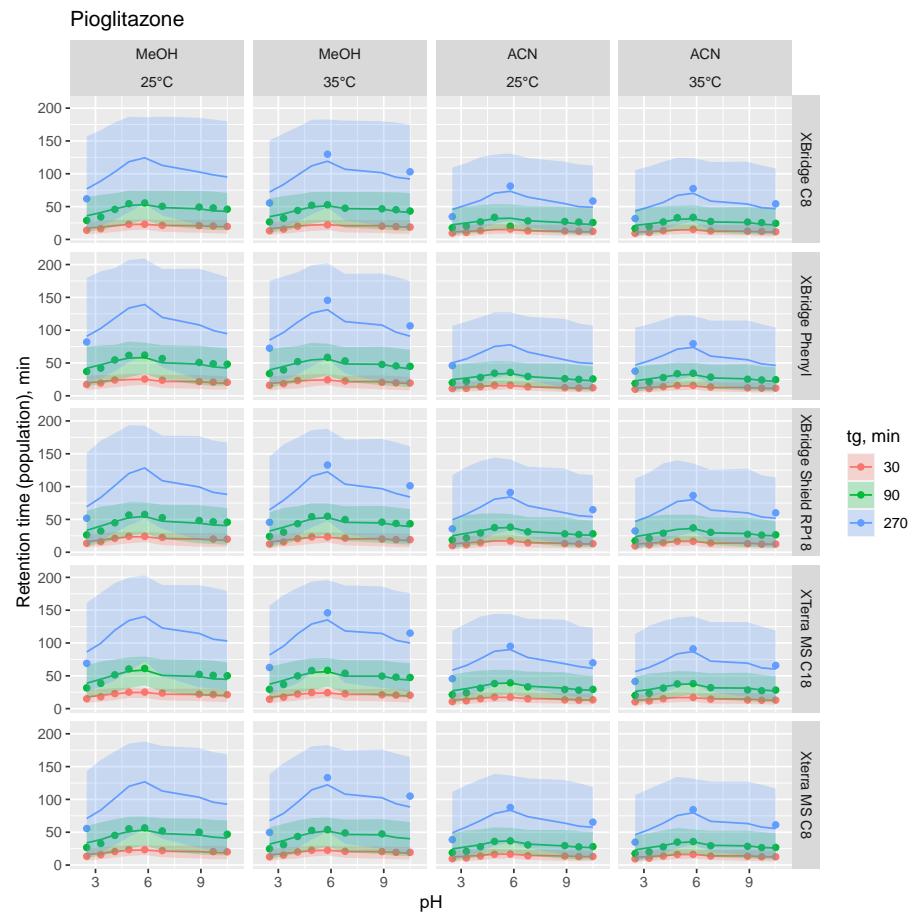
7 FIGURE S5. POPULATION GRADIENT PREDICTIONS.



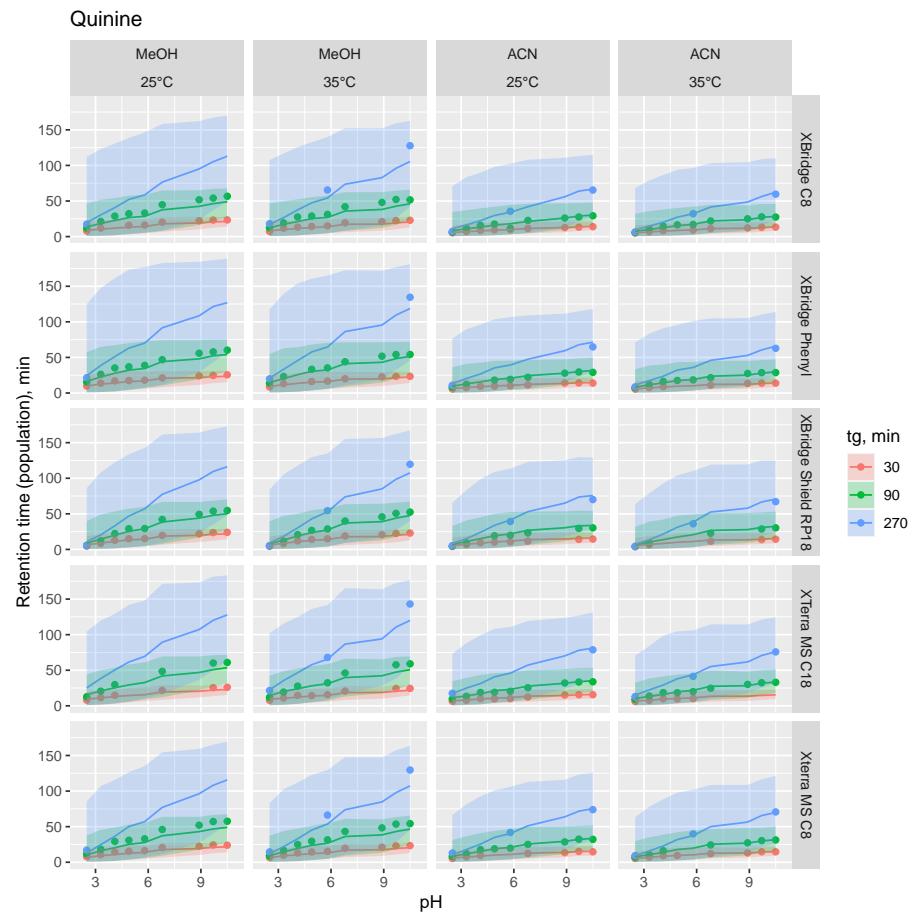
7 FIGURE S5. POPULATION GRADIENT PREDICTIONS.



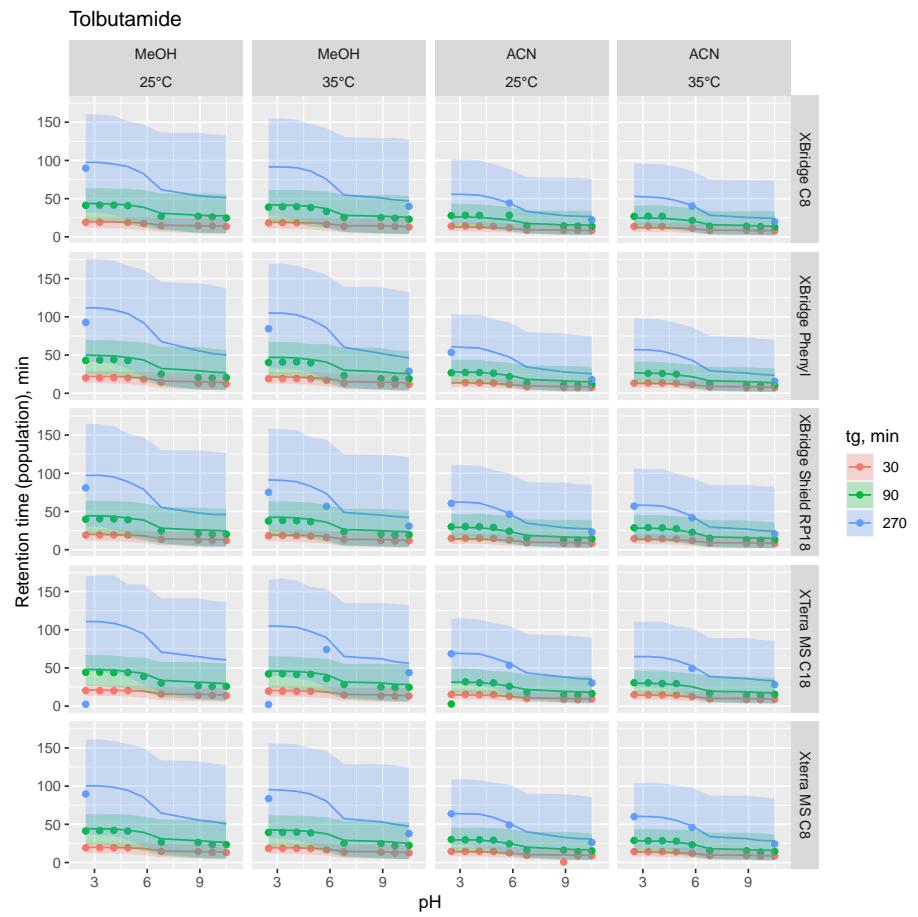
7 FIGURE S5. POPULATION GRADIENT PREDICTIONS.



7 FIGURE S5. POPULATION GRADIENT PREDICTIONS.



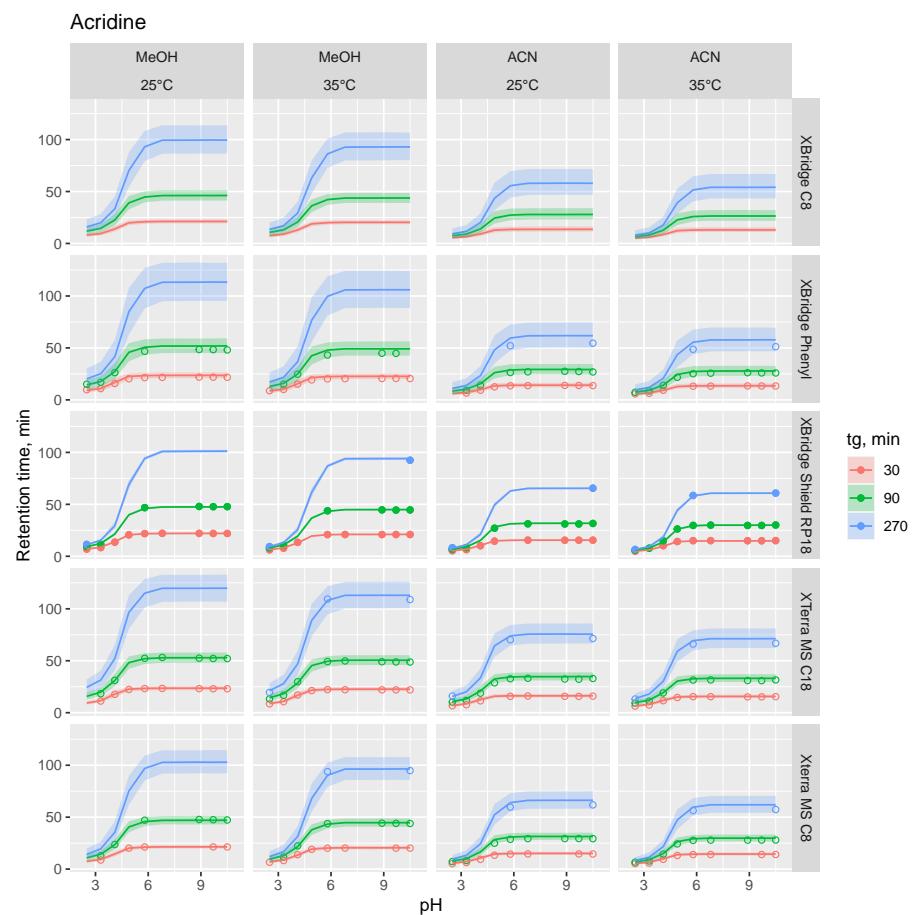
7 FIGURE S5. POPULATION GRADIENT PREDICTIONS.



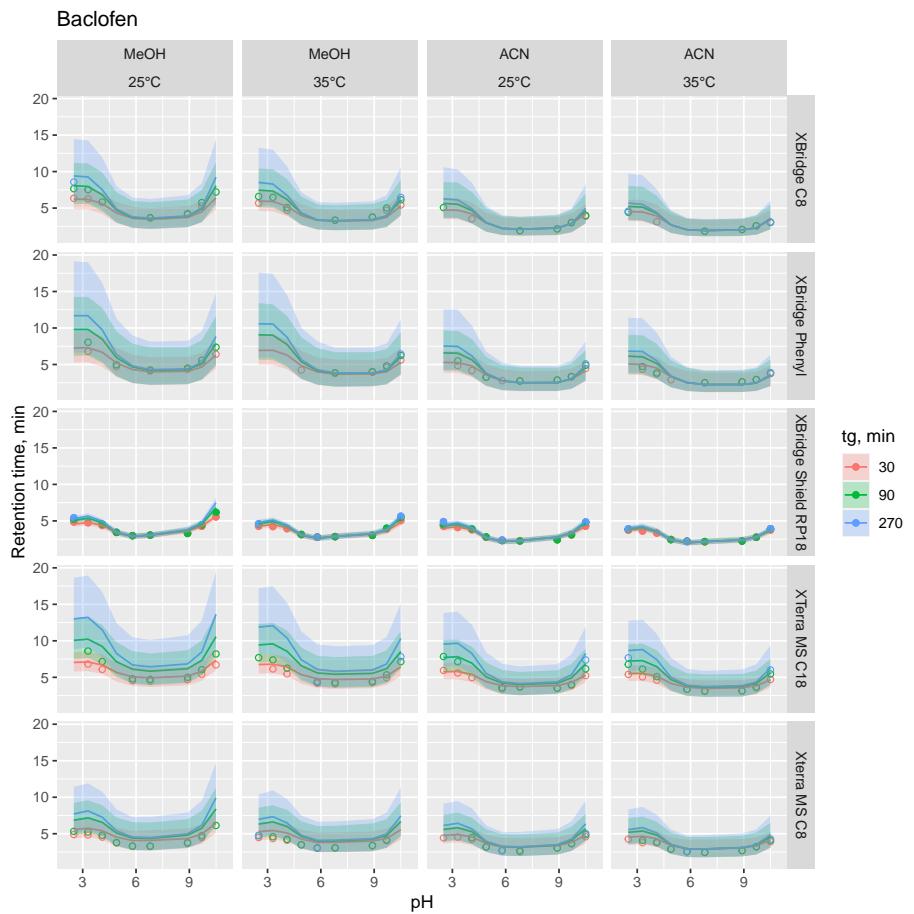
8 FIGURE S6. LIMITED DATA GRADIENT PREDICTIONS.

8 Figure S6. Limited data gradient predictions.

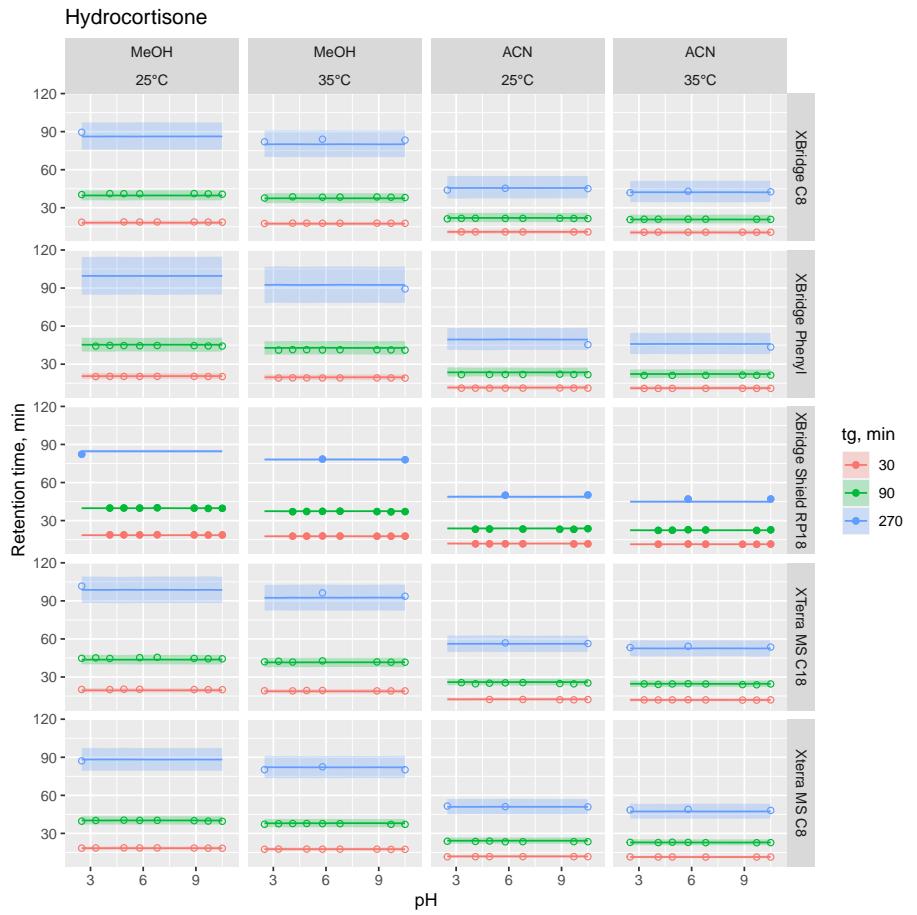
Predictions represented as posterior median (line) and 5th-95th percentiles (areas) for 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given population-level parameters and predictors (logP and pKa), and XBridge Shield RP18 data. Closed dots represent data used for predictions.



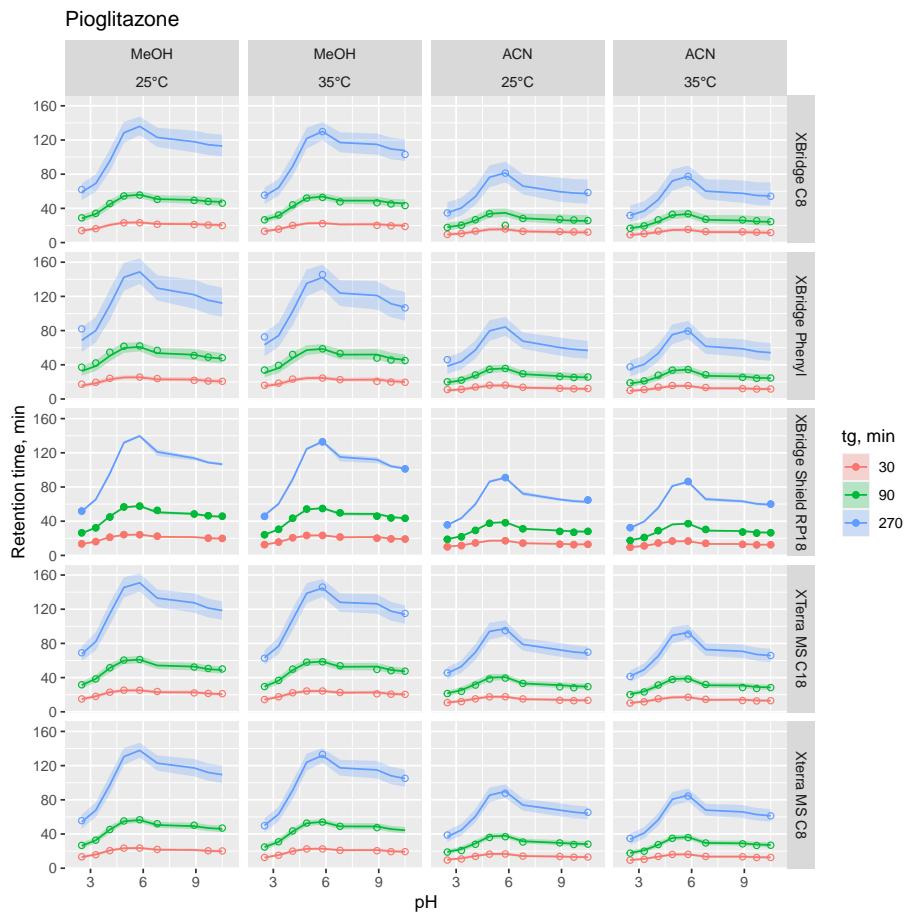
8 FIGURE S6. LIMITED DATA GRADIENT PREDICTIONS.



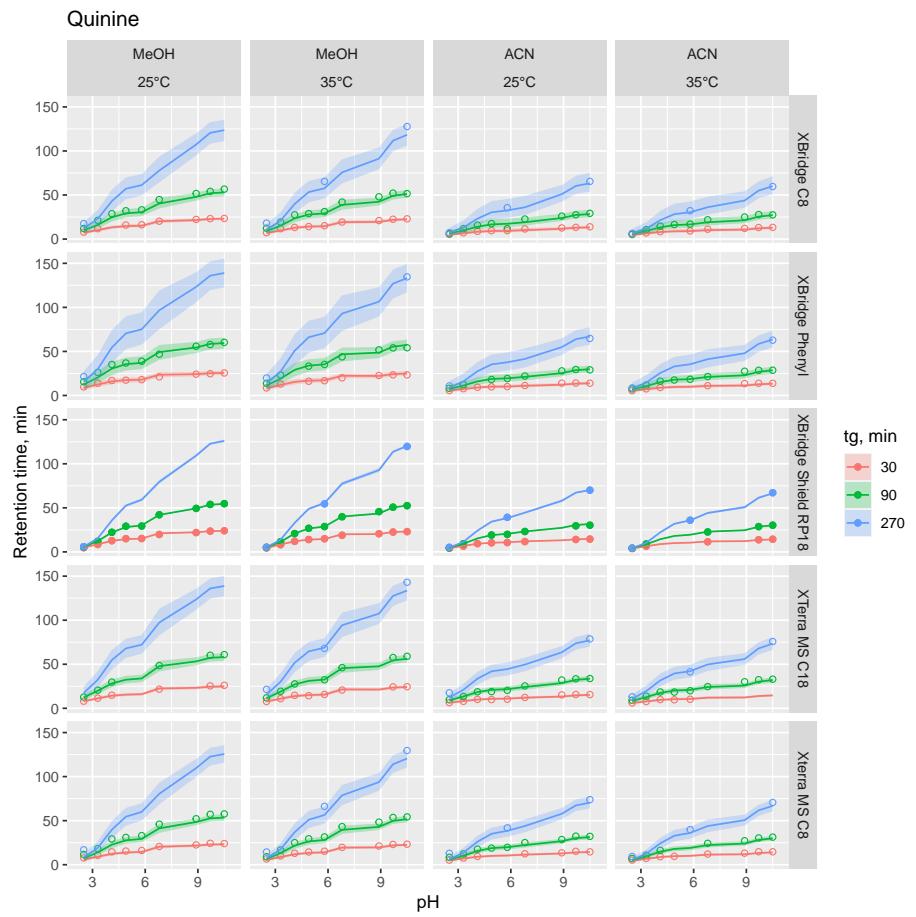
8 FIGURE S6. LIMITED DATA GRADIENT PREDICTIONS.



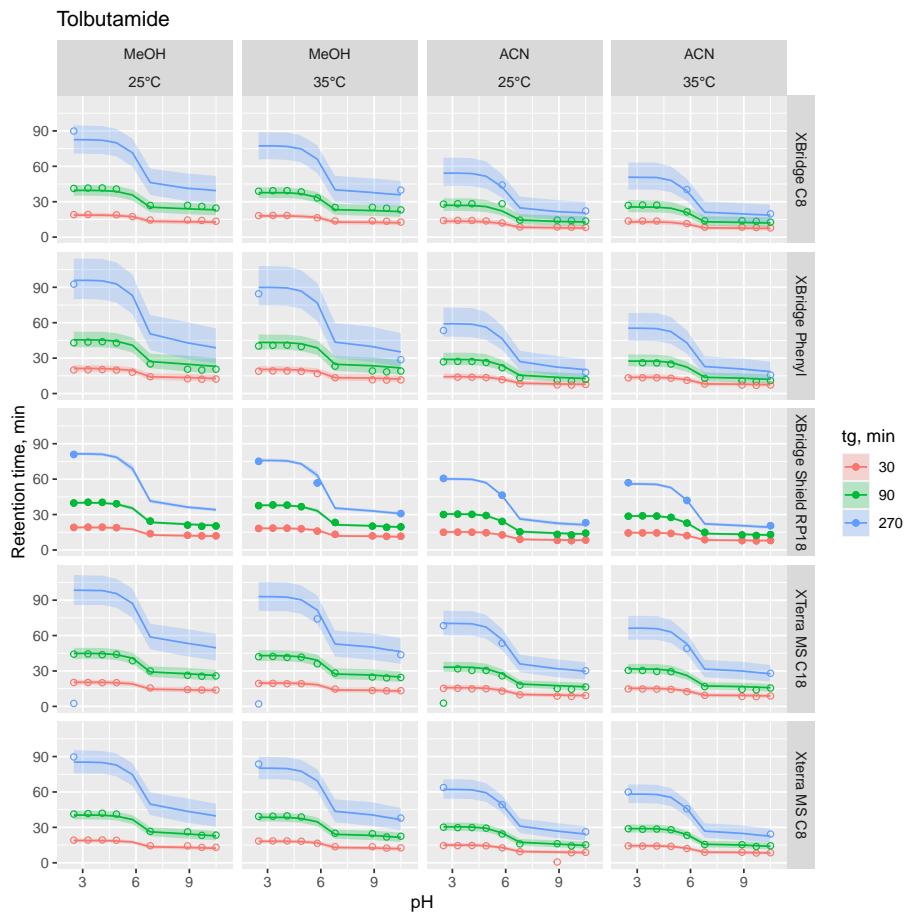
8 FIGURE S6. LIMITED DATA GRADIENT PREDICTIONS.



8 FIGURE S6. LIMITED DATA GRADIENT PREDICTIONS.



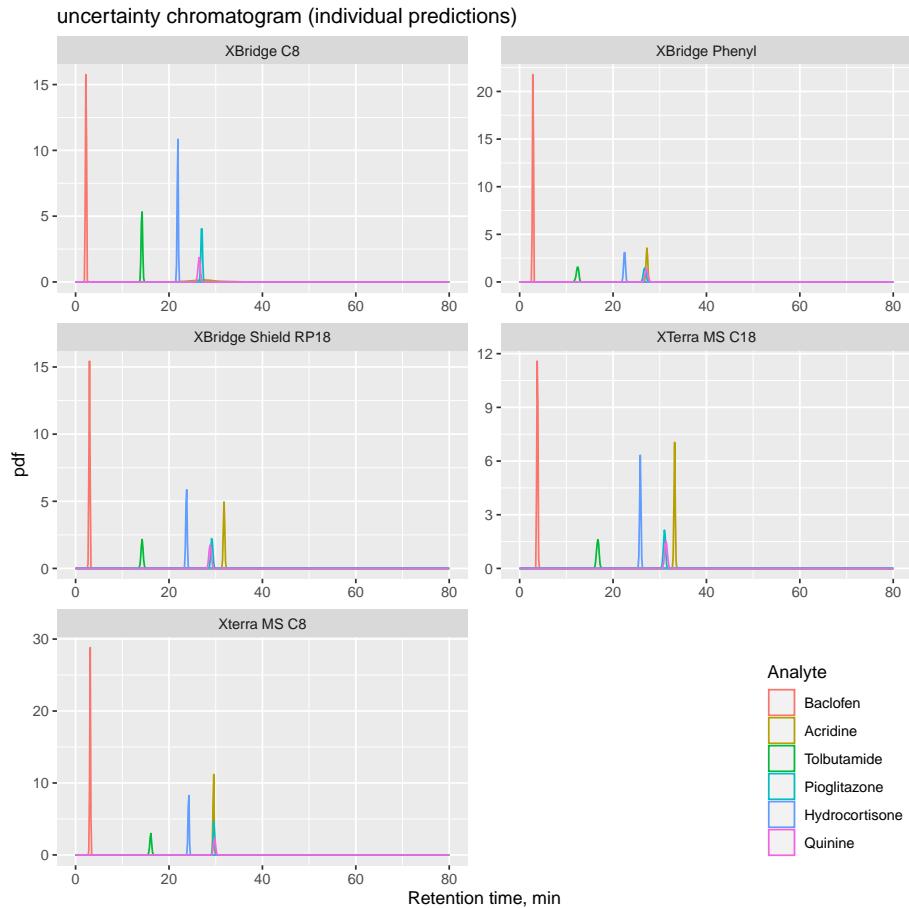
8 FIGURE S6. LIMITED DATA GRADIENT PREDICTIONS.



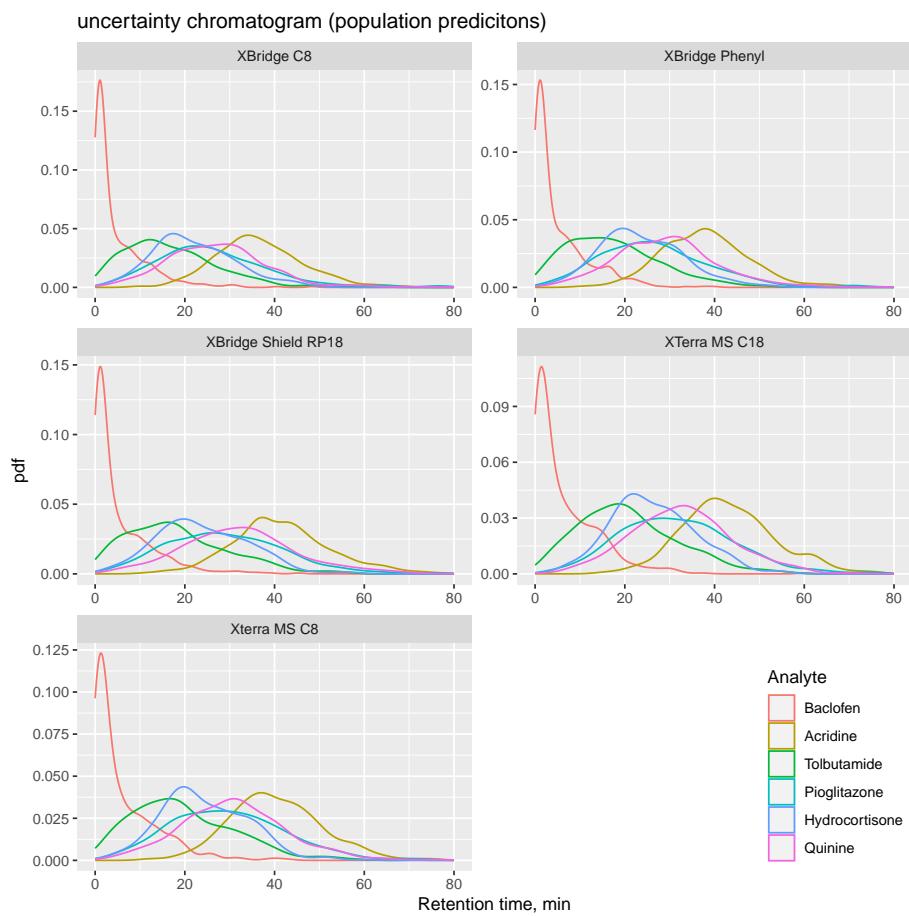
9 FIGURE S7. UNCERTAINTY CHROMATOGRAMS.

9 Figure S7. Uncertainty chromatograms.

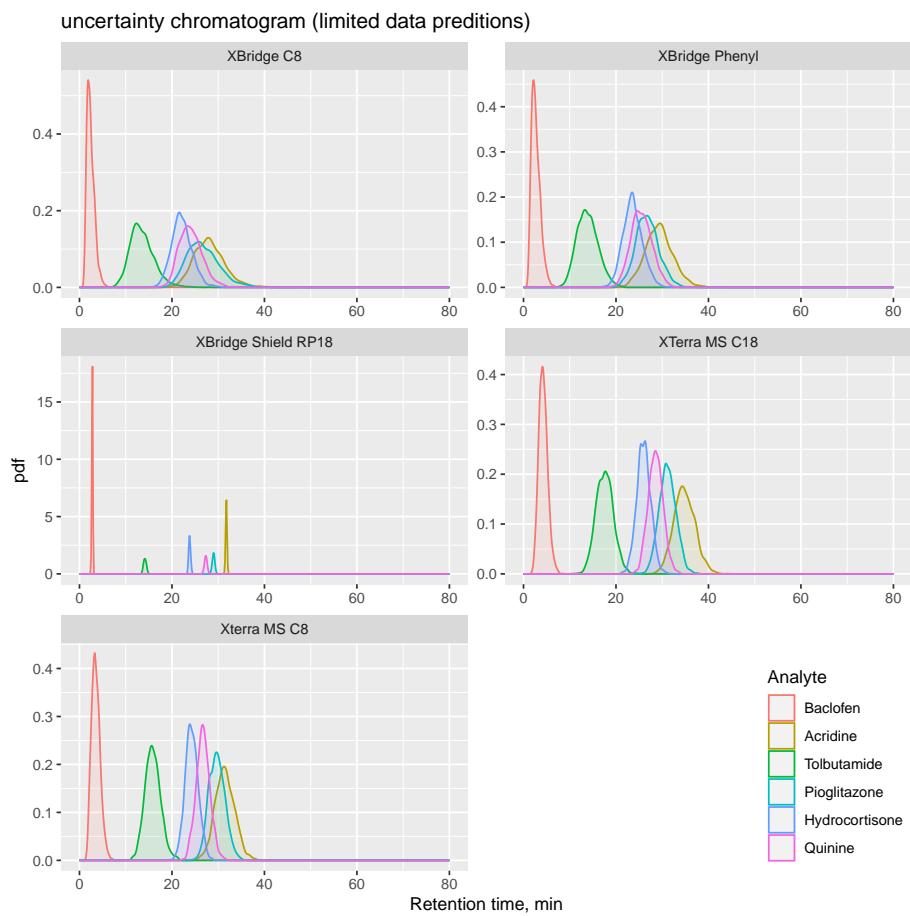
Uncertainty chromatograms displaying the predictions for 6 selected analytes using individual, population and limited data predictions. Each peak represents the range of analyte retention factors compatible with prior and preliminary data. The chromatographic conditions are pH_o=8.9, ACN, 25°C, tg = 90 min.



9 FIGURE S7. UNCERTAINTY CHROMATOGRAMS.

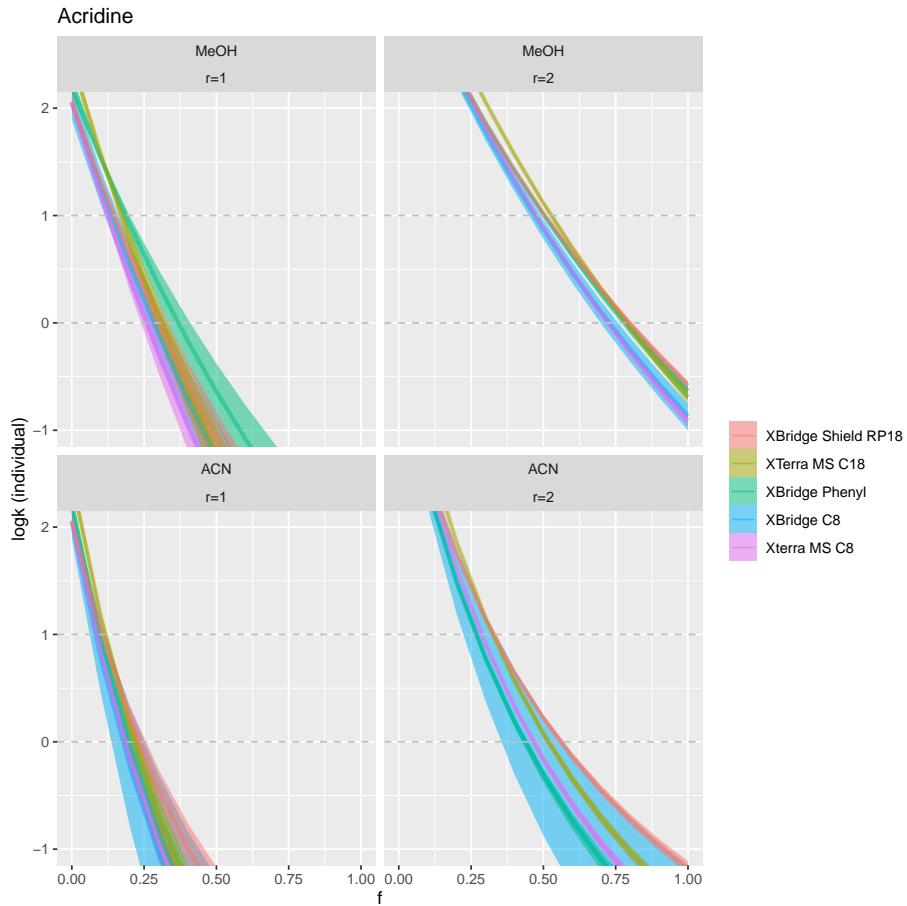


9 FIGURE S7. UNCERTAINTY CHROMATOGRAMS.

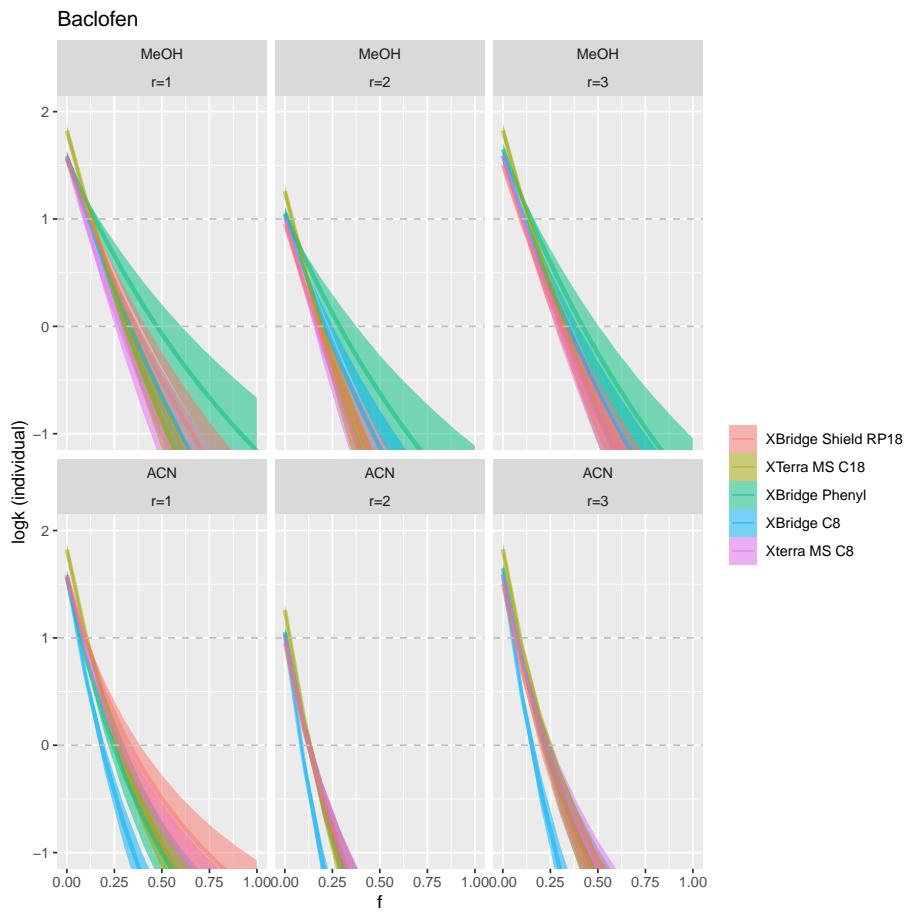


10 Figure S8. Individual izocratic predictions.

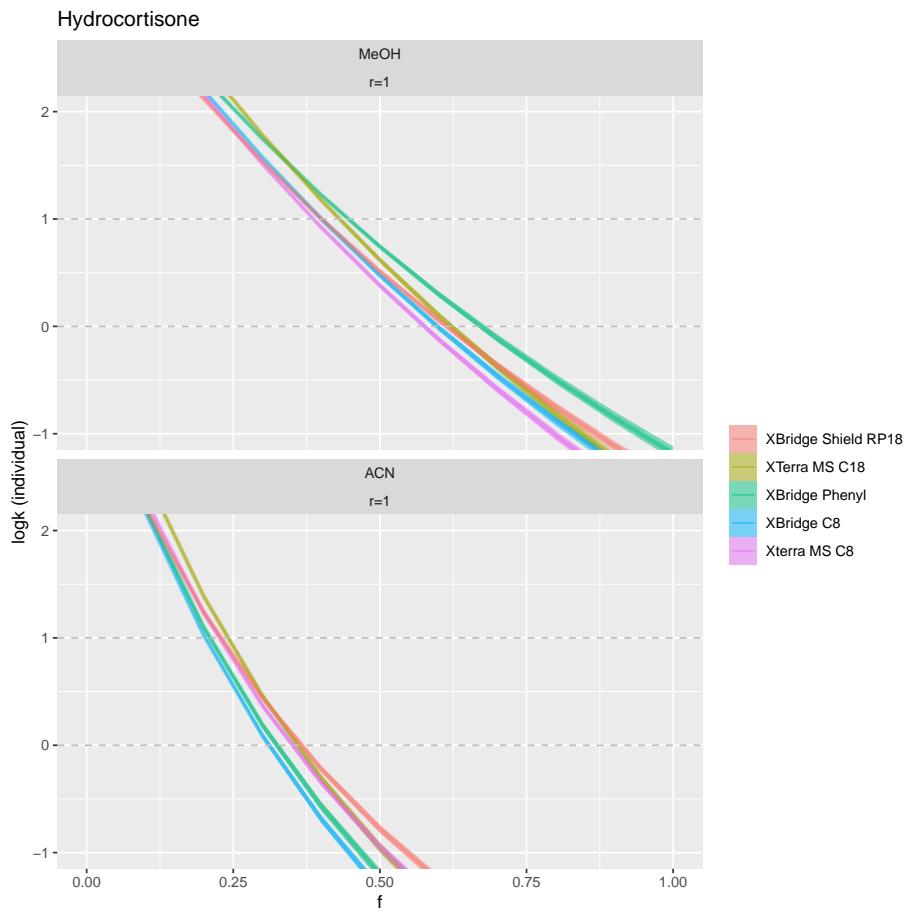
Predictions represented as posterior median (line) and 5th-95th percentiles (areas) for a 6 exemplary analytes. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.



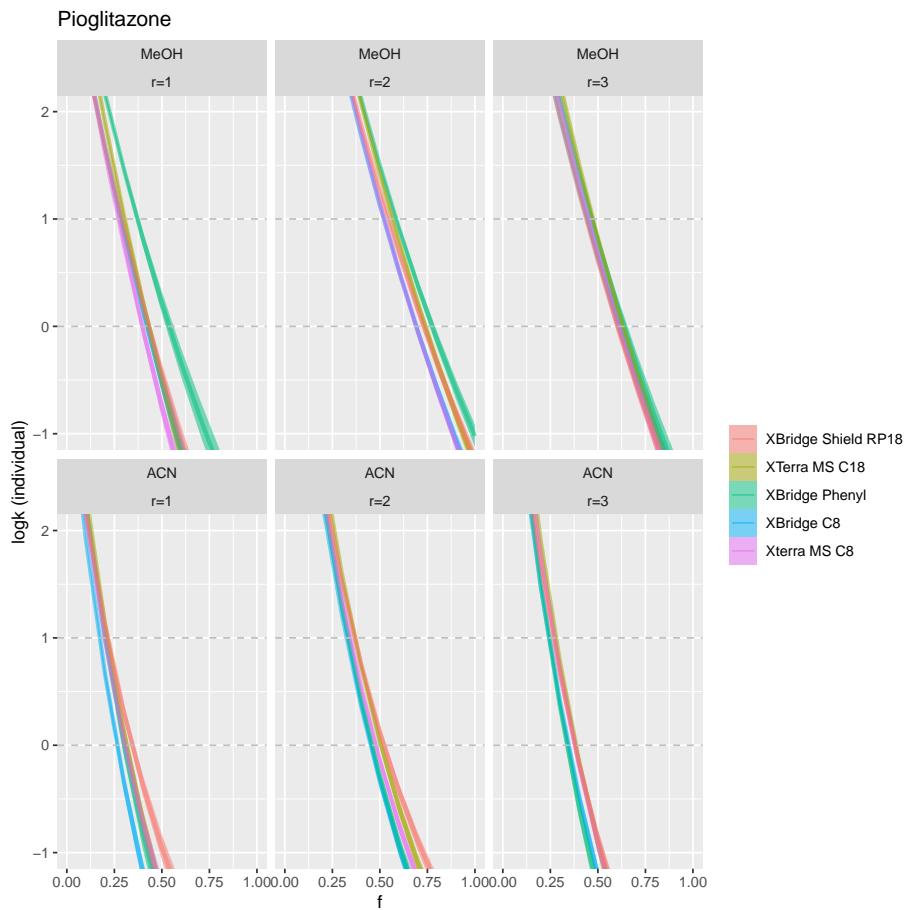
10 FIGURE S8. INDIVIDUAL IZOCRATIC PREDICTIONS.



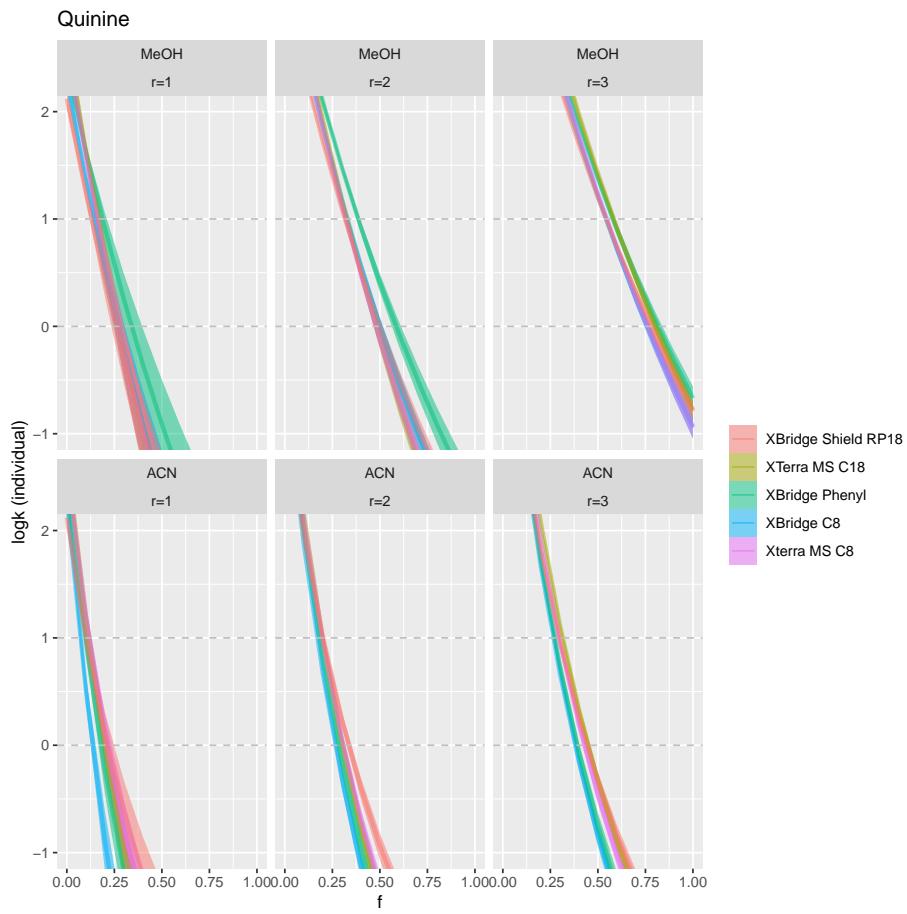
10 FIGURE S8. INDIVIDUAL IZOCRATIC PREDICTIONS.



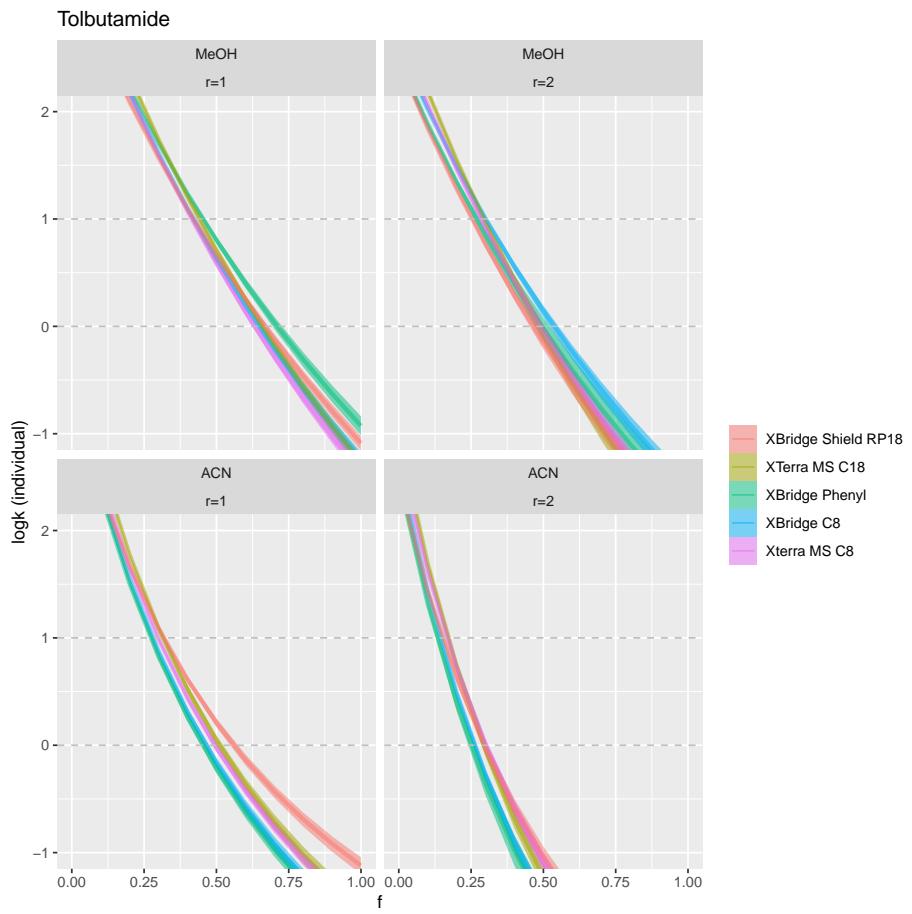
10 FIGURE S8. INDIVIDUAL IZOCRATIC PREDICTIONS.



10 FIGURE S8. INDIVIDUAL IZOCRATIC PREDICTIONS.

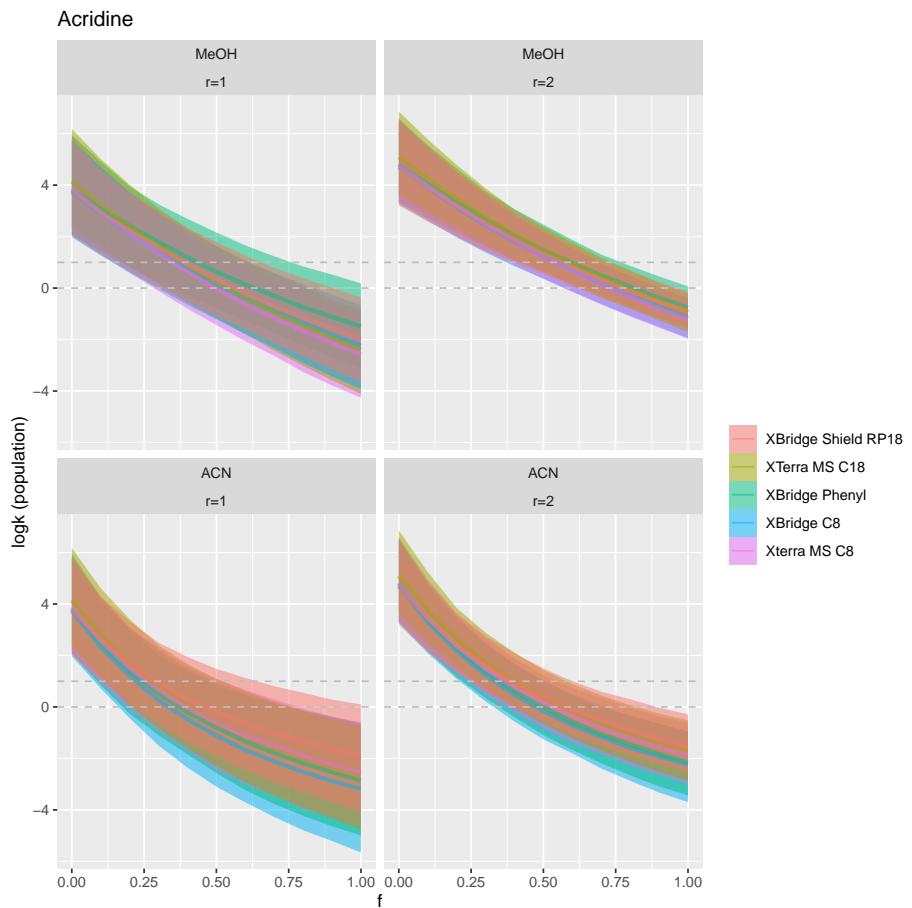


10 FIGURE S8. INDIVIDUAL IZOCRATIC PREDICTIONS.

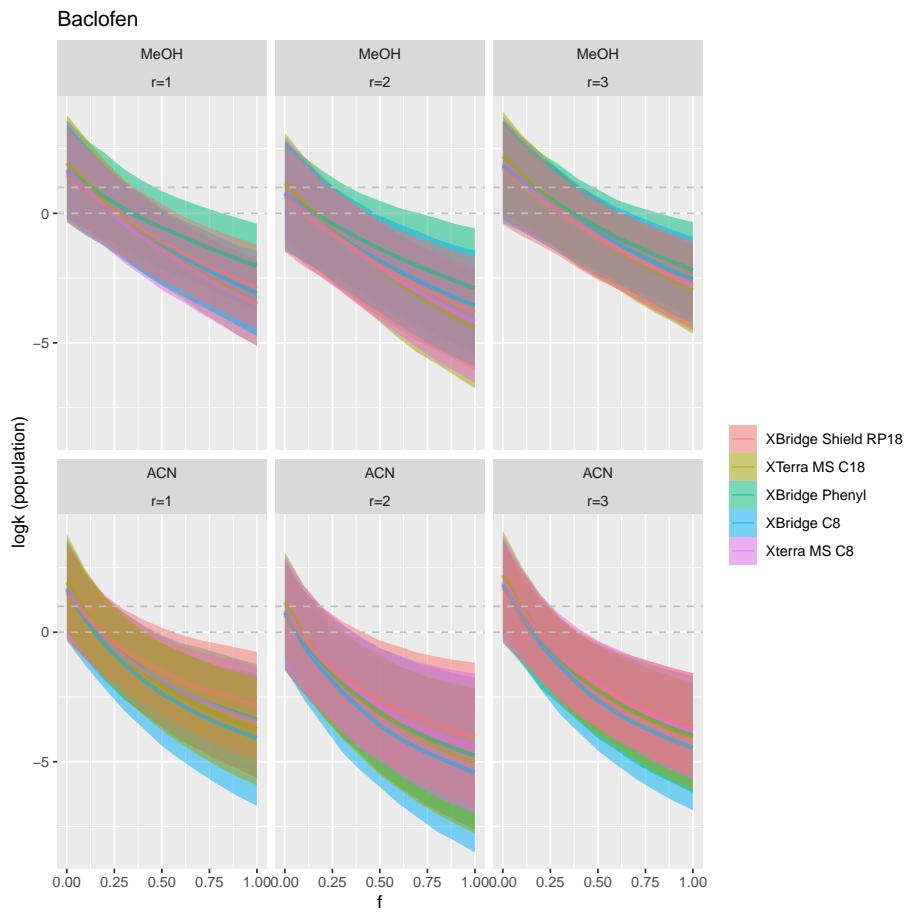


11 Figure S9. Population isocratic predictions.

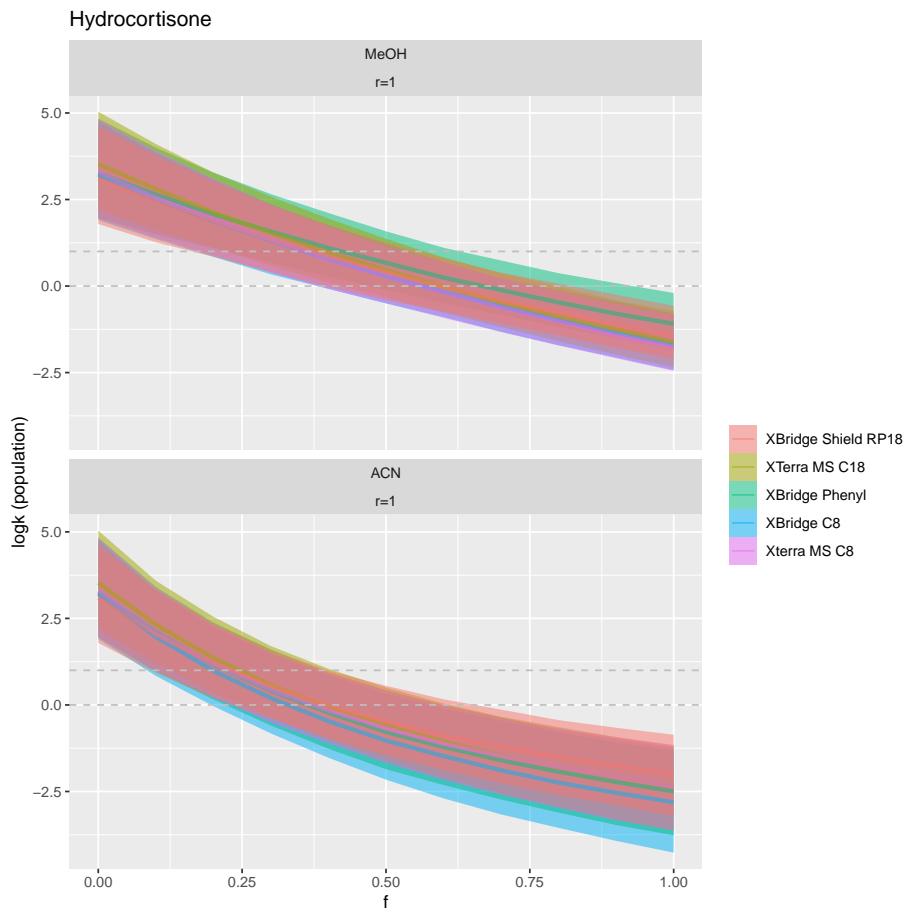
Predictions represented as posterior median (line) and 5th-95th percentiles (areas) for a 6 exemplary analytes. Predictions corresponding to future observations given only population-level parameters and predictors (logP and pKa).



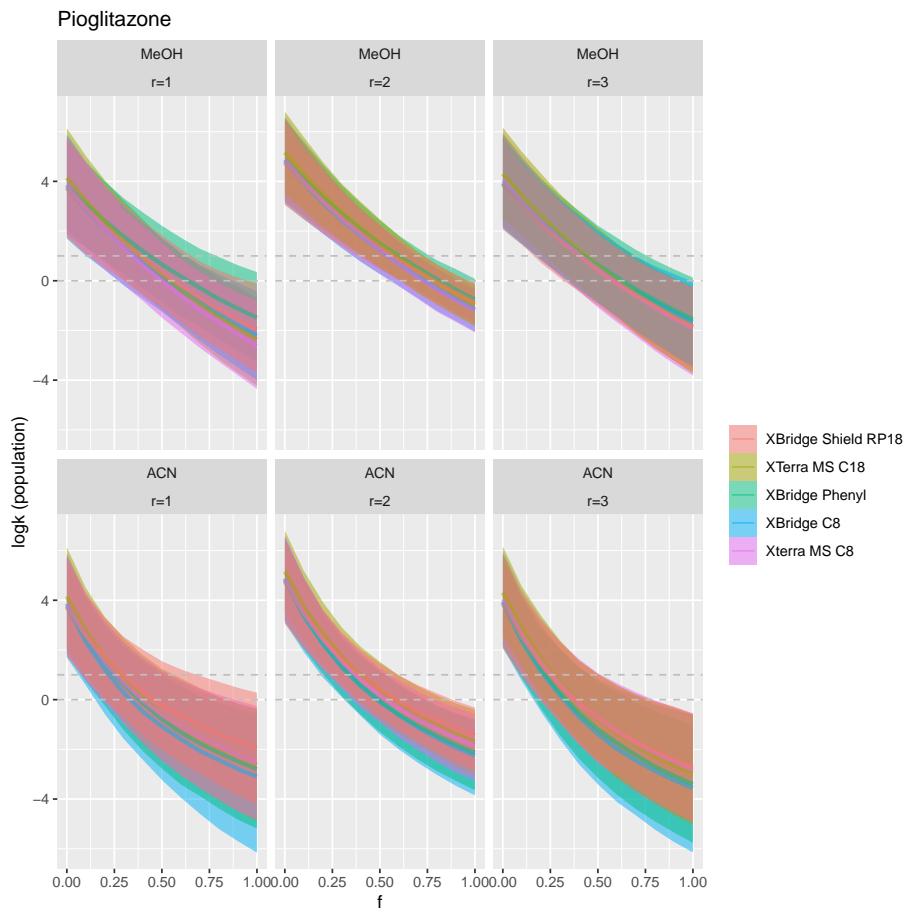
11 FIGURE S9. POPULATION ISOCRATIC PREDICTIONS.



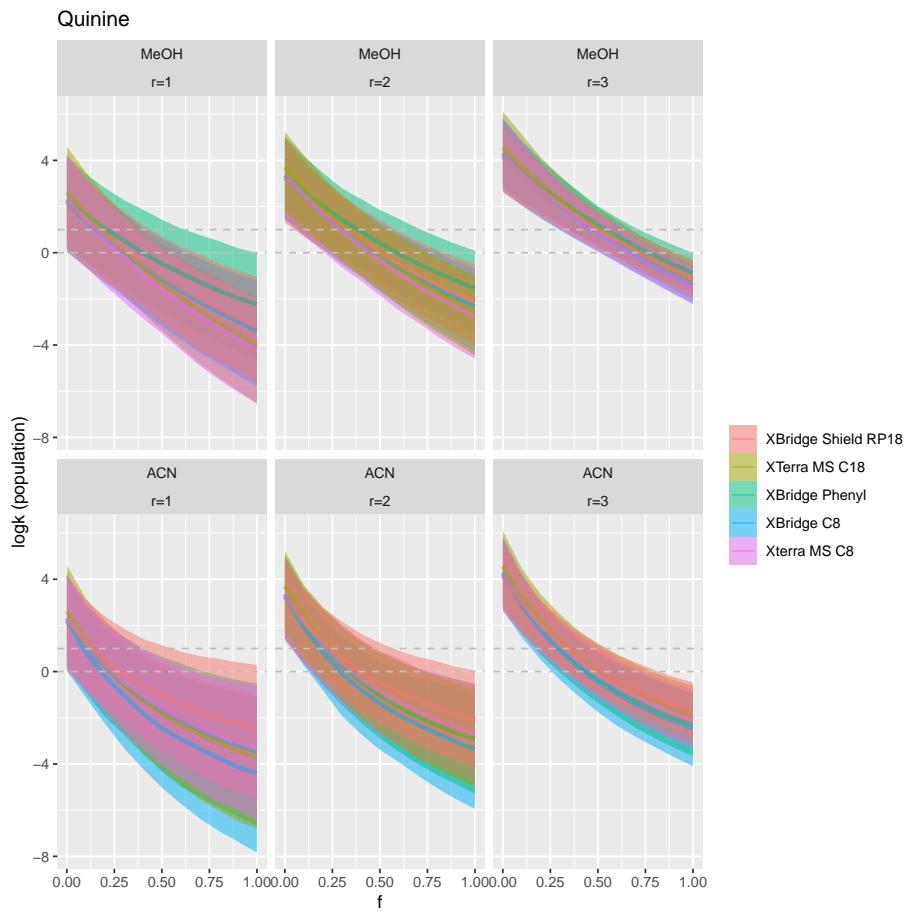
11 FIGURE S9. POPULATION ISOCRATIC PREDICTIONS.



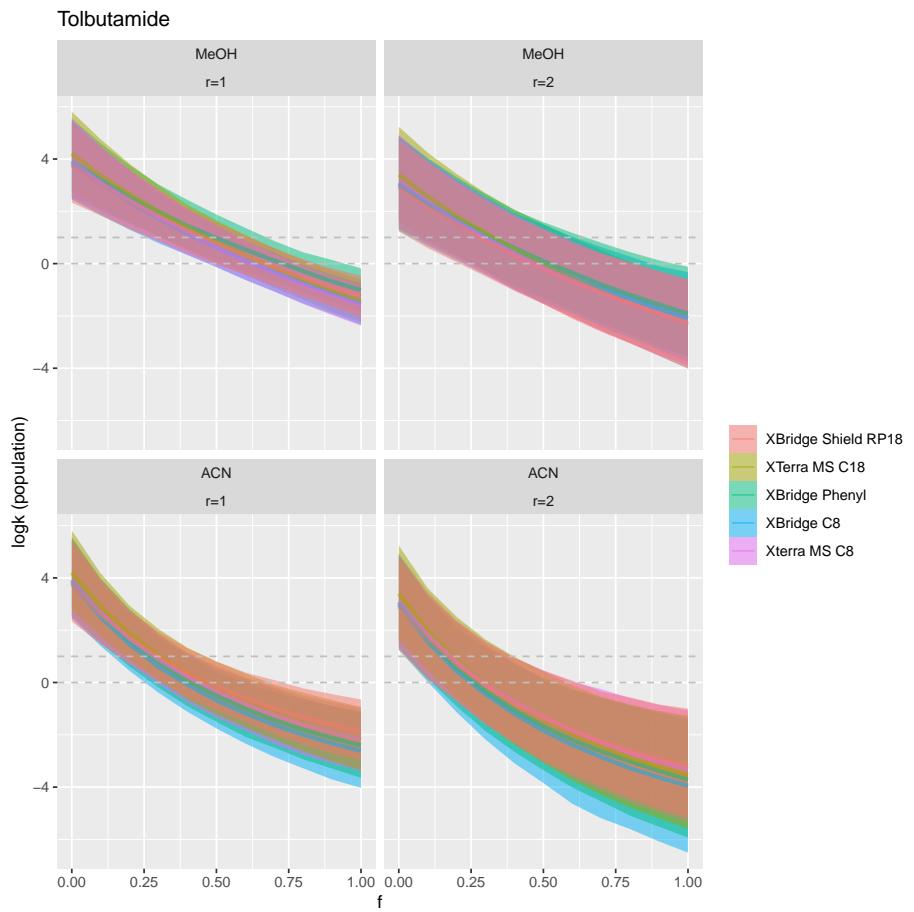
11 FIGURE S9. POPULATION ISOCRATIC PREDICTIONS.



11 FIGURE S9. POPULATION ISOCRATIC PREDICTIONS.



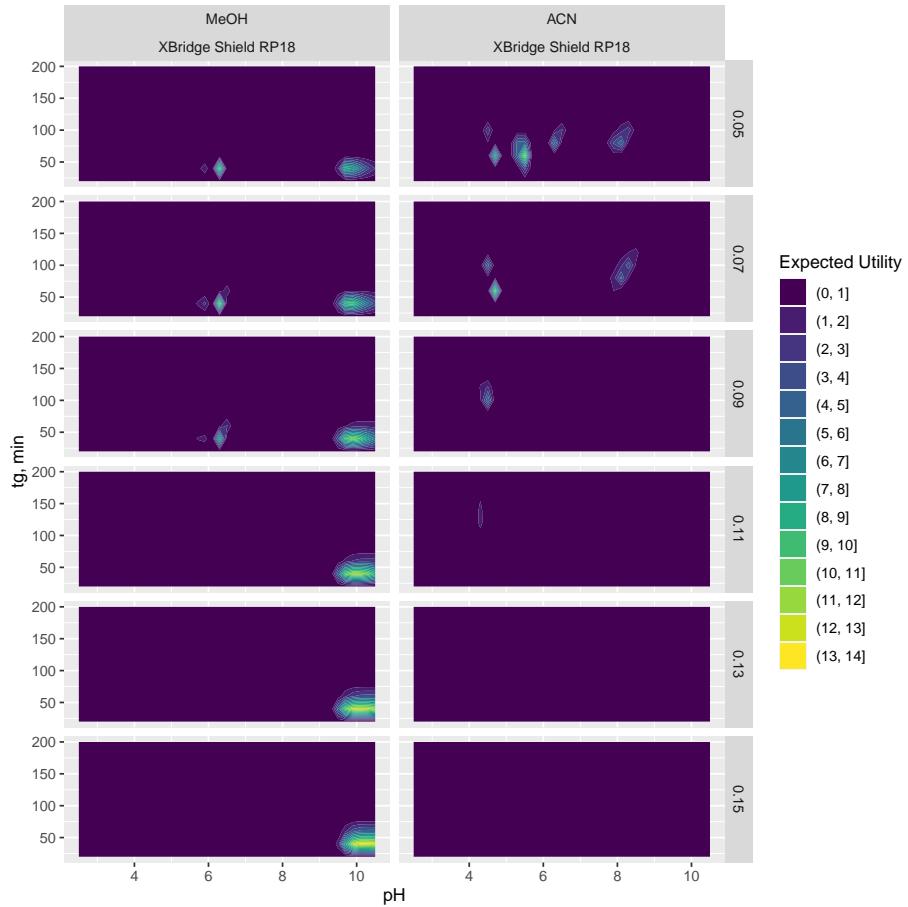
11 FIGURE S9. POPULATION ISOCRATIC PREDICTIONS.



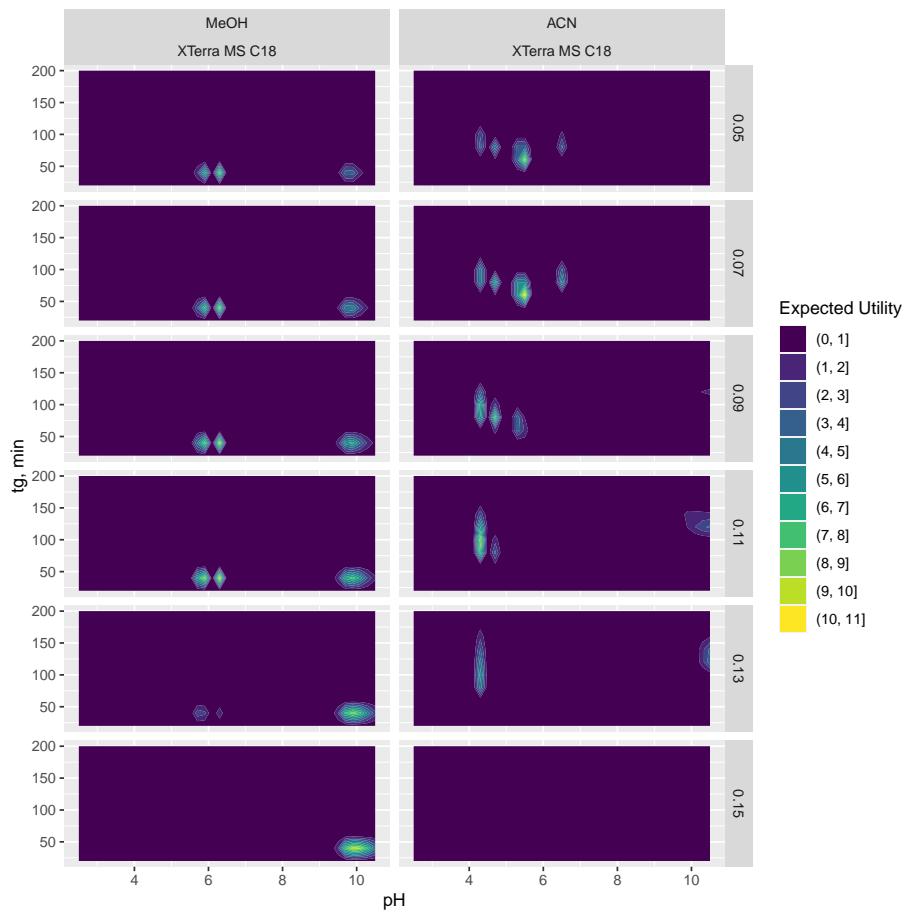
12 FIGURE S10. THE EXPECTED UTILITY MAPS BASED ON INDIVIDUAL PREDICTIONS

12 Figure S10. The expected utility maps based on individual predictions

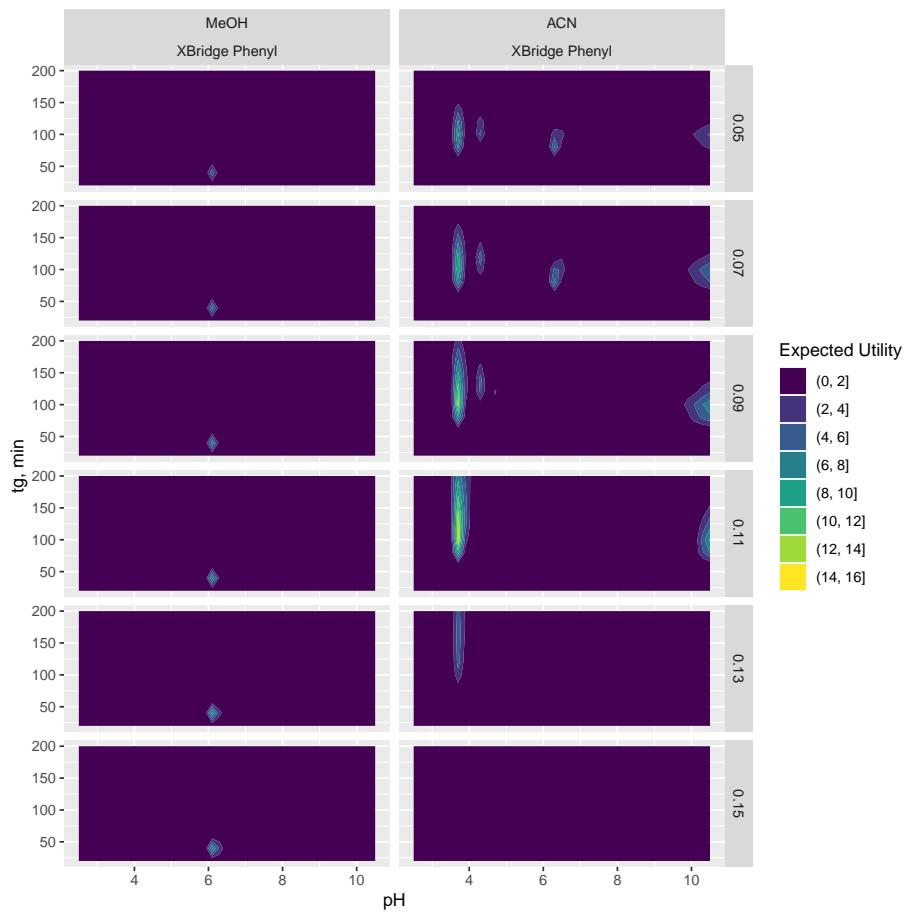
The expected utility map for each column:



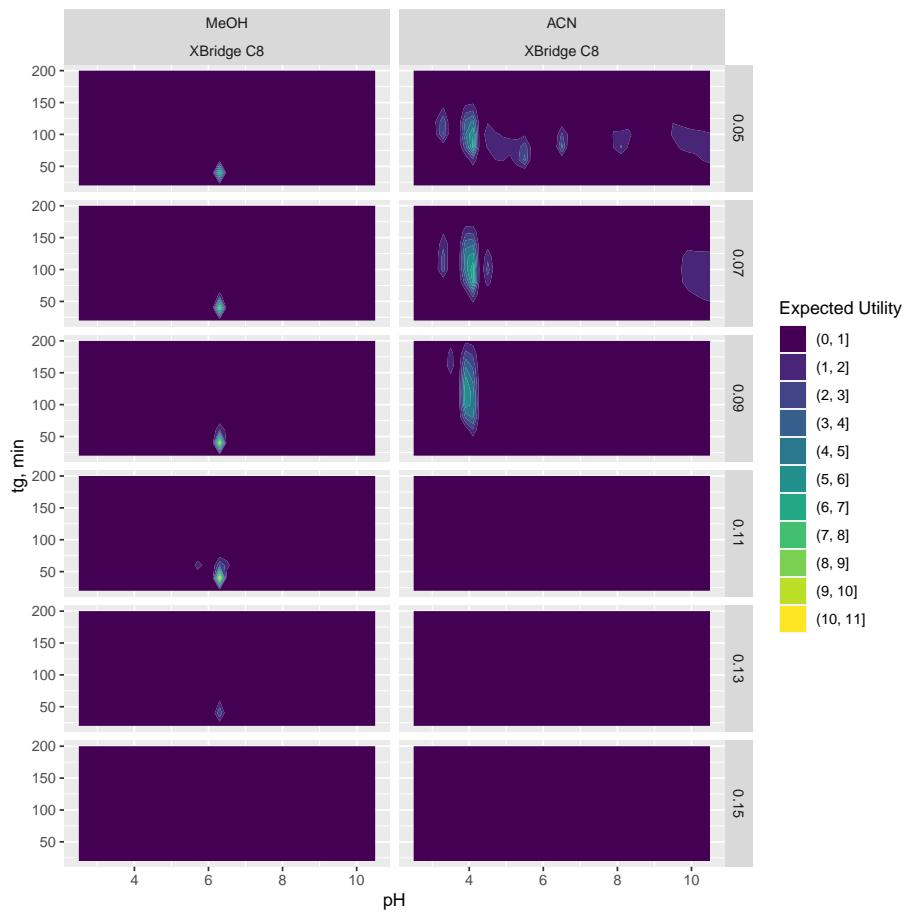
12 FIGURE S10. THE EXPECTED UTILITY MAPS BASED ON INDIVIDUAL PREDICTIONS



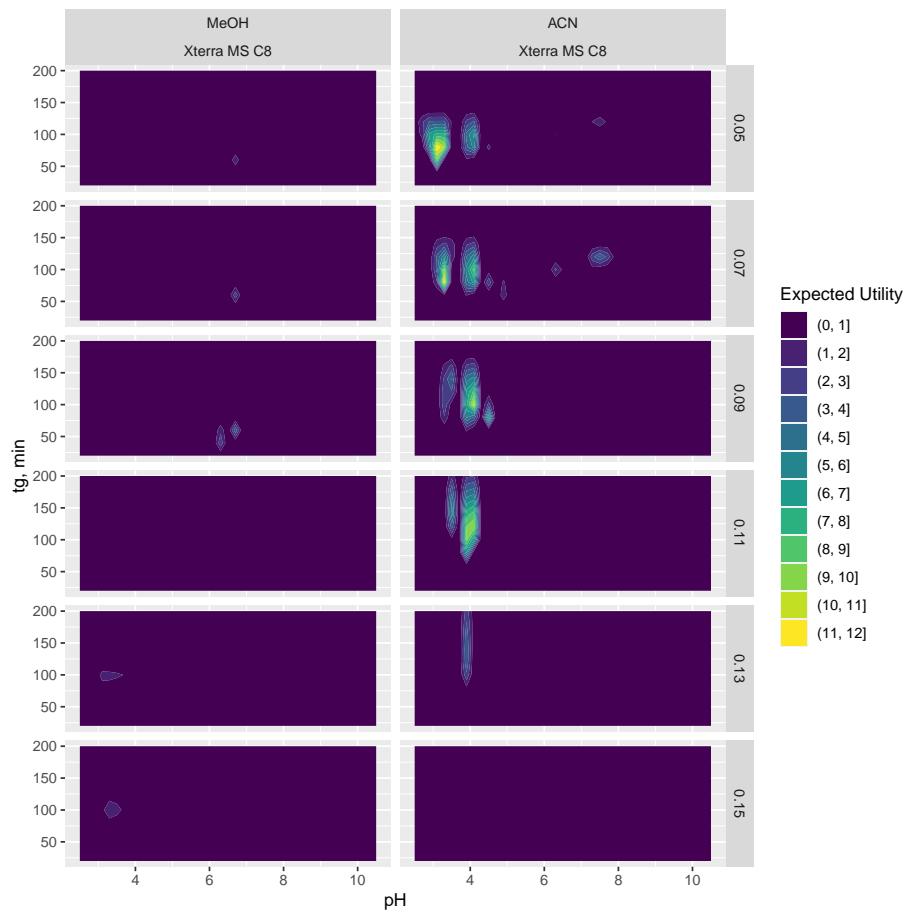
12 FIGURE S10. THE EXPECTED UTILITY MAPS BASED ON INDIVIDUAL PREDICTIONS



12 FIGURE S10. THE EXPECTED UTILITY MAPS BASED ON INDIVIDUAL PREDICTIONS



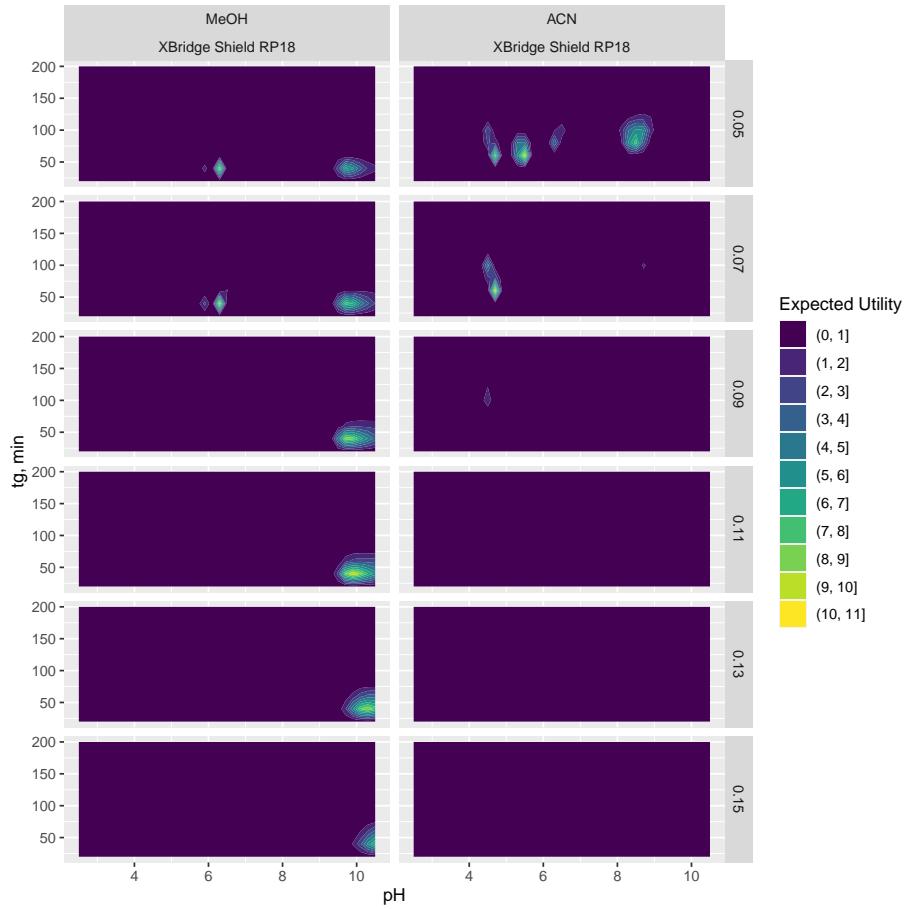
12 FIGURE S10. THE EXPECTED UTILITY MAPS BASED ON INDIVIDUAL PREDICTIONS



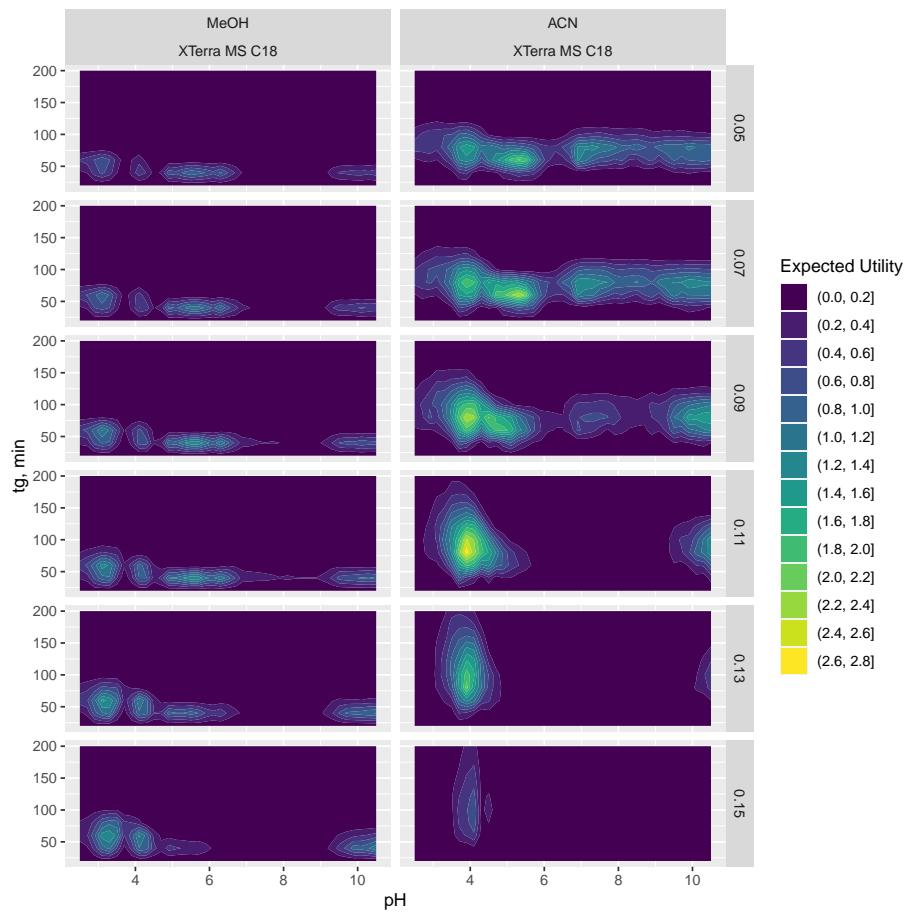
13 FIGURE S11. THE EXPECTED UTILITY MAPS BASED ON LIMITED DATA PREDICTIONS

13 Figure S11. The expected utility maps based on limited data predictions

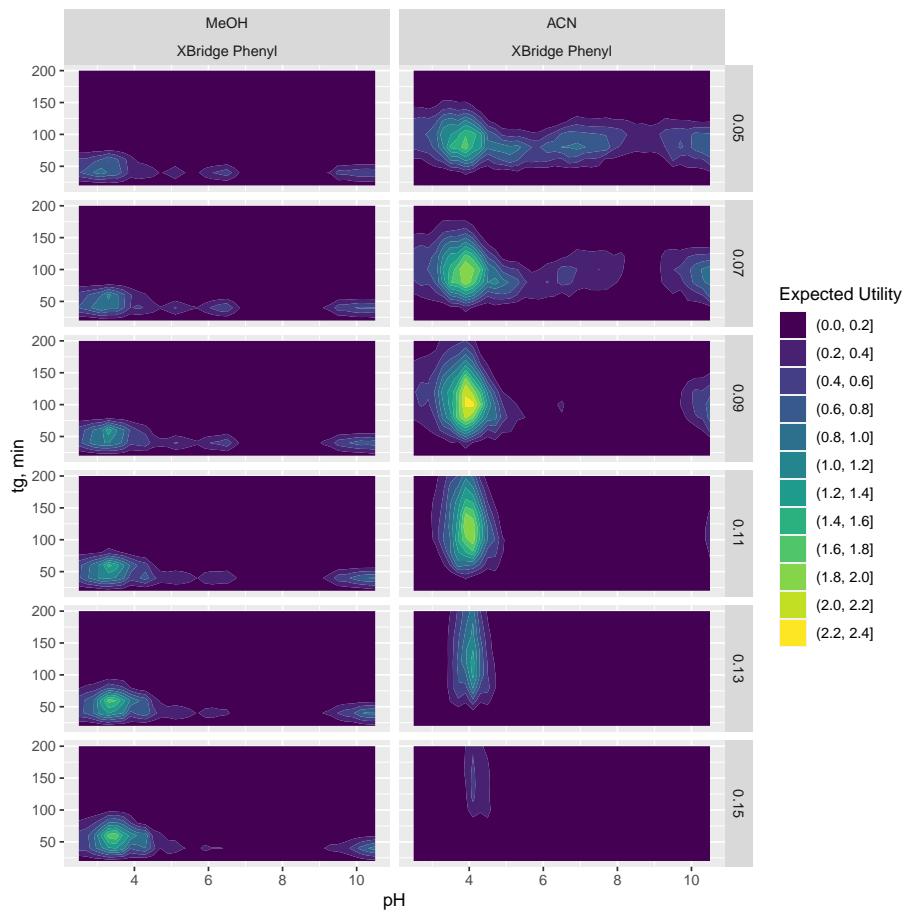
The expected utility map for each column:



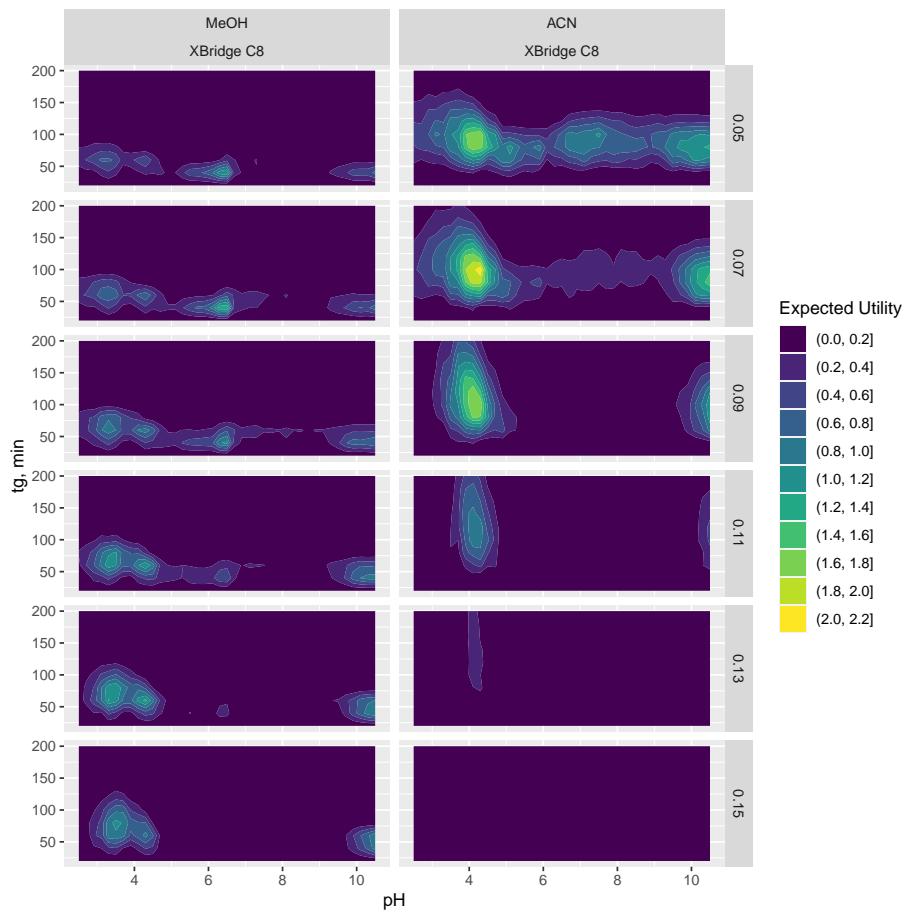
13 FIGURE S11. THE EXPECTED UTILITY MAPS BASED ON LIMITED DATA PREDICTIONS



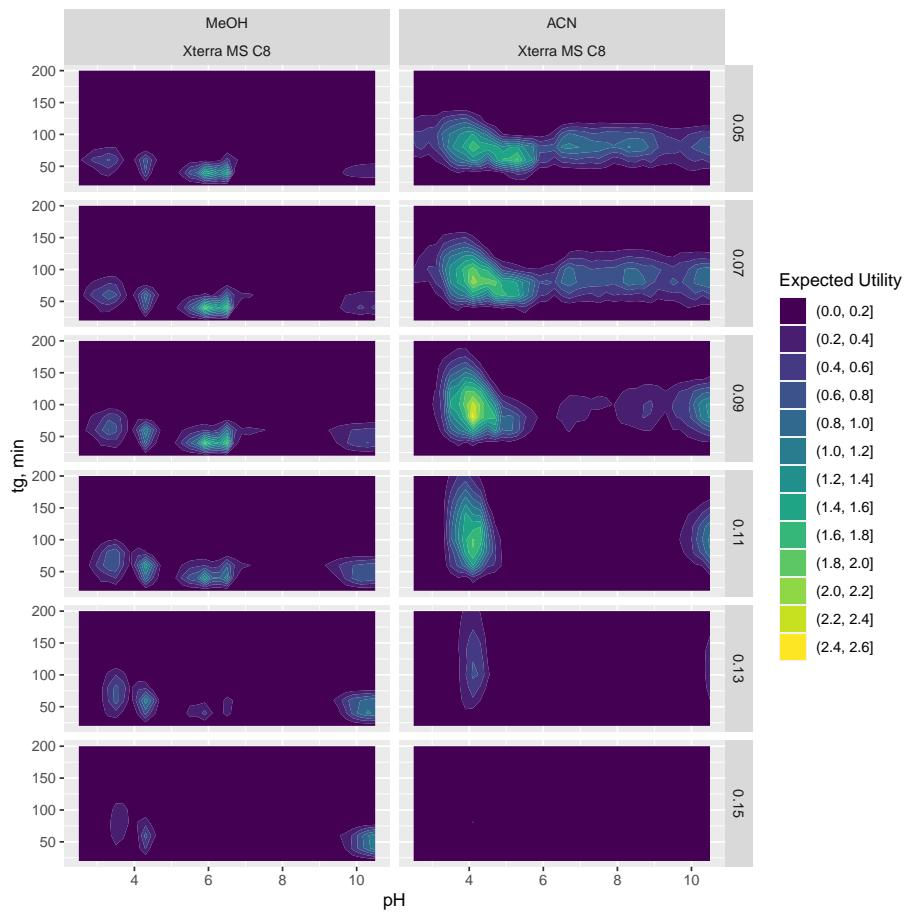
13 FIGURE S11. THE EXPECTED UTILITY MAPS BASED ON LIMITED DATA PREDICTIONS



13 FIGURE S11. THE EXPECTED UTILITY MAPS BASED ON LIMITED DATA PREDICTIONS



13 FIGURE S11. THE EXPECTED UTILITY MAPS BASED ON LIMITED DATA PREDICTIONS



Licenses

- Code & copy; 2023, Paweł Wiczling, licensed under BSD-3.
- Text & copy; 2023, Paweł Wiczling, licensed under CC-BY-NC 4.0.

Original Computing Environment

```
R version 4.1.3 (2022-03-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 22621)

Matrix products: default

locale:
[1] LC_COLLATE=Polish_Poland.1250  LC_CTYPE=Polish_Poland.1250
[3] LC_MONETARY=Polish_Poland.1250 LC_NUMERIC=C
[5] LC_TIME=Polish_Poland.1250

attached base packages:
[1] stats      graphics   grDevices utils      datasets   methods    base

other attached packages:
[1] kableExtra_1.3.4  GGally_2.1.2      posterior_1.3.1   bayesplot_1.9.0
[5] reshape2_1.4.4   knitr_1.40      cmdstanr_0.5.3   gridExtra_2.3
[9] ggplot2_3.4.2   dplyr_1.1.2      pracma_2.3.8

loaded via a namespace (and not attached):
[1] Rcpp_1.0.9          svglite_2.1.1      digest_0.6.29
[4] utf8_1.2.2          R6_2.5.1          plyr_1.8.7
[7] ggridges_0.5.3      backports_1.4.1    evaluate_0.16
[10] httr_1.4.5          pillar_1.9.0      rlang_1.1.0
[13] rstudioapi_0.14     data.table_1.14.2  checkmate_2.1.0
[16] rmarkdown_2.16       textshaping_0.3.6  webshot_0.5.4
[19] stringr_1.5.0       munsell_0.5.0     compiler_4.1.3
[22] xfun_0.32          pkgconfig_2.0.3    systemfonts_1.0.4
[25] htmltools_0.5.3     tidyselect_1.2.0   tibble_3.2.1
[28] tensorA_0.36.2     matrixStats_0.62.0 codetools_0.2-18
[31] reshape_0.8.9       fansi_1.0.3       viridisLite_0.4.1
[34] withr_2.5.0          grid_4.1.3        distributional_0.3.1
[37] jsonlite_1.8.4      gtable_0.3.1      lifecycle_1.0.3
[40] magrittr_2.0.3       scales_1.2.1      cli_3.4.0
[43] stringi_1.7.8       farver_2.1.1      xml2_1.3.3
[46] ragg_1.2.5          generics_0.1.3    vctrs_0.6.2
[49] RColorBrewer_1.1-3  tools_4.1.3       glue_1.6.2
```

13 FIGURE S11. THE EXPECTED UTILITY MAPS BASED ON
LIMITED DATA PREDICTIONS

```
[52] abind_1.4-5          fastmap_1.1.0        yaml_2.3.5
[55] colorspace_2.0-3      rvest_1.0.3
```