Supplementary Material for: Raman-guided Sample Subset Selection for Cost-efficient Offline Calibration in Bioprocesses

Supplementary material

Raman-guided Sample Subset Selection for Cost-efficient Offline Calibration in Bioprocesses

Terrance Wilms[a], Fabian Schwenke[a], Rudibert King[a], Steffi Knorn[a]

[a]   Technische Universität Berlin, Chair of Control, Hardenbergstr. 36a, 10623 Berlin, Germany

Contents

Supplementary Material

This document contains supplementary figures and tables referenced in the main manuscript as **Figures S1–S9** and **Tables S1–S2**.
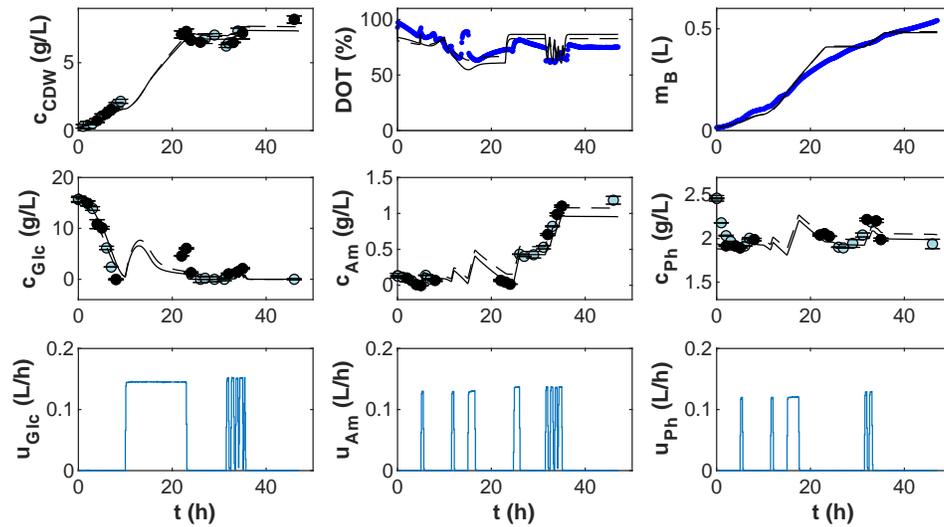
1   Supplementary Figures

Figure S1: PI auto-validation (F1) using FS-CVQ-10 is illustrated with selected samples represented as black dots with error bars, while the remaining samples are shown in green. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The simulation using FS-CVQ-10 is depicted by continuous black lines, and the simulation using No-CVQ is shown as dashed black lines.
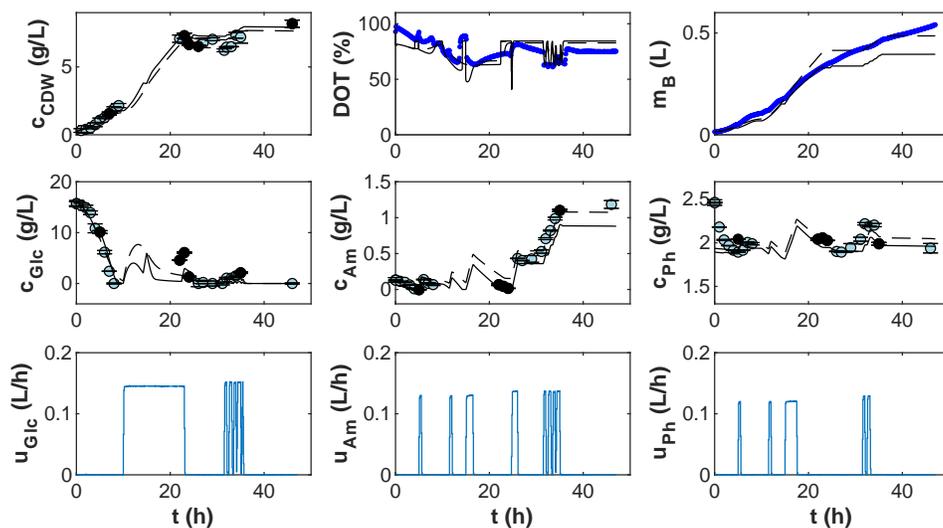
Figure S2: PI auto-validation (F1) using FS-CVQ-05 is illustrated with selected samples represented as black dots with error bars, while the remaining samples are shown in green. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The simulation using FS-CVQ-05 is depicted by continuous black lines, and the simulation using No-CVQ is shown as dashed black lines.

Figure S3: PI auto-validation (F1) using WS-CVQ-10 is illustrated with selected samples represented as black dots with error bars, while the remaining samples are shown in green. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The simulation using WS-CVQ-10 is depicted by continuous black lines, and the simulation using No-CVQ is shown as dashed black lines.
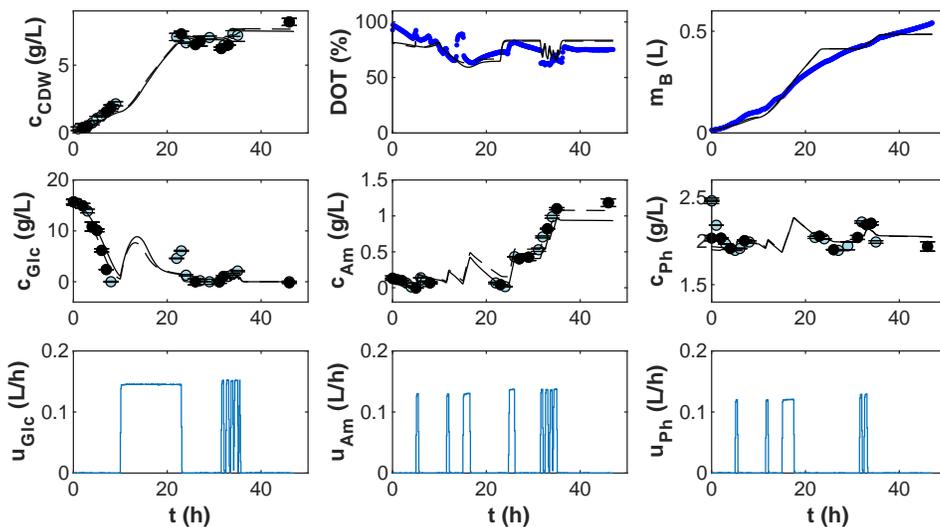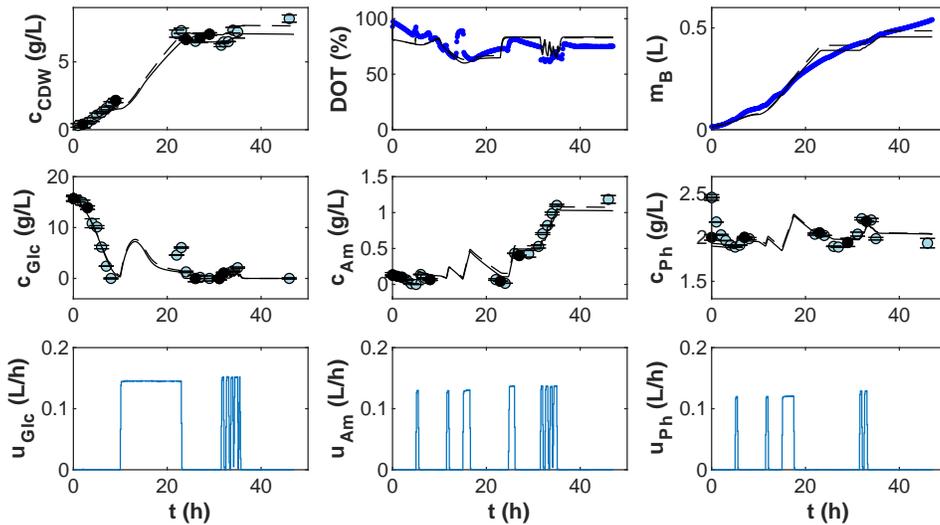
Figure S4: PI auto-validation (F1) using WS-CVQ-05 is illustrated with selected samples represented as black dots with error bars, while the remaining samples are shown in green. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The simulation using WS-CVQ-05 is depicted by continuous black lines, and the simulation using No-CVQ is shown as dashed black lines.

Figure S5: Cross-validation (F2) using No-CVQ is shown with all samples represented as green dots with error bars. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The PI cross-validation is illustrated by dashed black lines, while the cUKF cross-validation is depicted by continuous black lines.

Figure S6: Cross-validation (F2) using FS-CVQ-10 is shown with all samples represented as green dots with error bars. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The PI cross-validation is illustrated by dashed black lines.
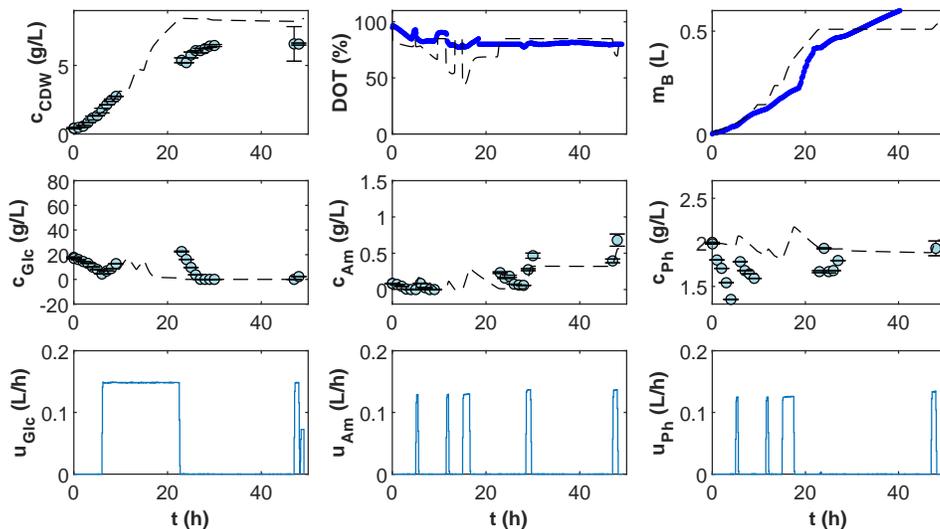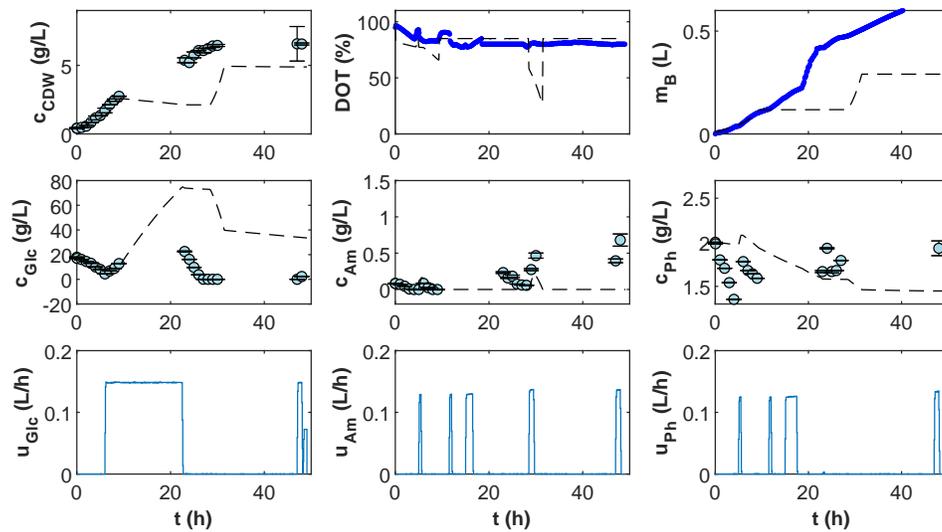

Figure S7: Cross-validation (F2) using FS-CVQ-05 is shown with all samples represented as green dots with error bars. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The PI cross-validation is illustrated by dashed black lines.
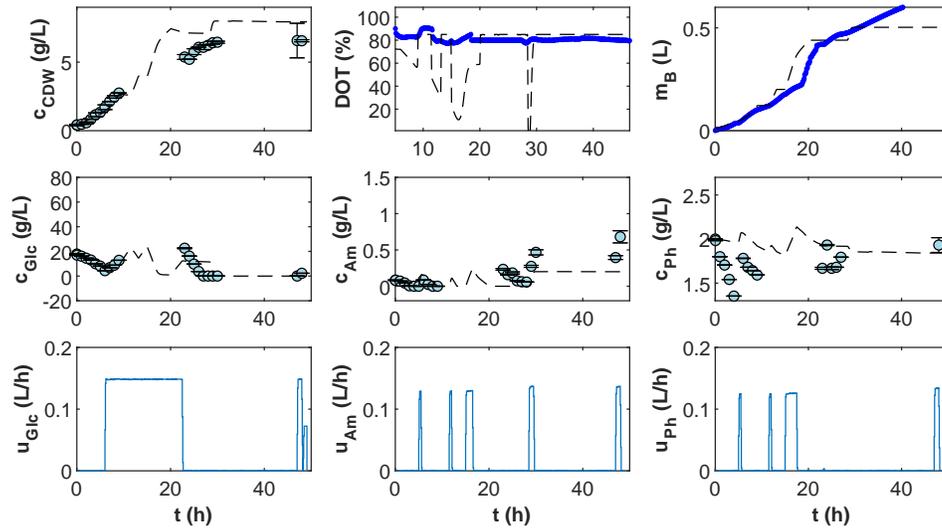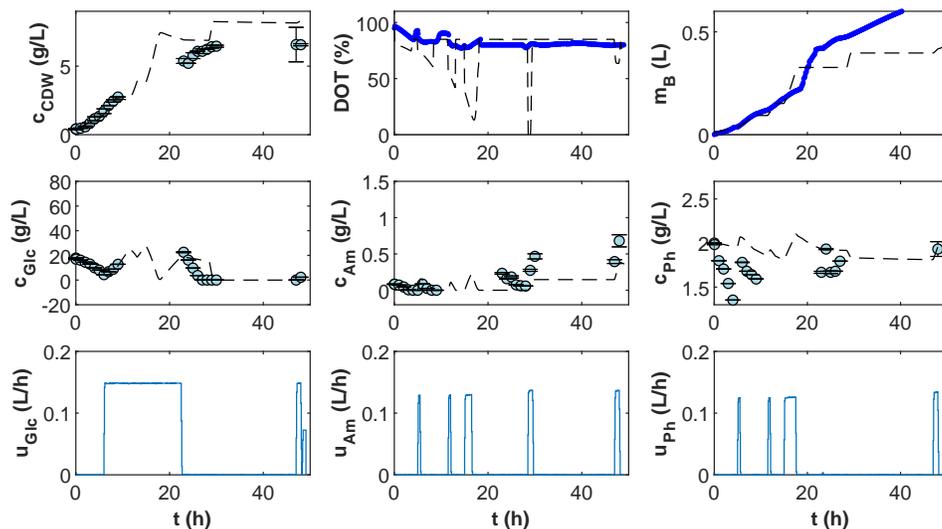
Figure S8: Cross-validation (F2) using WS-CVQ-10 is shown with all samples represented as green dots with error bars. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The PI cross-validation is illustrated by dashed black lines.

Figure S9: Cross-validation (F2) using WS-CVQ-05 is shown with all samples represented as green dots with error bars. The online data is displayed as thick blue lines and the feeding rates are also shown in blue. The PI cross-validation is illustrated by dashed black lines.
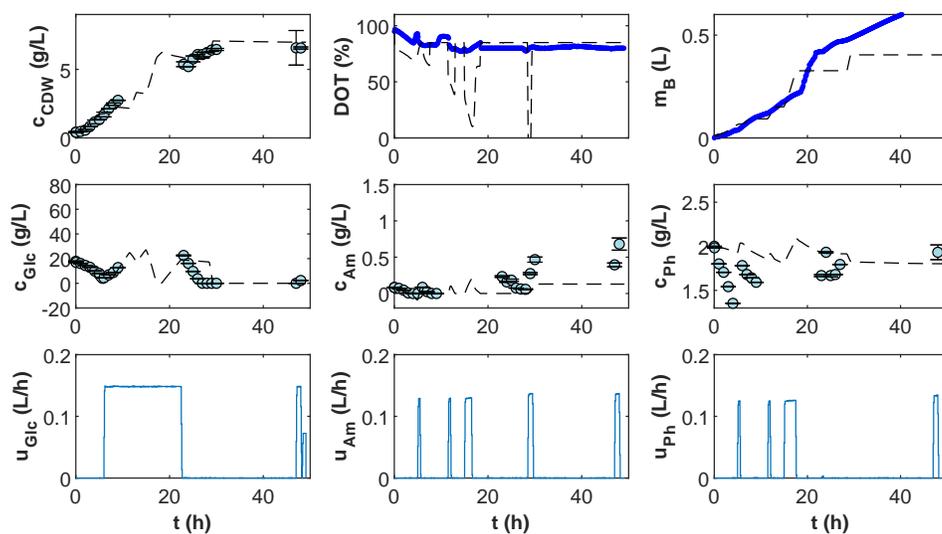
Supplementary Material for: Raman-guided Sample Subset Selection for Cost-efficient Offline Calibration in Bioprocesses

2 Supplementary Table

Table S1: States $\boldsymbol{x}$, measurements $\boldsymbol{y}$, inputs $\boldsymbol{u}$, parameters $\boldsymbol{\theta}$, and constants.

Symbol	Unit	Type	Description
m_{CDW}	g	state	Biomass mass (cell dry weight)
m_{Glc}	g	state	Glucose mass
m_{Am}	g	state	Ammonium mass
m_{Ph}	g	state	Phosphate mass
DOT	%	state	Dissolved oxygen tension
m_B	L	state	total volume of consumed base
V	L	state	Broth volume
$c_i = m_i/V$	$\mathrm{g\,L^{-1}}$	meas.	Concentrations ($i \in \{\mathrm{CDW}, \mathrm{Glc}, \mathrm{Am}, \mathrm{Ph}\}$)
$\boldsymbol{y} = [c_{\mathrm{CDW}}, c_{\mathrm{Glc}}, c_{\mathrm{Am}}, c_{\mathrm{Ph}}, DOT, m_B]^\top$	–	meas.	Measurement vector
$u_{\mathrm{Glc}}, u_{\mathrm{Am}}, u_{\mathrm{Ph}}, u_B, u_A, u_{\mathrm{sample}}$	$\mathrm{L\,h^{-1}}$	input	Feed/base/acid/sample flow rates
$c_{\mathrm{Glc,feed}}, c_{\mathrm{Am,feed}}, c_{\mathrm{Ph,feed}}$	$\mathrm{g\,L^{-1}}$	const.	Substrate concentrations in feed
r_{CDW}	$\mathrm{h^{-1}}$	rate	Specific growth rate
r_{Glc}	$\mathrm{h^{-1}}$	rate	Substrate limitation term
r_{Maint}	$\mathrm{h^{-1}}$	rate	Maintenance term
$\mu_{\mathrm{Glc,max}}$	$\mathrm{h^{-1}}$	param	Max. specific growth rate
$\mu_{\mathrm{Maint,max}}$	$\mathrm{h^{-1}}$	param	Max. specific maintenance rate
$Y_{\mathrm{Glc/CDW}}$	$\mathrm{g\,g^{-1}}$	param	Glucose consumption per biomass produced
$Y_{\mathrm{Am/CDW}}$	$\mathrm{g\,g^{-1}}$	param	Ammonium consumption per biomass produced
$Y_{\mathrm{Ph/CDW}}$	$\mathrm{g\,g^{-1}}$	param	Phosphate consumption per biomass produced
$Y_{\mathrm{O/Glc}}$	$\mathrm{g\,g^{-1}}$	param	Oxygen consumed per glucose utilized
$Y_{\mathrm{B/Am}}$	$\mathrm{g\,g^{-1}}$	param	Base formation per ammonium consumed
$k_{\mathrm{m,Glc}}$	$\mathrm{g\,L^{-1}}$	param	Monod constant for glucose uptake
$k_{\mathrm{m,Maint}}$	$\mathrm{g\,L^{-1}}$	param	Monod constant for maintenance metabolism
$k_{\mathrm{L}}a$	$\mathrm{h^{-1}}$	param	Volumetric oxygen transfer coefficient
DOT^*	% air sat.	const.	Gas–liquid equilibrium DOT (calibrated as 100%)
H	$\mathrm{\%\,L\,g^{-1}}$	const.	Henry constant
$\boldsymbol{\theta}$	–	param	Parameter vector
Φ_{PI}	–	cost	PI objective (modified MLE)
$W_{k,l}$	–	weight	Optional weighting for sampling heterogeneity
$\tilde{C}_{k,l}$	–	cov.	Measurement covariance used in PI

$\hat{\boldsymbol{\theta}}$	No-CVQ	FS-CVQ-10	FS-CVQ-05	WS-CVQ-10	WS-CVQ-05
$\mu_{\text{Glc,max}}$	0.42	0.35	0.31	0.35	0.33
$\mu_{\text{Maint,max}}$	0.19	0.02	0.07	0.12	0.06
$Y_{\text{Glc/CDW}}$	10.78	11.53	11.06	10.87	12.77
$Y_{\text{Am/CDW}}$	0.09	0.11	0.11	0.1	0.14
$Y_{\text{Ph/CDW}}$	0.04	0.04	0.03	0.03	0.05
$Y_{\text{O/Glc}}$	0.1	0.13	0.12	0.12	0.16
$Y_{\text{B/Am}}$	0.05	0.04	0.04	0.04	0.03
$k_{\text{m,Glc}}$	8.34	4.59	5.32	5.44	1.32
$k_{\text{m,Maint}}$	11.0	2.02	2.34	2.06	2.61
$k_{\text{L}}a$	532	405	524	513	571

Table S2: Model parameters ($\hat{\boldsymbol{\theta}}$) optimized on F1 data appyling No-CVQ, FS-CVQ-10, FS-CVQ-05, WS-CVQ-10, and WS-CVQ-05, rounded to significant figures.