

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE  
FAKULTA ELEKTROTECHNICKÁ



## DIPLOMOVÁ PRÁCE

Online detekce klíčových parametrů  
fermentačního výrobního procesu

Praha, 2014

Autor: Bc. Jan Kohout

## Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne \_\_\_\_\_

## Poděkování

Moc rád bych tímto velmi poděkoval vedoucímu mé práce Jiřímu Ciglerovi za jeho obětavou pomoc, pochopení pro mé optimalizační vrtochy, neutuchající podporu a množství času, který mi věnoval.

Dále bych chtěl poděkovat Janu Širokému, Marku Kulvejtovi, Petru Muchnovi, Pavlovi Pokornému, kolegům z VŠCHT a pracovníkům VUAB Pharma a. s.

Velký dík patří i Zdeňku Blažkovi. V neposlední řadě děkuji své rodině a svým přátelům za jejich podporu.

Pro mě osobně měla tato práce velký přínos. Naučil jsem se lépe oddělovat práci a osobní život, vidět, kdy hledání řešení nějakého detailu není podstatné a také více odpočívat. Zároveň jsem pronikl více do tajů systému Linux a velmi zajímavého programovacího jazyka Python.

## Abstrakt

Tato práce se zabývá monitorováním fermentačních procesů. Aplikační část bude prakticky využita pro monitorování a výzkum kultivace při výrobě Nystatinu ve firmě VUAB Pharma a. s. Práce stručně popisuje problematiku bioprocесů. Věnuje se dále přehledu možností a možných omezení při monitoringu fermentačních výrobních procesů. Ukazuje využití PCA analýzy pro získávání informací o fermentačním procesu. V další části se věnuje popisu implementace komunikačního rozhraní pro realizaci monitoringu. V neposlední řadě přináší analýzu přínosu online monitoringu klíčových parametrů pro řízení bioprocесů.

Klíčová slova: bioprocесy, fermentace, monitoring, MATLAB, Python, online

## Abstract

This work studies the monitoring of fermentation processes. The application part will be practically used for monitoring and research culture in the production of Nystatin in company VUAB Pharma a.s. The paper briefly describes the problem of bioprocesses. He is also a list of options and possible limitations in monitoring the fermentation production processes. It shows the use of PCA analysis for obtaining information about the fermentation process. The next section is devoted to a description of a communication interface for monitoring implementation. Finally, it presents an analysis of the benefits of online monitoring of key parameters for the control of bioprocesses.

Keywords: bioprocess, fermentation, monitoring, MATLAB, Python, online

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: **Bc. Jan Kohout**

Studijní program: Kybernetika a robotika  
Obor: Systémy a řízení

Název tématu: **Online detekce klíčových parametrů fermentačního výrobního procesu**

Pokyny pro vypracování:

1. Seznamte se s metodami statistického monitorování fermentačních procesů s důrazem na metody založené na principal component analysis (PCA)
2. Pro průmyslovou fermentační výrobu implementujte program pro monitorování klíčových parametrů ovlivňujících výtěžnost daného procesu. Přístup k samotnému procesu zajistí vedoucí práce.
3. Implementujte rozhraní mezi výpočetním nástrojem pro monitorování a řídicím systémem procesu (RcWare řídicí systém), tak aby bylo možné klíčové indikátory prezentovat obsluze procesu.
4. Analyzujte přínosy zobrazování online indikátorů obsluze a odhadněte vliv na produktivitu fermentačního procesu.

Seznam odborné literatury:

- [1] Lee, Jong-Min, Chang Kyoo Yoo, and In-Beum Lee. "Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis." *Journal of Biotechnology* 110.2 (2004): 119-136.
- [2] Albert, Sarolta, and Robert D. Kinley. "Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision." *TRENDS in Biotechnology* 19.2 (2001): 53-62.
- [3] Lennox, B., et al. "Process monitoring of an industrial fed&#8208;batch fermentation." *Biotechnology and Bioengineering* 74.2 (2001): 125-135.

Vedoucí: Ing. Jiří Cigler, Ph.D.

Platnost zadání: do konce letního semestru 2014/2015

prof. Ing. Michael Šebek, DrSc.  
vedoucí katedry



prof. Ing. Pavel Ripka, CSc.  
děkan

V Praze dne 3. 2. 2014

# Obsah

<b>Seznam obrázků</b>	<b>ix</b>
<b>Seznam tabulek</b>	<b>xi</b>
<b>1 Úvod</b>	<b>1</b>
1.1 Motivace a cíle . . . . .	2
1.2 Struktura práce . . . . .	3
1.3 Omezení . . . . .	3
<b>2 Bioproceny</b>	<b>4</b>
2.1 Názvosloví . . . . .	5
2.1.1 Dělení podle produkce . . . . .	5
2.1.2 Dělení podle fází . . . . .	5
2.1.3 Dělení podle typu kultivace . . . . .	7
2.2 Bioreaktory . . . . .	7
2.2.1 Sledované veličiny . . . . .	8
2.3 Výroba Nystatinu . . . . .	11
<b>3 Monitorování bioprocenů</b>	<b>13</b>
3.1 Typy měření . . . . .	13
3.2 Statistické monitorování bioprocenů . . . . .	14
3.3 Přehled metod statistického monitorování . . . . .	16
3.3.1 Statistický model . . . . .	18
3.3.2 PCA . . . . .	18

3.3.3	Faktorová analýza . . . . .	19
3.3.4	MPCA (multiway PCA) . . . . .	19
3.3.5	MICA (multiway independent component analysis) . . . . .	19
3.3.6	MPLS (multiway partial least square) . . . . .	20
3.3.7	LPP (locality preserving projection) . . . . .	20
3.3.8	Metody umělé inteligence . . . . .	20
3.4	Zhodnocení . . . . .	20
<b>4</b>	<b>Sledování klíčových parametrů online</b>	<b>21</b>
4.1	Analýza klíčových ukazatelů . . . . .	21
4.2	PCA analýza . . . . .	22
4.2.1	Dosažené výsledky . . . . .	22
4.2.1.1	Anotace související s Nystatinem . . . . .	22
4.2.1.2	Anotace související s cukry . . . . .	23
4.2.1.3	Anotace související s kultivací . . . . .	26
4.3	Online monitoring . . . . .	29
4.3.1	Rychlost poklesu viskozity . . . . .	29
4.3.2	Rychlost nárůstu produktu . . . . .	30
4.3.3	Délka kultivace . . . . .	30
4.3.4	Finální množství produktu . . . . .	30
4.4	Ukázka odhadů . . . . .	31
<b>5</b>	<b>Realizace rozhraní</b>	<b>34</b>
5.1	Architektura . . . . .	34
5.1.1	Komunikační server . . . . .	35
5.1.2	Workers . . . . .	35
5.1.3	SCADA . . . . .	35
5.2	Implementace . . . . .	36
5.2.1	Komunikační server . . . . .	37
5.2.1.1	Fronta úloh . . . . .	37



5.2.1.2	Lookup tabulka funkcí . . . . .	38
5.2.1.3	Stavy úlohy . . . . .	38
5.2.2	Workers . . . . .	40
5.2.2.1	Realizace přenosu dat . . . . .	40
5.2.2.2	Synchronizace dat z laboratoře . . . . .	41
5.2.2.3	Python worker . . . . .	42
5.2.2.4	MATLAB worker . . . . .	42
5.2.3	SCADA . . . . .	43
5.3	API . . . . .	43
5.3.1	Formát dat . . . . .	44
5.3.2	Zadání úlohy . . . . .	44
5.3.3	Výpočet úlohy . . . . .	45
5.3.4	Výsledek úlohy . . . . .	46
<b>6</b>	<b>Vyhodnocení vlivu a přínosů</b>	<b>47</b>
6.1	Přesnost odhadů . . . . .	47
6.1.1	Zkušenosti obsluhy . . . . .	51
<b>7</b>	<b>Závěr</b>	<b>52</b>
7.1	Další práce . . . . .	53
	<b>Literatura</b>	<b>55</b>
<b>A</b>		<b>I</b>
A.1	Anotace . . . . .	I
A.2	Příklad testovacích dat . . . . .	IV
<b>B</b>	<b>Obsah příloženého CD</b>	<b>V</b>

# Seznam obrázků

1.1	Ukázka stávajícího GUI. . . . .	2
2.1	Ilustrace fází bioprocessu. . . . .	6
2.2	Přehled typů kultivace. . . . .	7
2.3	Typický průmyslový fermentor. . . . .	8
2.4	Ukázka průběhu limitace kyslíkem. . . . .	11
3.1	Ukázka online měřených dat. . . . .	14
3.2	Ukázka laboratorně měřených dat. . . . .	15
3.3	Princip unfoldingu. . . . .	17
4.1	Vlastní čísla anotací souvisejících s Nystatinem. . . . .	23
4.2	PCA vytipovaných anotací souvisejících s Nystatinem. . . . .	24
4.3	Vlastní čísla anotací souvisejících s cukry. . . . .	24
4.4	PCA vytipovaných anotací souvisejících s příkrmy (2. a 3. komponenta). . . . .	25
4.5	PCA vytipovaných anotací souvisejících s příkrmy (1. a 2. komponenta). . . . .	26
4.6	PCA vytipovaných anotací souvisejících s příkrmy (1. a 3. komponenta) . . . . .	27
4.7	Vlastní čísla anotací souvisejících s kultivací. . . . .	27
4.8	Anotace související s kultivací. . . . .	28
4.9	Průběh viskozity a Nystatinu jedné šarže. . . . .	29
4.10	Průběh odhadu délky kultivace. . . . .	31
4.11	Průběh odhadu poklesu viskozity. . . . .	32
4.12	Průběh rychlosti nárůstu Nystatinu. . . . .	32
4.13	Průběh odhadu Nystatinu na konci kultivace. . . . .	33

5.1	Nezávislá implementace výpočetních nástrojů. . . . .	35
5.2	Komunikace uvnitř systému. . . . .	36
5.3	Propojení modulu Apache2 s Pythonem. . . . .	37
5.4	Schema možných statusů úlohy. . . . .	39
5.5	Schema práce workera. . . . .	40
5.6	Proces laboratorního měření a jeho napojení na výpočetní systém. . . . .	41
5.7	Nové GUI pro zobrazování údajů z výpočtů. . . . .	43
6.1	Přesnost odhadu ukazatele rychlosti poklesu viskozity. . . . .	48
6.2	Přesnost ukazatele konce kultivace. . . . .	48
6.3	Přesnost ukazatele rychlosti produkce. . . . .	49
6.4	Přesnost ukazatele koncové produkce. . . . .	49

# Seznam tabulek

3.1	Příklad offline dat. . . . .	16
5.1	Databázový záznam ve frontě úloh. . . . .	38
5.2	Předávané parametry úlohy. . . . .	44
5.3	Přehled HTTP požadavků. . . . .	44
A.1	Přehled anotací. . . . .	I
A.2	Přehled skupin anotací pro PCA analýzu. . . . .	II

# Kapitola 1

## Úvod

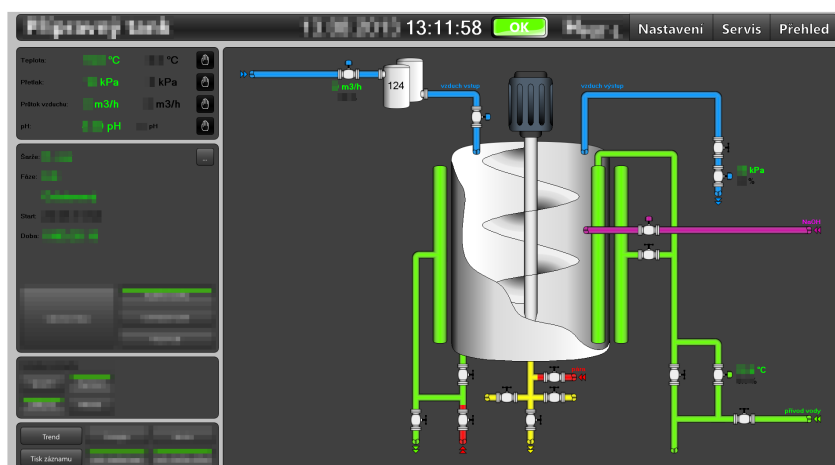
Průmyslová výroba, ve které se objevují živé mikroorganismy, patří z hlediska monitoringu a řízení k obtížnějším aplikacím řídicí techniky. Malé změny parametrů systému mívají za následek velké změny v celém procesu, které mohou být někdy z hlediska výstupního produktu přímo fatální.

Aplikační část této práce je úzce spjata s výrobou antibiotika Nystatinu, které vyrábí firma VUAB Pharma a. s. Ta se výrobou antibiotik zabývá již od roku 1949 [23]. V roce 1966 proběhlo sloučení dvou subjektů a vznikl Výzkumný ústav antibiotik a biotransformací (odtud dnešní název VUAB). Postupem času společnost pronikla na trh USA a od poloviny roku 2004 je společnost členem skupiny Safina.

Samotný proces výroby je v současné době řízen poloautomaticky. Značná část rozhodovacího procesu leží na zkušenosti obsluhy a vnitřních předpisech. To v sobě skrývá výhodu i nevýhodu - nesporná výhoda je, že člověk s dlouholetou zkušeností se již hodně řídí jakousi intuicí a je schopen velmi efektivně vývoj procesu odhadovat. Nevýhodou je, že tento nedeterminismus představuje určitý problém z hlediska automatizace a řízení. Stejně tak může být zdrojem většího rozptylu výsledného produktu z jednotlivých šarží.

Tato práce je dílčí část víceletého projektu TAČR [22] (na kterém spolupracuje mimo VUAB Pharma a. s. také ČVUT, VŠCHT a Energocentrum Plus s.r.o.), který má za cíl snížení variability výroby. V roce 2013 v jeho rámci probíhala analýza charakteristických ukazatelů (tzv. anotací), které charakterizují probíhající fermentaci. Byl zkoumán jejich vliv na samotný proces a to, jak se vzájemně ovlivňují. Na základě těchto analýz byly vytipovány klíčové ukazatele, které nejvíce korelují s výsledným výstupem (množstvím produktu).

Postupně tak vyvstal požadavek na zobrazení některých výsledků výzkumu obsluze na řídicím pracovišti výroby. K tomu bylo nutné vyřešit dostupnost laborator-



Obrázek 1.1: Ukázka stávajícího GUI. Sledují se běžné provozní parametry (teplota, tlak, pH, průtok vzduchu...). Ilustrace sledovaných parametrů, skutečné hodnoty úmyslně rozostřeny.

Zdroj: VUAB Pharma a. s.

ních měření v co nejkratším čase od samotné analýzy a zpřístupnit a propojit tato data s těmi, které se měří online. Nakonec zastřešit vše vhodnou komunikační strukturou, která umožní vizualizaci online výpočtů nad takto propojenými daty.

## 1.1 Motivace a cíle

Hlavní motivací této práce je přinést komunikační rozhraní, které umožní využívání laboratorních offline dat pro online výpočty, měření a vizualizaci klíčových parametrů během probíhající fermentace. Nástroj, který bude připraven na další rozšiřování o potřebné výpočty nad daty, které jistě v rámci dalšího studia procesu ještě vyplynou.

Snahou je přinést další informace o probíhající fermentaci a rozšířit podporu rozhodování obsluhy. Stávající GUI (obrázek 1.1) zobrazuje pouze provozní ukazatele a většina rozhodování je tak na bedrech a letitých zkušenostech obsluhy. Využíváno je přitom laboratorních měření a osobního dohledu nad průběhem kultivace.

## 1.2 Struktura práce

Druhá kapitola slouží k seznámení čtenáře s bioprocesy a veličinami, které je nejvíce charakterizují. Třetí kapitola obsahuje přehled statistického monitorování fermentačních procesů. Čtvrtá kapitola se věnuje offline PCA analýze a popisu jednoduchého rozhraní pro monitoring klíčových parametrů procesu, ovlivňujících výtěžnost daného procesu.

V páté kapitole je popsána praktická realizace komunikačního rozhraní mezi výpočetním nástrojem pro monitorování a řídicím systémem procesu (RcWare řídicí systém), aby bylo možné klíčové indikátory prezentovat obsluze procesu. Šestá kapitola obsahuje shrnutí a analýzu přínosů zobrazování online indikátorů obsluze a odhad vlivu na produktivitu fermentačního procesu.

## 1.3 Omezení

Jelikož se jedná o výrobní tajemství firmy VUAB Pharma a. s., jsou veškeré průběhy na obrázcích normovány do rozsahu  $\langle 0, 1 \rangle$ .

Pozornému čtenáři jistě také neunikne, že některé formulace jsou velmi obecné. Avšak vzhledem k mlčenlivosti o tajemství výroby bohužel v některých případech skutečně není možné zacházet do větších detailů, i když s konkrétními průběhy, principy či hodnotami samozřejmě seznámeni jsme.

# Kapitola 2

## Bioproceny

Bioproceny obecně je označován výrobní děj, kdy vlivem působení mikroorganismů dochází ke vzniku či přeměně látek. Tyto mikroorganismy mohou být například různé kmeny kvasinek nebo například bakterií.

Bioproceny dnes najdeme v řadě výrobních procesů – v potravinářství (výroba etanolu, piva, vína, destilátů, pekařského droždí a krmné biomasy), medicíně (výroba antibiotik, enzymů, organických kyselin, aminokyselin, vitaminů, růstových faktorů, steroidních hormonů) [9]. Jsou využívány v mnoha dalších odvětvích průmyslu, zemědělství či ekologie.

Díky rozvoji molekulární biologie, genetického a proteinového inženýrství, fermentačních technik, bioinženýrství, separačních a purifikačních technik atd. se technologie stále rozvíjejí. Velké budoucnost patrně patří vývoji bezodpadových technologií či třeba ochraně životního prostředí.

Biologická povaha s sebou přináší ne zcela deterministické chování celého systému. Malá změna v parametrech prostředí může vést k velkým změnám v chování kultury. Živá povaha mikroorganismů v systému totiž přináší velmi citlivou závislost na byť jen malých změnách v životních podmínkách. [7] Bioproceny jsou tedy typicky silně nelineární systémy a jejich vlastnosti nejsou časově invariantní.

Zvídavé čtenáře, které by zajímalo více aplikací bioprocenů, než je zdejší stručný přehled, odkážeme například na [14].



## 2.1 Názvosloví

*Fermentace* (kvasný proces) je jedním z biotechnologických procesů, kterým se budeme zabývat dále.

*Kultivací* pak budeme v dalším textu označovat samotný biologický proces, během kterého dochází k udržování a zejména růstu požadovaných mikroorganismů.

*Šarže* je označení jedné kompletní kultivace.

### 2.1.1 Dělení podle produkce

Obecně můžeme bioproceny rozdělit do dvou velkých skupin [9] podle toho, na jaký produkt se zaměřují:

**produkce biomasy** – hlavním produktem je biomasa (tedy látky tvořící těla organismů).

Příkladem může být výroba pekařského droždí či kultivace řas pro krmné účely.

**produkce metabolitů** – hlavním produktem je vedlejší metabolit mikroorganismu. Dále rozlišujeme *primární* (metabolit vzniká současně s růstem kultury) či *sekundární* (metabolit vznikající ve stacionární fázi, tj. když je již mikroorganismus dostatečně narostlý a množení kultury již probíhá minoritně).

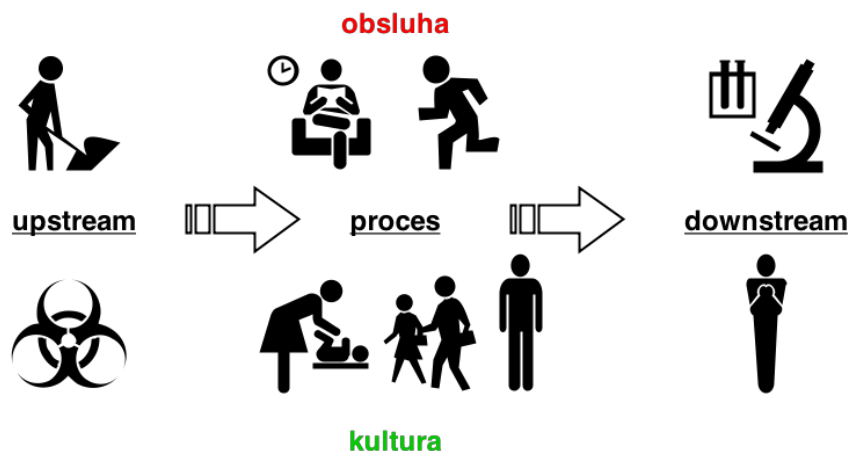
Příkladem mohou být různé typy antibiotik či třeba výroba lihovin.

### 2.1.2 Dělení podle fází

Bioproceny z hlediska jednotlivých fází procesu lze rozdělit do následujících tří skupin:

**Upstream processing** – probíhá sterilizace fermentoru, příprava média (substrátu), aeračních plynů, živin. Důležitá je vysoká sterilita, v této fázi je systém totiž vysoce citlivý na kontaminace.

**Biologický proces** – probíhá příprava *inokula* (násada, jakási „sazenička“ celé fermentace) a jeho adaptace na provozní podmínky. V této fázi potom probíhá samotná fermentace, za působení mikroorganismů dochází k přeměně chemických látek.



Obrázek 2.1: Ilustrace fází bioproduktu.

Zdroj: autor

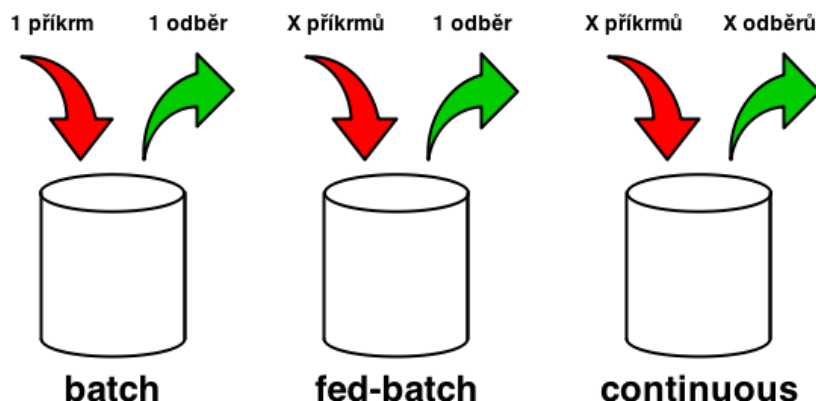
Při realizaci bioproduktů v průmyslovém měřítku se nejčastěji setkáváme se **submerzními** procesy (probíhajícími v kapalině, např. ve vodě). Kapalina je typicky sycena potřebnými plyny.

Procesy vyžadující kyslík nazýváme *aerobní*, pokud kyslík mikroorganismy ke svému životu nepotřebují, nazýváme pak takové procesy *anaerobní*.

Mimo submerzních kultivací existují i kultivace, při kterých mikroorganismy rostou na povrchu nějakého média (ať již pevného či kapalného).

**Downstream processing** – oddělení produktu z fermentační směsi, izolace a purifikace (chemické čištění) výsledného produktu. Často zde vzniká množství odpadních látek z proběhlých procesů, které se již dále nezpracovávají.

Na obrázku 2.1 jsou popisované fáze bioproduktu názorně ilustrovány s jejich vlivem na práci obsluhy. *Upstream* znamená přípravu sterilního prostředí pro kulturu, pro obsluhu to znamená mnoho práce s přípravou substrátu a fermentoru. *Proces* znamená zrození kultury a její růst. Pro obsluhu znamená sledování procesu, ale místy i čekání na odezvy systému (např. při změně provozních parametrů, přidání příkrmu apod.). *Downstream* většinou pro kulturu znamená její smrt. Pro obsluhu potom vytěžování a purifikaci požadovaného produktu.



Obrázek 2.2: Přehled typů kultivace.

Zdroj: autor

### 2.1.3 Dělení podle typu kultivace

V praxi existují tři typy kultivace, které se liší hlavně v strategii přidávání živin (příkrmu – více viz kapitola 2.2.1). Možností je také kombinace přístupů, jako je tomu v případě sledované výroby Nystatinu.

**batch (vsádková)** – nedochází k přidávání příkrmu, fermentace běží do doby, než se mikroorganismus vyčerpá.

**fed-batch (přítoková)** – fermentor je na začátku naplněn určitým množstvím substrátu a v průběhu kultivace dochází k postupnému doplňování živin.

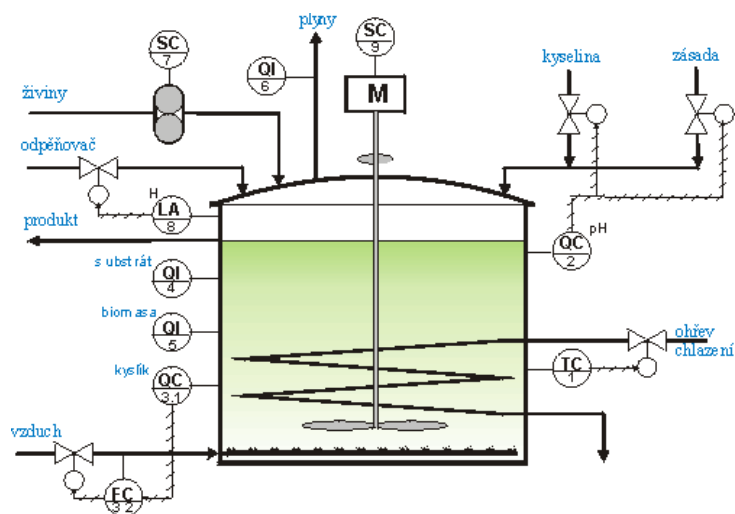
**continuous (kontinuální)** – živiny jsou do systému přidávány nepřetržitě za stálého odebírání kultivačního média s produktem.

Rozdíl mezi jednotlivými typy kultivace je přehledně shrnut na obrázku 2.2.

## 2.2 Bioreaktory

Samotný bioprocés probíhá typicky v **bioreaktoru** (v případě fermentace se označuje jako *fermentor*), což je zařízení, které je schopné udržovat pro růst a vývoj organismů příčinné podmínky.

Během průmyslové výroby dochází k přestupu hmoty skrze několik různě velkých fermentorů (typicky malý fermentor pro nárůst *inokula*, očkovací směsi, která se



Obrázek 2.3: Typický průmyslový fermentor.

Zdroj: <http://uprt.vscht.cz/kminekm/mrt/F3/F3k39-o336.gif>

aplikuje do většího, kde dochází k jeho zmnožení a následně je přesunut do výrobního, fermentačního tanku, který může mít objem již několik desítek  $m^3$ ).

Na obrázku 2.3 je vidět typický představitel průmyslového fermentoru s popisem jeho nejdůležitějších funkčních částí. Ve fermentoru je možné udržovat určité pH (kyseliny, zásady), teplotu (ohřev/chlazení), aeraci (kyslík, vzduch), množství látek pro podporu růstu (živiny, substrát), případně sledovat množství biomasy a pěny, která fermentací vzniká. Mísení směsi je zajišťováno rotačním míchadlem.

### 2.2.1 Sledované veličiny

Bioprocés, jak již bylo řečeno výše, je velmi komplexní děj, na jehož chování má vliv řada interních i externích činitelů (teplota okolí, roční období...). My se v našem přehledu zaměříme zejména na veličiny, které se přímo sledují nejčastěji:

**teplota** – většinou se udržuje konstantní či v určitém úzkém rozsahu, neboť mikroorganismy jsou na teplotní změny vysoce citlivé. Na teplotu uvnitř fermentoru má vliv například aktivita mikroorganismů a ohřev či chlazení fermentoru.

Měření probíhá většinou online, teplotními senzory.

Regulační zásahy se konají ohřevem či chlazením. Regulací je nutno teplotní rozsah udržovat prakticky konstantní.

**tlak** – tlak uvnitř fermentoru.

Měření probíhá většinou online, tlakovými čidly.

Regulace je možná například změnou množství plynů ve fermentoru.

**koncentrace biomasy** – měření koncentrace biomasy (těla mikroorganismů) není triviální a ve výrobě se z tohoto důvodu přistupuje k odběrům vzorků a laboratornímu měření. Měření nárůstu biomasy je podstatné pro sledování samotného růstu a stavu mikroorganismů. Z koncentrace biomasy lze například vyvozovat i aktuální životní fázi organismů.

V případě studovaného procesu probíhá offline v laboratoři. I když byla snaha toto měření automatizovat a koncentraci biomasy odhadovat na základě měření elektrické kapacity, výsledky nebyly uspokojivé.

Koncentrace biomasy souvisí s životními cykly mikroorganismů. Nedá se tedy přímočaře regulovat.

**objem** – zaplnění fermentoru fermentační směsí.

Měření probíhá většinou online, např. bezkontaktními sondami na základě výšky hladiny, přepočtem na základě tlaku apod. Potíž může nastat v případě pění (což je ve fermentačních procesech častý jev), když použitá čidla nemusí ukazovat skutečnou hodnotu objemu samotné fermentační směsi.

V případě překročení žádaného objemu nebo silném pění lze přistoupit k tzv. *odtahům*, kdy se část fermentační směsi odebere.

Objem souvisí s množstvím hmoty ve fermentoru, s příkrmy a růstem mikroorganismů.

**produktivita, množství produktu** – měří se obvykle laboratorně. Jedná se o přepočet množství produktu na jednotku objemu, hmoty či jeho koncentrace, za jednotku času, případně množství produktu v definovaných jednotkách. Dle aplikace.

Měří se většinou laboratorně.

Řízení je velmi obtížné. Jedná se o hlavní cíl monitoringu a řízení bioprocusů.

**pH** – důležitá veličina, která se v řadě aplikací udržuje v určitém povoleném rozsahu. Velmi ovlivňuje životní prostředí mikroorganismů. Ovlivňují jej příkrmy i samotné metabolické procesy probíhající v mikroorganismech. Hodnota pH je také spjata s teplotou.

Měření probíhá většinou online, např. pH senzory.

Řídí se obvykle na určitém pro mikroorganismy žádoucím intervalu, například přidáváním kyselých či zásaditých přísad.

**příkrm (koncentrace)** – živiny, které udržují mikroorganismus v růstu či produkční fázi. Ve fed-batch kultivaci je jej třeba pravidelně doplňovat.

Měření probíhá většinou online (měřením množství, objemu či hmotnosti přidaného materiálu).

Řízení probíhá například průběžným přidáváním materiálu s příkrmem (např. různé druhy sirupů, sacharidů v pevném skupenství apod.). Buď v pevně daných intervalech nebo na základě koncentrace biomasy či produktivity.

**substrát (koncentrace)** – živná půda, která je pro přežití mikroorganismů zásadní hlavně v první fázi kultivace.

Měření probíhá většinou na základě laboratorních rozborů.

Řízení je základním know-how při zakládání kultivace. Substrát musí mít definované množství a složení.

**viskozita** – viskozita ovlivňuje prostup kyslíku k mikroorganismům i výslednou extrakci produktu. Ovlivňuje ji podíl různých hmot ve fermentoru a zejména činnost mikroorganismů.

Měření probíhá offline v laboratoři.

Vzhledem k tomu, že jde hlavně o výsledek činnosti mikroorganismů, řízení opět není přímočaré.

**koncentrace rozpuštěného  $O_2$ ,  $CO_2$**  – koncentrace kyslíku a oxidu uhličitého (v případě aerobních mikroorganismů), případně další plyny. Podstatná je zejména aerace fermentační směsi, mikroorganismy totiž ke svému životu využívají hlavně rozpuštěné plyny. Proto je nezbytně nutné zajistit kvalitní promíchávání směsi.

Měření probíhá většinou online.

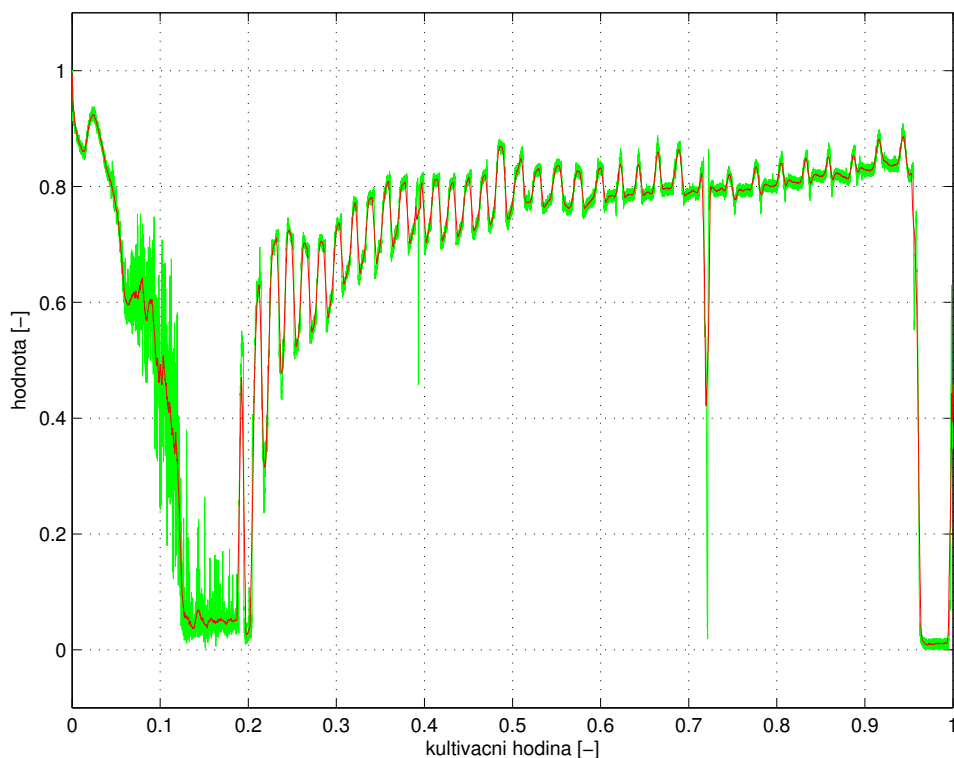
Řízení je možné změnou množství plynu vstupujícího a vystupujícího z fermentoru.

**pěna** – při fermentačních procesech vzniká velmi často množství pěny, která může bránit kvalitnímu vzdušnění a je proto žádoucí její množství ve fermentoru sledovat.

Měření obvykle probíhá online.

Řízení přidáváním přísad proti pění nebo výše zmiňovanými odtahy.

V konkrétním fermentačním procesu může být samozřejmě sledovaných veličin více, dle povahy systému.



Obrázek 2.4: Ukázka průběhu limitace kyslíkem. Zelená čára je průběh rozpuštěného kyslíku, červená klouzavý průměr.

Zdroj: autor

## 2.3 Výroba Nystatinu

Bioprocес v případě produkce Nystatinu probíhá v několika fázích. Jednotlivým fázím odpovídají tři typy tanků: předočkovací, očkovací a fermentační. Samotná produkce Nystatinu vzniká ve fermentačním tanku, proto se naše hlavní pozornost směřuje na něj.

Výroba Nystatinu probíhá kombinací *batch* a *fed-batch* fáze. Batch fáze slouží k nárůstu kultury a často zde dochází k tzv. limitaci kyslíkem (*oxygen limitation*), což je velmi důležitý, ale také velmi špatně ovlivnitelný parametr ovlivňující chování systému v další fázi. Průběh limitace kyslíkem je naznačen na obrázku 2.4. Do přibližně kultivační hodiny 0.2 trvá fáze batch, později pak již mluvíme o fázi fed-batch.

V nejvíce zjednodušeném modelu se na fermentaci budeme dívat jako na silně nelineární systém s mnoha stavy a jedním výstupem (množství finálního Nystatinu). Vstupy je však velmi obtížné přímočaře definovat, už jen z toho prostého faktu, že akční zásahy jsou velmi omezené (celý proces výroby se řídí velmi striktními pravidly vyvinutými na základě dlouhodobé zkušenosti).

V naší práci jsme se zaměřili na odhad a zobrazování parametrů fed-batch fáze,

pro kterou je k dispozici více laboratorních měření.

Pro batch fázi se v současné době vyvíjí model, který podrobnější popis batch fáze také umožní.



# Kapitola 3

## Monitorování bioprocесů

Monitorování bioprocесů patří k poměrně složitým výzvám. Vzhledem k biologické povaze bioprocесů vyvstává otázka, jak efektivně měřit potřebné veličiny (viz kap. 2.2.1). Veličiny jako teplota, tlak, objem nepředstavují v praxi z hlediska měření velký problém, ale veličiny úzce spjaté se samotným monitoringem růstu mikroorganismů mohou představovat určitou komplikaci.

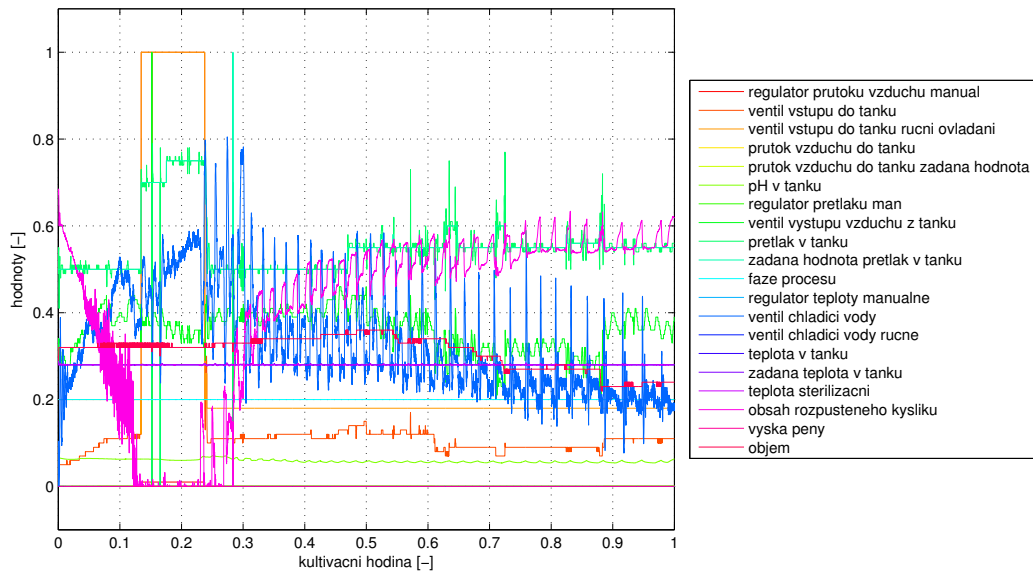
Například pro sledování růstu buněk je možné využít rozličné typy mikroskopie. Existují sice také různé přístupy online monitorování například pomocí tzv. *in-situ* (z lat., na místě) senzorů [2], ale v zavedeném provozu je investice do takovýchto senzorů obtížná. Proto se přistupuje k měření offline – z fermentoru je ručně odebrán vzorek, který se poté analyzuje v laboratoři, například právě optickou mikroskopií či chemickým rozbořením.

Tento postup z hlediska řízení procesu přináší různá zpoždění a v praxi také určitou nepřesnost, protože organismy odebrané z hlavního fermentoru mají po určitém čase trochu odlišné chování. Podobný problém nastává například u měření biomasy, koncentrace živin apod. To do sledování průběhu kultivaci přináší nejistoty (laboratorní měření jsou zpožděná, navíc zde hraje roli i lidský faktor a neopakovatelnost takového měření).

### 3.1 Typy měření

Co se zdrojů dat týče, měli jsme při naší práci dva hlavní zdroje:

**online data** – jsou získávána přímo ze senzorů ve fermentoru. Pomocí sběrnic (záleží na konkrétním hardwarovém řešení) jsou data ze senzorů shromažďována do centrálního úložiště (databáze). Prakticky v reálném čase jsou tak získané informace



Obrázek 3.1: Ukázka online měřených dat.

Zdroj: autor

připraveny k použití. Z databáze mohou být vyčítány pro zobrazení uživateli (například na dispečerském pracovišti).

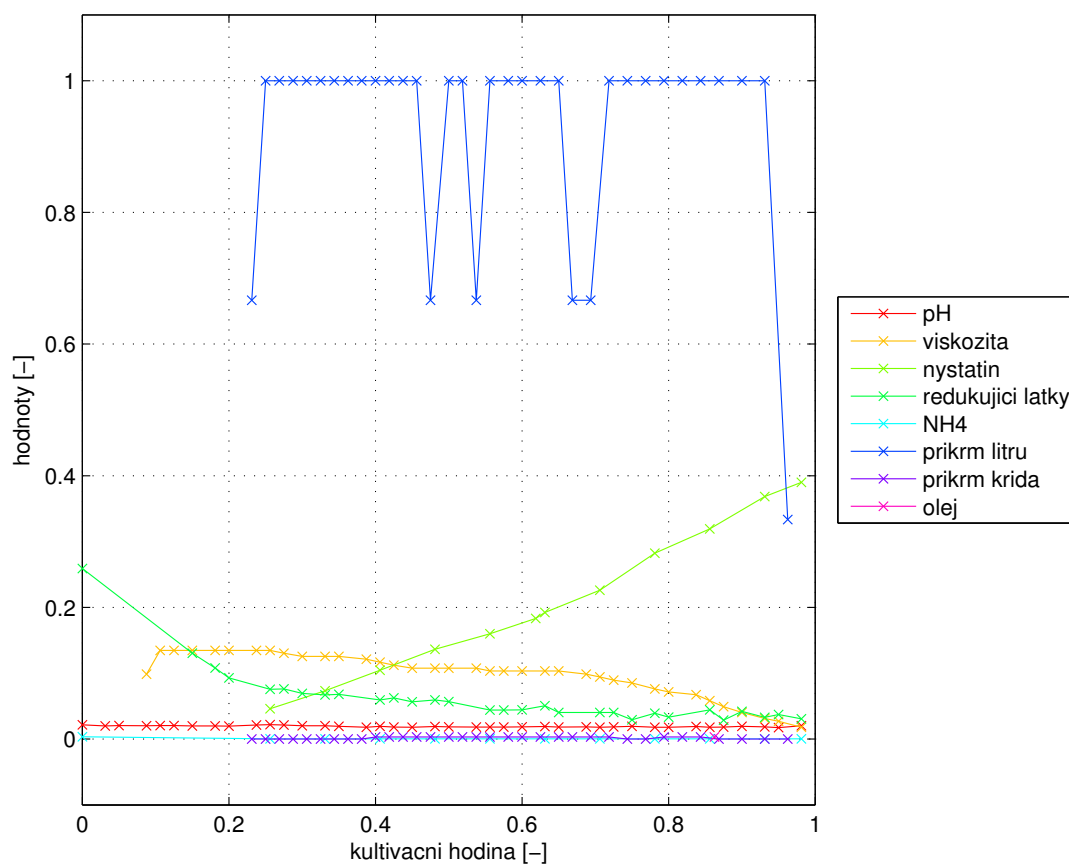
Ukázka online dat z jedné šarže je vidět na obrázku 3.1. Lze snadno nahlédnout, že se jedná zejména o technické procesní parametry.

**offline data** – jsou získávána ručním odběrem vzorků z fermentoru. Vzorky jsou následně dopraveny do laboratoře, kde jsou analyzovány. Tento proces může trvat různě dlouho dobu, dle povahy konkrétní analýzy řádově i hodiny. Následně jsou výsledky analýz zaznamenány do datových souborů.

Příklad offline měřených dat s typickými průběhy je uveden na obrázku 3.2. Popis nejvýznamnějších parametrů obsahuje tabulka 3.1.

## 3.2 Statistické monitorování bioprocésů

Monitorování a řízení bioprocésů je úzce spjato se zkušeností operátorů výroby. Množství procesních dat je ale tak vysoké, že tento úkol je velmi obtížný. Zjednodušení může přinést statistický monitoring dat [1], [5], [13]. Výsledkem mohou být statistiky, upozorňující na možné chyby parametrů procesu či celé expertní systémy.



Obrázek 3.2: Ukázka laboratorně měřených dat.

Zdroj: autor

Tabulka 3.1: Příklad offline dat.

viskozita	hodnota dynamické viskozity směsi
Nystatin	množství produktu ve fermentoru
křída, olej	látky potlačující či snižující tvorbu pěny
příkrm, redukující látky	množství přidaného příkrmu (živin)

Výrobní fermentační procesy v průmyslovém měřítku trvají typicky několik dní. Obvykle běží několik kultivací paralelně a každá kultivace (šarže) obsahuje mnoho dat popisujících průběh procesu (kap. 4.3). Až po jejich proběhnutí je k dispozici informace o úspěšnosti dané šarže (například na základě množství finálního produktu). Využití těchto historických dat pro monitoring a detekci chyb je velmi žádoucí.

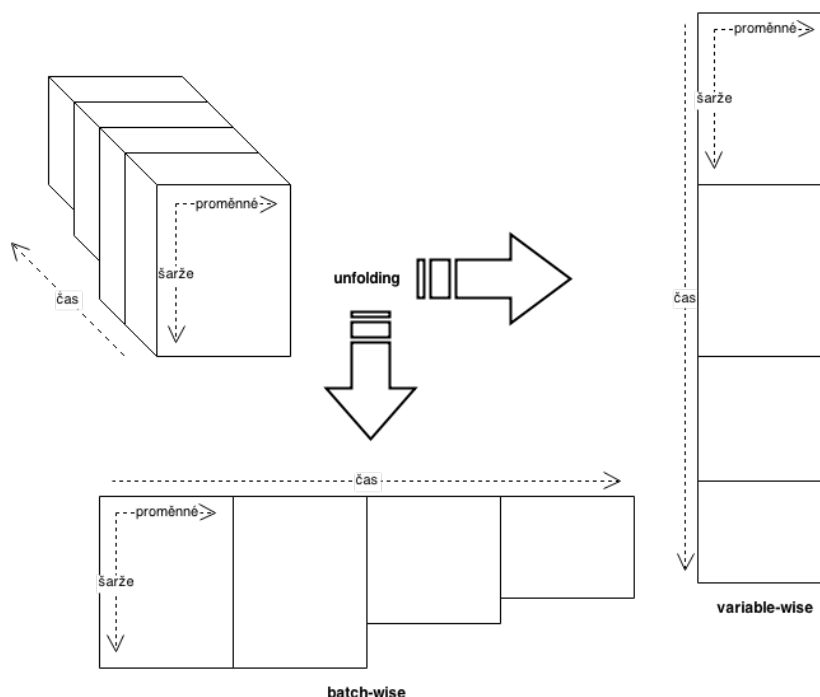
K tomu je ale zapotřebí z historických dat vytěžit podstatné informace. A zde je právě vhodná aplikace metod založených na PCA analýze – tzv. *MSCP* (Multivariate Statistical Process Control) [7]. Jedná se o metody, které jsou schopny zredukovat dimenzi vstupních dat – *vícerozměrná analýza dat*. Na základě apriorní informace z historických dat o úspěšnosti či neúspěšnosti dané šarže je pak vytvořen model, který data statisticky popisuje.

Model lze potom s větší či menší úspěšností využít při online monitoringu kultivace. Na základě statistik (např. *SPE*) je pak možné analyzovat, zda sledovaná šarže vykazuje známky úspěchu či neúspěchu již v průběhu její kultivace. Takto je možné například podchytit možné chybové stavy v procesu a upozornit na ně operátory na velícím pracovišti.

### 3.3 Přehled metod statistického monitorování

Vnímání světa v 3D prostoru, vlastnosti objektů jako teplota, tlak, objem – to vše je pole působnosti vícerozměrné analýze dat. Zejména pak v případě, že máme k dispozici velké množství takových měření.

Data, která je možné v bioprocesech měřit, mají jednorozměrný charakter (například časová osa a jednotlivá měření teploty). Při detailnějším zkoumání ale zjistíme, že jednotlivá měření se často vzájemně silně ovlivňují, a je proto výhodnější dívat se na



Obrázek 3.3: Princip unfoldingu pro rozdílné délky šarží (vlevo nahoře).

Zdroj: autor

ně jako na vícerozměrná data.

Smysl vícerozměrné analýzy oceníme zejména v případech, kdy je k dispozici dostatečné množství měřených hodnot a jednotlivých parametrů (proměnných) je již takové množství, že se již špatně nejen zobrazují, ale i analyzují. Vícerozměrná analýza slouží ke zjednodušení analýz. Snaží se nalézt smysluplné pohledy a řezy mnoharozměrným prostorem dat.

Vícerozměrná analýza však vyžaduje předzpracování dat. V případě fermentačních procesů, kde je k dispozici měření pro více šarží, se na data můžeme dívat jako na trojrozměrnou matici. Šarže (X rozměr) obsahují časovou osu (Y rozměr) a samotné proměnné (Z rozměr). To je vhodné zejména pro online monitoring a detekci chyb [6] ve sledovaných proměnných.

Pro analýzy je však výhodnější 2D pohled na data. Toho je docíleno pomocí tzv. **unfoldingu** (složení matice dat vstupujících do analýz, [3]). V oblasti studia fermentačních procesů jsou obvyklé tři přístupy:

**batch-wise** – popisuje variabilitu mezi jednotlivými kultivacemi. Data jsou převedena do struktury, ve které jsou vytvořeny časové bloky s jednotlivými proměnnými.

**variable-wise** – popisuje variabilitu napříč jednotlivými proměnnými. Poskytuje tak lepší pohled na dynamiku celého procesu.

**time-wise** – popisuje variabilitu napříč časem. V praxi se využívá nejméně [1].

Na obrázku 3.3 je znázorněn princip variable-wise a batch-wise unfoldingu. Všimněme si, že v případě variable-wise unfoldingu dostáváme kompletní 2D matici i v případě rozdílných délek šarží.

Obecně se s chybějícími daty lze vyrovnat [6]: vyplněním nulami, nahrazením interpolovanou (extrapolovanou hodnotou) nebo využitím schopnosti PCA se chybějícími daty se vyrovnat.

### 3.3.1 Statistický model

Statistický model předpokládá soubor měření jednotlivých šarží. Při vytváření modelu je nutné vypořádat se s chybějícími daty a zejména nestejnými délkami kultivace.

### 3.3.2 PCA

PCA (principal component analysis) umožňuje zjednodušit datový prostor tak, že je popsán pomocí tzv. komponent (vlastní vektory) [11]. Jedná se o přepočítání původního prostoru do nových souřadnic, které se snaží co nejvíce podchytit původní data. PCA hledá vlastní vektory a vlastní čísla, vypočítaná nad maticí dat. Prvky vlastních vektorů váhují originální proměnné.

V novém prostoru, který vytvoří hlavní komponenty, popisují vlastní čísla množství rozptylu, který daná komponenta vysvětluje.

Předpoklady: mnohorozměrné normální rozdělení, nezávislost pozorování, proměnné kvantitativní (a musí být možné vypočítat pro ně kovarianci či korelaci).

**Centrovaná PCA** – původní proměnné jsou nejprve centrovány (součet vlastních hodnot kovarianční matice je roven součtu rozptylů proměnných).

**Standardizovaná PCA** – původní proměnné jsou normovány na nulový průměr a jednotkový rozptyl (součet vlastních hodnot matice korelačních koeficientů je roven řádu matice – počtu proměnných).

### 3.3.3 Faktorová analýza

Faktorová analýza vznikla v oblasti psychologie a až v posledních desetiletích začala více pronikat do dalších vědních oborů. Využívá taktéž PCA metodu s tím rozdílem, že je zde předpoklad závislosti proměnných vlivem neměřitelných na pozadí stojících faktorů (tzv. *common factors*).

Společný (hlavní) faktor se faktorová analýza snaží nalézt stejně jako PCA tak, aby vysvětloval co největší část rozptylu dat. Rozdílný je však další krok – faktorová analýza se snaží hlavní faktory rotovat, aby proměnné popisovaly co nejjednodušeji. Rotace faktorů zde slouží k usnadnění jejich interpretace.

Předpokladem je normální rozdělení dat.

### 3.3.4 MPCA (multiway PCA)

Jedná se o rozšíření PCA [6]. Algoritmus MPCA zjednodušuje dimenzi vstupních dat výraznějším způsobem. S použitím statistického modelu z historických dat (tzv. *NOC model* – model získaný z úspěšných šarží) ji lze využít k online monitorování biologických procesů.

Při použití k online monitoringu vykazuje ale velmi špatnou detekci chyb v počátečních fázích kultivace, které bývají ovšem těmi nejpodstatnějšími.

### 3.3.5 MICA (multiway independent component analysis)

Zatímco PCA se snaží zachytit varianci dat zachycením nekorelovaných proměnných, ICA se snaží o zachycení nezávislých komponent. Metoda hledá i negaussovske komponenty. Pracuje podobně jako MPCA s NOC modelem a statistikami pro detekci odchylek.

Ve srovnání s MPCA dosahuje při použití pro online monitoring procesu a detekci chyb mnohem rychlejší odezvy [13]. Její výhoda je tedy v časnější detekci chyb. V provedené studii ([13]) například zachytila chybu v průběhu procesní proměnné 160 h před tím, než problém detekovala MPCA (studie byla prováděna na kultivacích délky 400 h).

### 3.3.6 MPLS (multiway partial least square)

Další z technik, vhodná k zjištění lineárních vztahů mezi proměnnými [7]. Je vhodná i pro špatně podmíněné matice, které se v analýze bioprocenů vyskytují. Pro jeden výstup však oproti metodám odvozeným od PCA nenabízí prakticky žádné výhody [1].

### 3.3.7 LPP (locality preserving projection)

Metoda založená na hledání vnitřní informace, která je při vícedimenzionálním pozorování ve srovnání s konvenční PCA skryta [5]. Algoritmus LPP hledá transformační matici, která transformuje prostor původních proměnných do nového prostoru nižší dimenze, přičemž minimalizuje tzv. objektivní funkci. Výhodou této metody je robustnost proti outlierům a šumu, nicméně není příliš vhodná pro popis nelineárních vztahů mezi původními proměnnými, které se u dat z bioprocenů dají předpokládat.

### 3.3.8 Metody umělé inteligence

Mezi další techniky, které je možné k monitorování fermentačních procesů použít, patří modelování pomocí metod umělé inteligence [12]. Příkladem mohou být nejrozličnější expertní systémy, založené na databázi znalostí obsluhy procesu, využití fuzzy logiky, metod strojového učení, rozpoznávání (pattern recognition) či například aplikace neuronových sítí.

## 3.4 Zhodnocení

K nalezení klíčových vztahů mezi ukazateli využijeme analýzu PCA a pro prvotní fázi (implementaci a otestování online rozhraní) použijeme zobrazení a jednoduchou predikci ukazatelů na základě statistického modelu.

Z provedené rešerše vyplývá, že pro komplexní online monitoring je nejvíce vhodným kandidátem na analýzu online dat MICA. Jeho implementace bude navazovat na vytvořené rozhraní jako další práce.



# Kapitola 4

## Sledování klíčových parametrů online

Před samotnou implementací programu na sledování klíčových parametrů bylo zapotřebí analyzovat celý bioprocess a klíčové parametry vytipovat.

K účelu lepšího popisu procesu byly ve spolupráci s VŠCHT vytipovány ukazatele, tzv. **anotace**, které měly popis celého bioprocessu zjednodušit. Cílem bylo vytvořit sérii ukazatelů, které budou schopny kvalitativně danou šarži popsat. Při výpočtu anotací byla využita jak data ukládaná do databáze, tak data z provozních záznamů (laboratorní data). Přehled všech anotací včetně jejich stručného popisu naleznete v tabulce A.1.

### 4.1 Analýza klíčových ukazatelů

Celkem bylo vytvořeno přes 40 různých anotací, proto vyvstala otázka, jak určit ty opravdu klíčové. Pro vytipování klíčových parametrů procesu byla tedy provedena PCA analýza. Analýza měla posloužit především k lepšímu pochopení vztahů a vytipování klíčových veličin, které jsou pro sledování kultivace nejpodstatnější.

Protože analýza všech anotací najednou by byla ale velmi nepřehledná, v první řadě byly anotace rozděleny do různých skupin, dle předpokládaných souvislostí jednotlivých ukazatelů (například souvislost s cukry, příkrmem, kyslíkem; dále kupříkladu skupiny podle dostupnosti jednotlivých anotací v průběhu procesu – ne všechny anotace je totiž možné určit až na konci procesu, některé lze počítat již v jeho průběhu). Přehled skupin, které byly analyzovány, naleznete v tabulce A.2.

## 4.2 PCA analýza

Proměnné (anotace) byly nejdříve před výpočtem normalizovány na jednotkový rozptyl a nulovou střední hodnotu (použita byla standardizovaná PCA). Poté byla aplikována funkce MATLABu *pca*.

*Biplot* je graf, který zobrazuje původní objekty i proměnné v jednom ordinačním diagramu [11]. Na ose  $x$  a  $y$  se nacházejí hlavní komponenty (s procentuálním vyjádřením vysvětlovaného rozptylu). Pro vytvoření biplotů byly použity komponenty s vlastním číslem větším než 1 (tzv. *Kaiserovo kritérium*).

Čím delší je vektor proměnné, tím větší je její vliv na danou hlavní komponentu. Úhel mezi vektorem proměnné a hlavní komponentou je úměrný korelaci mezi proměnnou a hlavní komponentou. Čím menší je tento úhel, tím více daná proměnná hlavní komponentu ovlivňuje. Kosinus úhlu mezi proměnnými (jejich vektory) popisuje vztahy (korelace) mezi proměnnými.

### 4.2.1 Dosažené výsledky

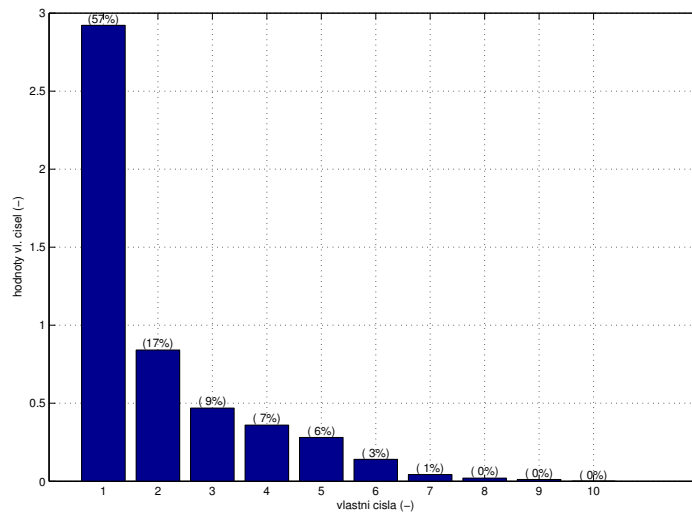
Z výsledků lze vyzdvihnout například zajímavý vztah množství Nystatinu ve 40. hodině kultivace (tedy konec batch fáze a začátek fed-batch) a počet *mia* jednotek (jednotka množství produktu) na konci kultivace, které vykazují silnější korelaci. Naopak poměrně silnou nezávislost vykazoval objem při sklizni a počet či objem celkového příkrmu. Z toho lze dovozovat, že velikost výsledné produkce ovlivňují hlavně další faktory.

Pro přehlednost zde uvádíme pouze ukázkou těch nejzajímavějších výsledků analýz. Červené tečky v obrázcích znamenají původní objekty přepočítané do prostoru biplotu. Procenta u jednotlivých proměnných značí, jakou částí daná proměnná přispívá k hlavní komponentě (hodnotu nutno ještě rozložit promítnutím do os).

Nutno předeslat, že interpretace takovýchto výsledků je velmi zrádná, neboť je třeba dávat velký pozor na procento rozptylu, který daná hlavní komponenta popisuje. Výsledky tak mohou vypadat na první pohled velmi nadějně, ale často se jedná o vysvětlení jen malé části rozptylu daných dat.

#### 4.2.1.1 Anotace související s Nystatinem

Obrázek 4.1 ukazuje vývoj vlastních čísel pro skupinu anotací, souvisejících s Nystatinem. Vlastní čísla klesají exponenciálně, což je pro analýzu žádoucí. Větší než



Obrázek 4.1: Vlastní čísla anotací souvisejících s Nystatinem.

Zdroj: autor

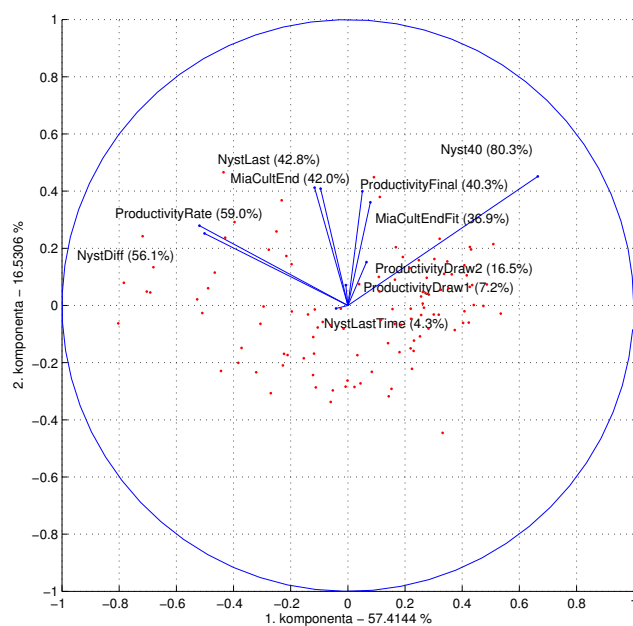
jedna je jen první vlastní číslo, proto obdržíme biplot pouze jeden.

Na obrázku 4.2 je zobrazen biplot anotací souvisejících s Nystatinem. Všimněme si, že 1. komponenta popisuje přes 57 % rozptylu proměnných, což ukazuje silnou relevanci této komponenty pro popis dat. Je vidět silná souvislost rychlosti produkce (ProductivityRate) a rozdílu Nystatinu na začátku a na konci (NystDiff) a naopak nezávislost na množství Nystatinu ve 40. kultivační hodině (Nyst40). Z toho plyne, že množství Nystatinu ve 40. hodině nemá velký vliv na výslednou produktivitu. Očividně jde tedy častěji o produktivnější kulturu, než o kulturu s vyšším množstvím Nystatinu na počátku (Nyst40).

Dále si všimněme silných korelací množství Nystatinu na konci (NystLast) a mia na konci (MiaCultEnd). To není nijak překvapivé, jedná se o velmi podobný ukazatel, jen v jiných jednotkách. Za povšimnutí stojí i čas poslední analýzy Nystatinu (NystLast-Time), jehož rozptyl zobrazované hlavní komponenty vysvětlují jen ze 4 %. To ukazuje, že tento ukazatel s ostatními příliš nesouvisí a vysvětlují jej až další hlavní komponenty.

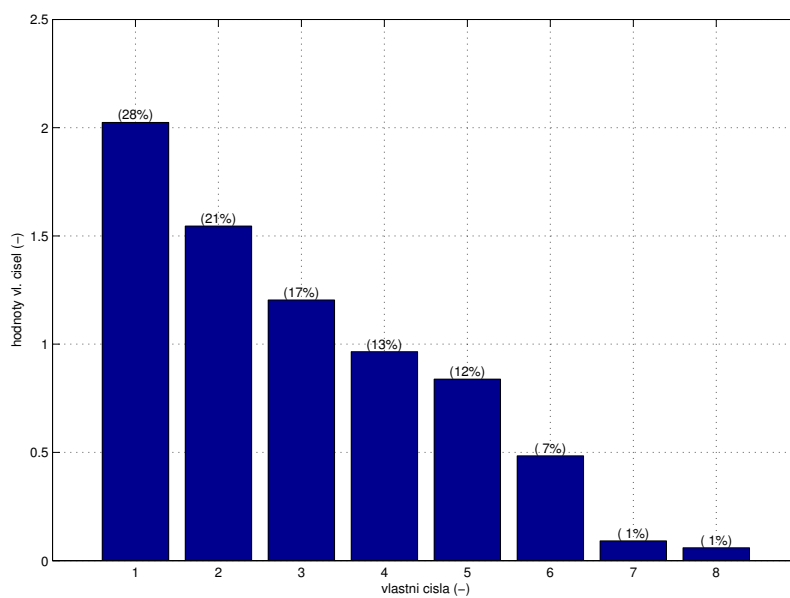
#### 4.2.1.2 Anotace související s cukry

Obrázek 4.3 ukazuje vývoj vlastních čísel pro skupinu anotací souvisejících s cukry (a příkrmy). Všimněme si, že pokles hodnot vlastních čísel je více pozvolný, pro zobrazení tedy bude potřeba více biplotů, ale jejich relevance již nebude tak vysoká jako v případě anotací souvisejících s Nystatinem. Větší než jedna jsou první tři vlastní



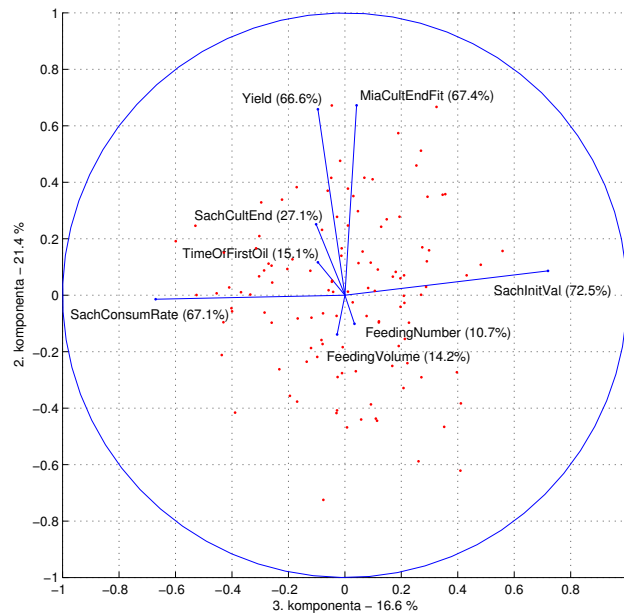
Obrázek 4.2: PCA vytipovaných anotací souvisejících s Nystatinem.

Zdroj: autor



Obrázek 4.3: Vlastní čísla anotací souvisejících s cukry.

Zdroj: autor



Obrázek 4.4: PCA vytipovaných anotací souvisejících s příkrmy (2. a 3. komponenta).

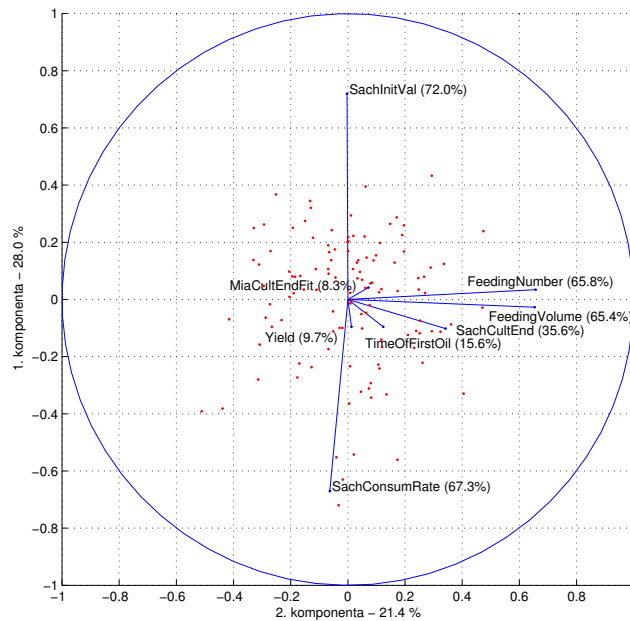
Zdroj: autor

čísla, proto dostáváme tři biploty (v případě čtyř vlastních čísel by jich bylo potřeba šest, jedná se o kombinaci bez opakování).

Na obrázku 4.4 je počáteční množství cukrů (SachInitVal) a mia na konci fermentace (MiaCultEndFit) vykazují částečnou nezávislost. To znamená, že množství živin na počátku procesu patrně nestojí za úspěšností dané šarže (která se měří právě podle množství finálního produktu na konci kultivace). Rychlost spotřeby cukrů (SachConsumRate) a počáteční množství cukrů (SachInitVal) i na těchto komponentách vykazují hezkou zápornou korelaci, což je vcelku logické – mikroorganismy mající na počátku více živin je spotřebovávají rychleji.

Z obrázku 4.5 vyplývá silná nezávislost počátečního množství cukrů (SachInitVal) a počtu (FeedingNumber), respektive objemu příkrmů (FeedingVolume). To ukazuje, že kulturu je nutné přikrmovat nezávisle na počátečním množství živin. Nutno ovšem podotknout, že hodnota ukazatele počátečního množství cukrů byla velmi málo variabilní. Rychlost spotřeby cukrů (SachConsumRate) a počáteční množství cukrů (SachInitVal) naopak vykazují opět zápornou korelaci.

4.6 je vidět poměrně silnou nezávislost množství mia na konci fermentace (MiaCultEndFit) na množství, resp. objemu příkrmů (FeedingVolume). To ukazuje, že množ-



Obrázek 4.5: PCA vytipovaných anotací souvisejících s příkrmy (1. a 2. komponenta).

Zdroj: autor

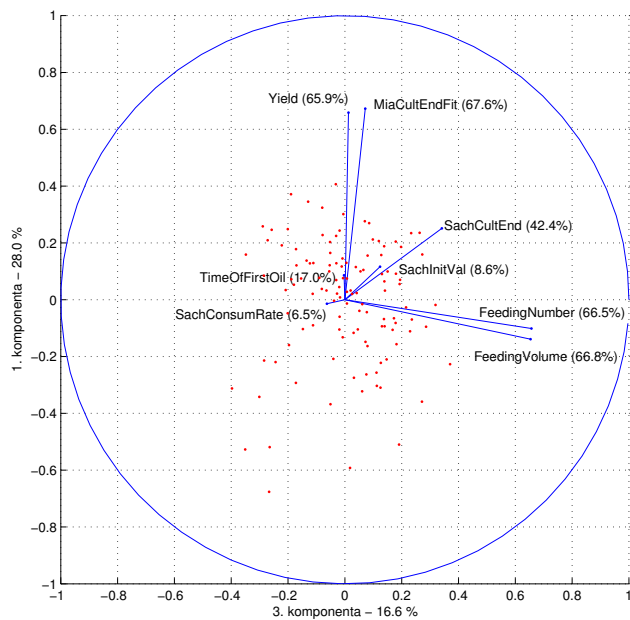
ství příkrmu nemá rozhodující vliv na velikost výsledné produkce.

#### 4.2.1.3 Anotace související s kultivací

Obrázek 4.7 ukazuje vývoj vlastních čísel pro skupinu anotací, souvisejících s délkou kultivace. Opět jsou zde dvě hlavní vlastní čísla, která jsou větší než jedna, proto dostáváme jen jeden biplot.

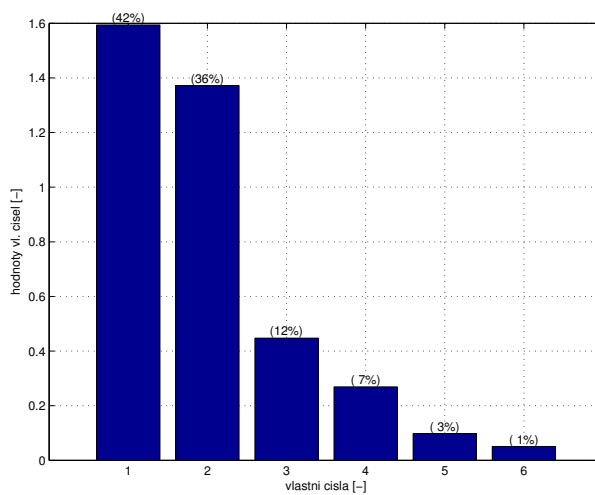
Na obrázku 4.8 vidíme biplot anotací souvisejících s kultivací. V první řadě si všimněme, že obě komponenty jsou docela silné (přes 40 %, resp. přes 35 %). Velmi pěkně se ukazuje záporná korelace délky limitace kyslíkem (OxygenLimitationLength) na končném množství Nystatinu (NystLast). To znamená, že čím kratší je limitace kyslíkem, tím větší je množství produktu na konci.

Souvislost rychlosti produkce (ProductivityRate) s Nystatinem na konci (NystLast) není nijak překvapivá, i když úplně zřejmá být nemusí. Ukazuje, že množství Nystatinu na konci ovlivňuje zejména průběh kultivace a nikoli lepší nebo horší počáteční podmínky. Podobně rychlost poklesu viskozity (ViscosityDecreaseRate) a maximální hodnota viskozity (ViscosityMaxValue) vykazuje zápornou korelaci, což je již více očekávatelné (čím



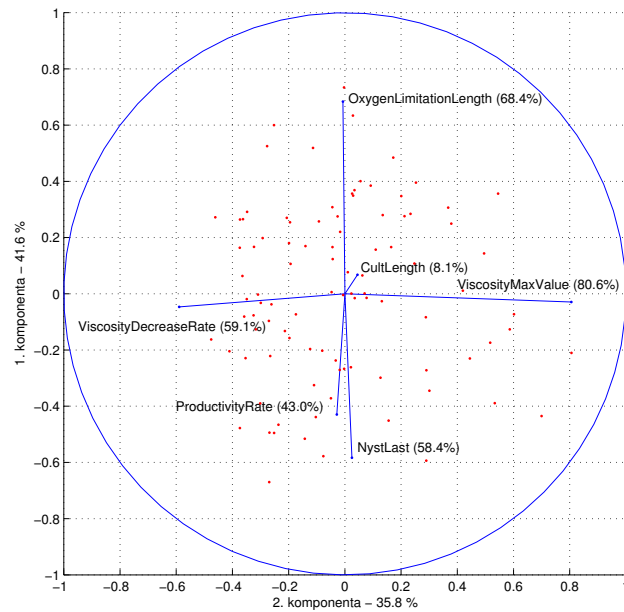
Obrázek 4.6: PCA výtípaných anotací souvisejících s příkrmy (1. a 3. komponenta).

Zdroj: autor



Obrázek 4.7: Vlastní čísla anotací souvisejících s kultivací.

Zdroj: autor



Obrázek 4.8: Anotace související s kultivací.

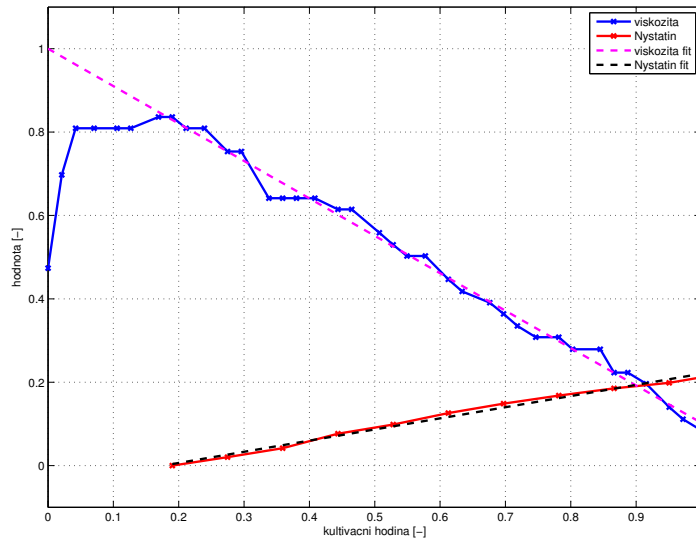
Zdroj: autor

větší je maximum, tím rychlejší musí být pokles, tj. tím menší je hodnota ukazatele poklesu).

Zajímavá je nezávislost poklesu viskozity (*ViscosityDecreaseRate*) na množství Nystatinu na konci (*NystLast*). Podtrhává to pozorovaný fakt, že k poklesu viskozity dochází vždy. Lze z toho tedy usuzovat, že použití viskozity pro predikci délky kultivace by mohlo být výhodné, neboť ji neovlivňuje výsledný produkt. Kdyby tomu bylo naopak, dal by se pokles viskozity použít přímo i k odhadu finální produkce, takto to ale nelze. Vliv na finální produkt musí mít zřejmě jiné proměnné systému.

Pro popis procesu se ukázaly anotace jako relevantní, nicméně jejich hlavní nevýhodou je, že většinu z nich je možné získat až po skončení fermentace (tedy offline). Proto jsme přistoupili k implementaci programu pro online monitorování některých významných parametrů.





Obrázek 4.9: Průběh viskozity a Nystatinu jedné šarže.

Zdroj: autor

### 4.3 Online monitoring

Na základě našich zkušeností, výše uvedené PCA analýzy (kap. 4.2.1) a s přihlédnutím k zobrazovaným parametrům stávajícího GUI (obr. 1.1) a dostupným datům (obr. 3.1 a 3.2) byly zvoleny níže uvedené ukazatele, které jsme se rozhodli v první fázi implementovat do výpočetního systému. Umožní nám to mimo jiné otestovat online komunikaci s řídicím systémem výroby.

Na obrázku 4.9 je zobrazena ukázková šarže pro ilustraci implementovaných ukazatelů. Všimněme si zejména průběhu viskozity, která zpočátku narůstá a poté, co dosáhne svého maxima, klesá přibližně lineárně. Toho bylo při implementaci využito. Množství Nystatinu narůstá přibližně od 0.2 délky kultivace, což je přesně začátek fed-batch fáze, kdy kultura začíná produkovat svůj metabolit (v našem případě právě Nystatin).

#### 4.3.1 Rychlost poklesu viskozity

Rychlost poklesu viskozity je klíčový ukazatel, který je velmi úzce spjatý s koncem kultivace. Při překročení jisté hodnoty, která patří mezi výrobní know-how, totiž již není možné jednoduchým způsobem získat výsledný produkt z fermentační směsi. I když je pravděpodobné, že kultura by ještě dále Nystatin produkovala, je žádoucí kultivaci

před dosažením této hodnoty ukončit.

Směrnice poklesu viskozity tak ukazuje na přibližnou délku kultivace. Je zřejmé, že čím rychlejší pokles viskozity, tím kratší patrně kultivace bude. Rychlost poklesu je vyjádřena v mPas/h.

### 4.3.2 Rychlost nárůstu produktu

Rychlost nárůstu produktu je podstatný ukazatel, který může v relativním předstihu říci, jak úspěšná daná šarže bude.

Rychlost je udávána v počtu jednotek Nystatinu za hodinu.

### 4.3.3 Délka kultivace

Samotná délka kultivace je parametr podstatný hned z několika důvodů. V první řadě platí přibližně přímá úměra – čím delší je délka kultivace, tím větší je výtěžnost.

Ovšem vzhledem k průběhu výroby není možné držet kultivaci donekonečna. Produkující organismus není nevyčerpatelný, dochází k různým procesům, které mohou vzniklý produkt degradovat, ať již mechanickou či chemickou cestou. Délka kultivace je silně spjatá s produkcí a s poklesem viskozity.

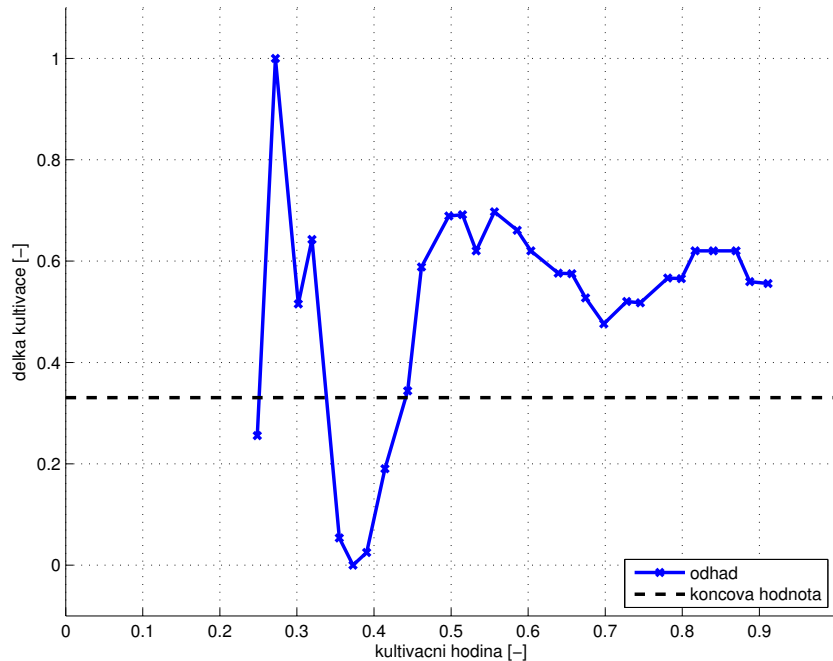
Odhad je vyjádřen v kultivačních hodinách (čas od začátku kultivace).

Odhad délky kultivace by mohl být v budoucnu využit například pro optimalizaci zakládání očkovacích a předočkovacích tanků.

### 4.3.4 Finální množství produktu

Jedná se o klíčový parametr, neboť podle něj se určuje úspěšnost celé fermentace. Přesný odhad finálního množství produktu by mohl být velmi užitečný například ke zkrácení kultivace, pro kterou je již v jejím průběhu predikováno nízké množství finálního produktu.

Finální množství produktu je odhadováno v jednotkách Nystatinu.



Obrázek 4.10: Průběh odhadu délky kultivace. Čárkovaná čára značí finální hodnotu.

Zdroj: autor

## 4.4 Ukázka odhadů

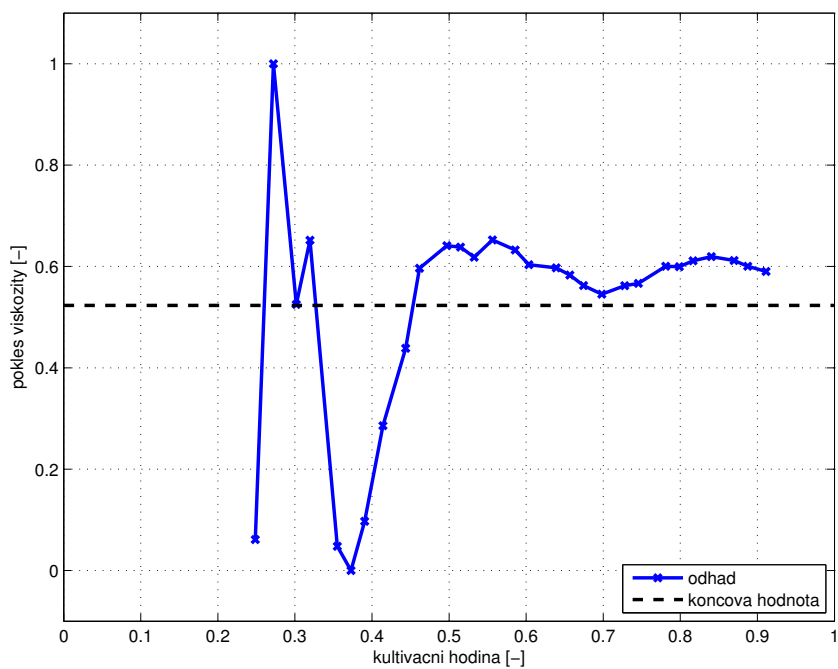
Křížky v obrázcích níže značí čas odhadu – estimátor svůj odhad prováděl na datech, která byla do tohoto času dostupná.

Na obrázku 4.10 je zobrazena ukázka odhadu délky kultivace. Lze vidět, že přibližně od poloviny kultivace se již odhad přibližuje ke skutečné hodnotě, nicméně stále s poměrně velkou chybou. Je to způsobeno závislostí na přesnosti odhadu poklesu viskozity (obrázek 4.11) a extrapolací do finální hodnoty.

Průběh odhadu rychlosti poklesu viskozity zobrazuje graf 4.11. Zde je již přesnost větší, hlavně díky tomu, že pokles viskozity je téměř lineární (obrázek 4.9). Přibližně od hodnoty 0.5 délky kultivace je již odhad v uspokojivé blízkosti finální hodnoty.

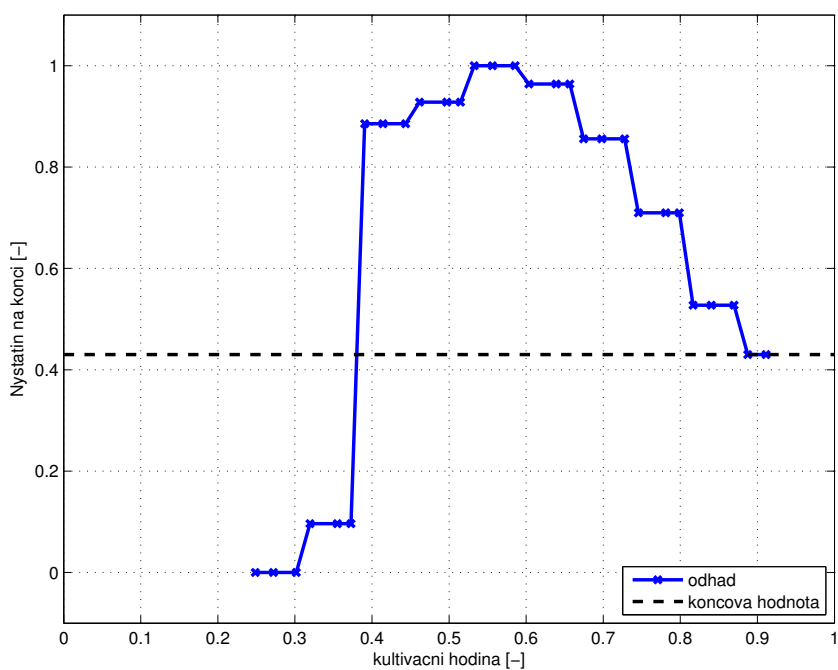
Průběh odhadu rychlosti nárůstu Nystatinu zobrazuje graf 4.12. Přibližně od poloviny kultivace se blíží k hodnotě na konci. Obrázek 4.13 představuje průběh odhadu koncového množství Nystatinu. Protože je ukazatel velmi těsně spjat s odhadem délky kultivace, nepřesnost odhadu délky kultivace se projevuje i v něm.

Podrobnějšímu vyhodnocení ukazatelů se věnuje kapitola 6.



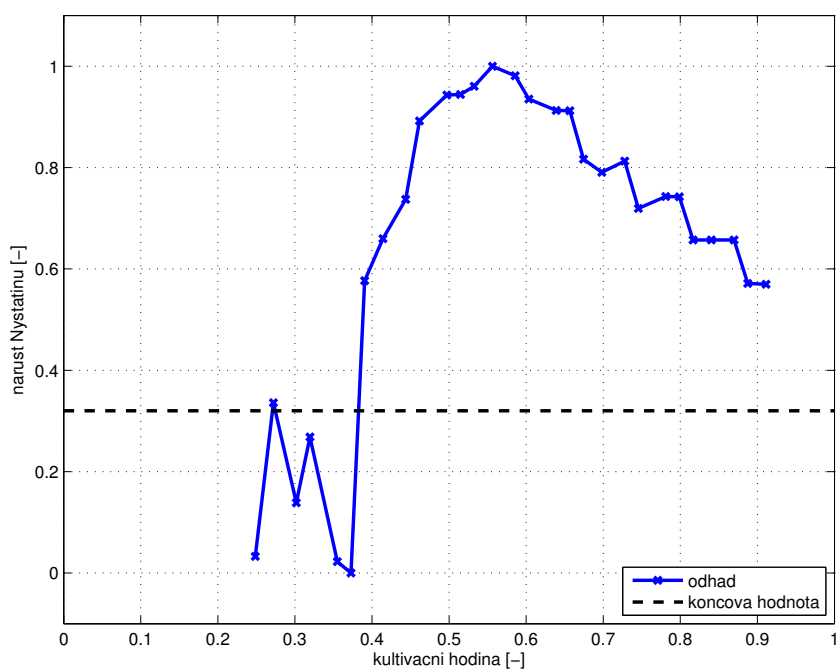
Obrázek 4.11: Průběh odhadu poklesu viskozity. Čárkovaná čára značí finální hodnotu.

Zdroj: autor



Obrázek 4.12: Průběh rychlosti nárůstu Nystatinu. Čárkovaná čára značí finální hodnotu.

Zdroj: autor



Obrázek 4.13: Průběh odhadu Nystatinu na konci kultivace. Čárkovaná čára značí finální hodnotu.

Zdroj: autor

# Kapitola 5

## Realizace rozhraní

Komunikační rozhraní má sloužit k online výpočtům nad daty ze samotné výroby. Pro realizaci rozhraní byl hlavní požadavek nezávislost na vnitřní síti provozu. Z vývojových důvodů je totiž výhodné mít oddělenou implementaci výpočetní strany od samotného provozu. Nabízí to také flexibilitu pro použití v dalších projektech.

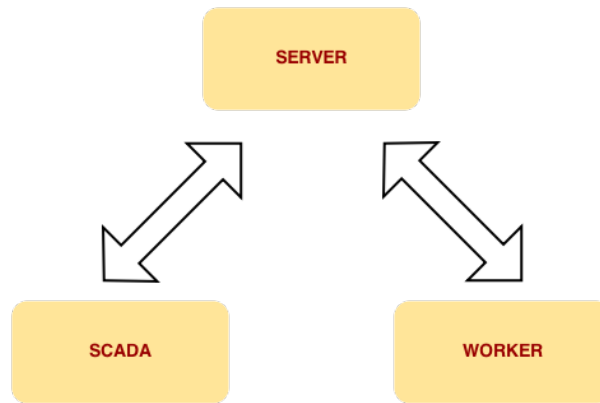
Dalším požadavkem byla modulárnost samotné výpočetní strany (workerů), aby bylo možné snadno implementovat další funkce, které nebudou závislé pouze na jednom programovacím jazyce.

Bylo třeba také vzít v potaz, že zapisování dat z laboratoře má již zavedený systém a strukturu. Zaznamenávání dat je realizováno do sešitů programu MS Excel. Bylo třeba tedy vyřešit transfer dat z laboratoře na výpočetní server a automatické načítání těchto dat.

### 5.1 Architektura

Navržená struktura má tedy tu výhodu, že zákazník (výroba) je zcela oddělen od poskytovatele (komunikační server, workeri). Tato struktura je žádoucí zejména z důvodu výzkumného záměru projektu – nákup nového programového vybavení pro experimentální výpočty by byl velkou ekonomickou zátěží.

Architekturu celého systému schematicky zobrazuje obrázek 5.1. SCADA představuje klientskou část, WORKER výpočetní a SERVER komunikační.



Obrázek 5.1: Nezávislá implementace výpočetních nástrojů.

Zdroj: autor

### 5.1.1 Komunikační server

Komunikační server běží na adrese <https://buildinglab.felk.cvut.cz>. Stará se o uchování zadání úloh od klienta, jejich poskytování workerům, ukládání výsledků a následného poskytování výsledků klientovi.

Klienta v tomto případě představuje systém SCADA, workera potom výpočetní funkce v MATLABu nebo Pythonu.

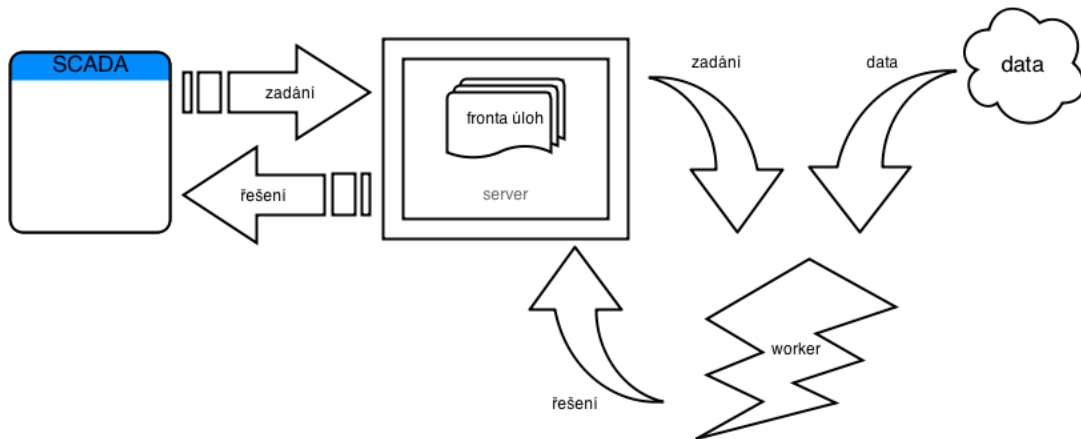
Rozhodovací pravomoc o tom, který worker bude úlohu zpracovávat, je na straně komunikačního serveru. V jednoduché lookup tabulce (viz kap. 5.2.1.2) je dané funkci přidělen patřičný worker, takže je možné měnit výpočetní kapacitu jednoduchou změnou workera.

### 5.1.2 Workers

Workers (workeri, pracovníci, výpočetní agenti) představují samostatnou funkční jednotku. Dotazují se fronty na komunikačním serveru na případná zadání výpočetních úloh a v případě kladné odezvy vykonají potřebné výpočty. Výsledek opět uloží zpět do fronty na komunikačním serveru.

### 5.1.3 SCADA

RcWare[20] je řídicí systém, použitý v provozu (systém pro monitorování procesů), zatímco SCADA vizualizační rozhraní. V našem případě SCADA zobrazuje požá-



Obrázek 5.2: Komunikace uvnitř systému. SCADA, server a worker tvoří prakticky nezávislé celky.

Zdroj: autor

dované výsledky výpočtů na velině výroby.

Systém pro zobrazování obsluze je vyvíjen firmou *Energocentrum Plus s. r. o.* Ukázka jednoduché vizualizace GUI je na obrázku 5.7.

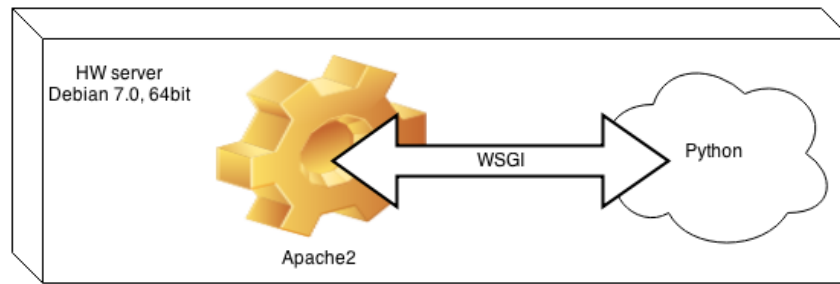
Ve zjednodušeném pohledu se SCADA chová jako klient – zadává úlohy dle požadavků na zobrazení uživateli (kterým může být např. pracovník na velině provozu). Po zadání úlohy se po určitém čase dotazuje se na výsledek, který následně zobrazí.

Obrázek 5.2 podrobně rozkrývá popisovanou komunikaci mezi systémem SCADA a výpočetními nástroji. SCADA na základě obrazovky, kterou požaduje uživatel zobrazit, zašle požadavek na výpočet dané úlohy na komunikační server. Ten požadavek zařadí do fronty úloh. Paralelně běžící worker mezitím v pravidelných intervalech sleduje frontu úloh a pokud nalezne úlohu, kterou je schopen vyřešit, stáhne si její zadání a případně další potřebná data. Zpracuje potřebné výpočty a výsledné řešení uloží zpět na serveru do fronty. Odtud je opět systémem SCADA výsledek výpočtu stažen a vizualizován.

## 5.2 Implementace

Tato část popisuje samotné technické řešení komunikačního rozhraní. Jednou z možností, kterou umožňuje přímo MATLAB, by bylo použití produkčního serveru [19], který umožňuje podobnou funkcionalitu. Jeho pořizovací cena je ale velmi vysoká, a proto jsme přistoupili k vlastnímu řešení.





Obrázek 5.3: Propojení modulu Apache2 s Pythonem.

Zdroj: autor

V MATLABu je tak pouze worker, serverovou část obsluhuje rozhraní napsané v jazyce Python.

### 5.2.1 Komunikační server

Celá filosofie komunikačního serveru je založena na REST (Representational State Transfer)[8]. Jedná se o modulární prostředí komunikace pomocí HTTP. REST určuje, jak se přistupuje k datům zdrojů. Ty mají vlastní URI a REST k nim umožňuje přistupovat čtyřmi základními metodami (GET, POST, PUT, DELETE).

Vlastní obsluha HTTP požadavků je realizována modulem Pythonu FlaskRESTful [17], který velmi usnadňuje obsluhu HTTP požadavků. Spojení Pythonu s Apache2 je potom realizováno pomocí balíku WSGI[16], tak, že spuštěním Apache2 služby dochází k automatickému zavolání skriptů v Pythonu obsluhujících komunikaci. Na obrázku 5.3 je popisovaná architektura uvnitř hardwarového serveru ilustrována.

Komunikace probíhá pomocí HTTP požadavků přes zabezpečené připojení SSL a to na portu 5052. Zároveň jsou požadavky ověřovány pomocí tokenu v tělu JSON dat, respektive parametrů předávaných komunikujícími stranami, takže nemůže dojít k jednoduchému odposlechnutí zadávaných úloh či jejich řešení.

Samotný komunikační server běží na linuxovém serveru s OS Linux Debian 7.1, 64 bit s modulem Apache2.

#### 5.2.1.1 Fronta úloh

Fronta pro udržování zadání a řešení je realizována záznamy v databázi SQLite [15] (výhodou této databáze je její jednoduchost – ukládání se realizuje do jednoho souboru, na serveru se nespouští žádné další procesy, jako je to u většiny ostatních databází).

Tabulka 5.1: Databázový záznam ve frontě úloh.

databázový záznam	popis
created	časová značka vytvoření úlohy
data	JSON struktura se vstupními daty
finished	časová značka s dokončením výpočtu
function	název funkce pro provedení výpočtu nad daty
id	identifikační číslo úlohy
result	uložení výsledku ve formátu JSON
started	časová značka pro počátek výpočtu
status	status úlohy
version	verze volané funkce

Vlastní spojení s komunikační částí v Pythonu zajišťuje modul SQLAlchemy [21].

V tabulce 5.1 je vidět struktura ukládaných dat jednotlivých záznamů zadání úlohy v databázi. Časové značky slouží zejména ke kontrole stáří úlohy, případně je možné je využít k výkonostním testům.

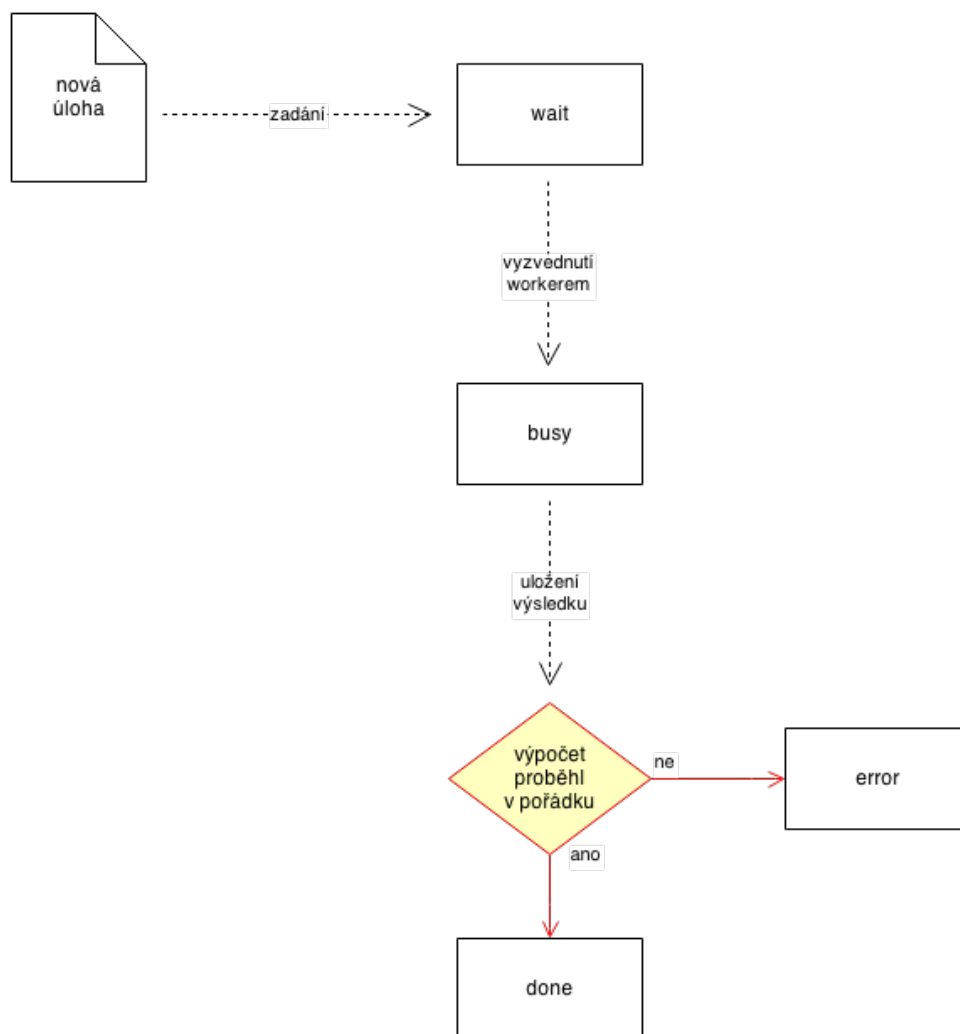
Na straně serveru se v pravidelných intervalech kontroluje, zda úlohy stáří úloh ve frontě nepřekročilo kritickou mez. V takovém případě je úloha z fronty odstraněna.

### 5.2.1.2 Lookup tabulka funkcí

Při přichozím požadavku na zpracování úlohy je v **lookup tabulce** funkcí na serverové straně nejdříve ověřena existence požadované funkce. Pokud funkce s danou verzí existuje, je jí přiřazen worker. Lookup tabulka tak slouží pro jednodušší implementaci nových funkcí (včetně jejich verzí). Představuje tak snadnou cestu ke změně workera bez nutnosti měnit klienta (tj. implementaci uživatelské strany; v našem případě systém SCADA).

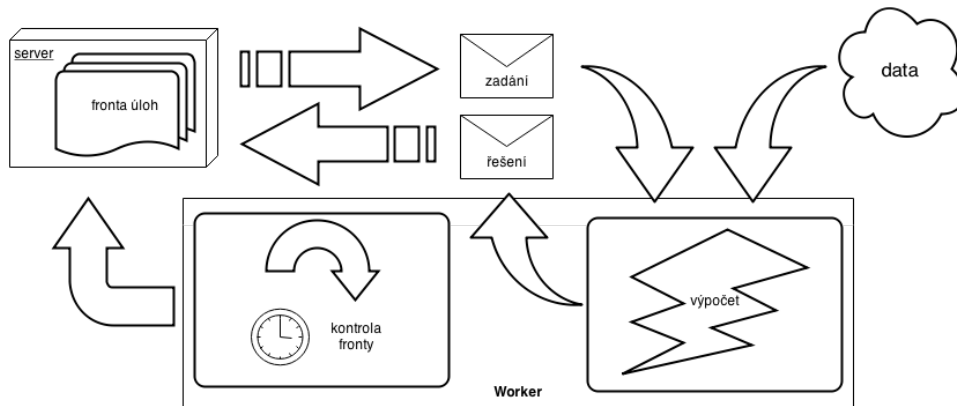
### 5.2.1.3 Stavby úlohy

K rozlišení, v jakém stavu se výpočetní úloha nachází, bylo použito několika definovaných stavů úlohy. Možné stavy úlohy ve frontě jsou znázorněny na obrázku 5.4. Po zadání nové úlohy je úloha označena stavem „wait“ (čekající na zpracování). V okamžiku, kdy je zadání vyzvednuto workerem, je úloha ve frontě označena statutem „busy“ (úloha je právě zpracovávána). Po ukončení výpočtu je úloha označena statutem „done“ (v pří-



Obrázek 5.4: Schema možných statusů úlohy.

Zdroj: autor



Obrázek 5.5: Schema práce workera.

Zdroj: autor

padě, že výpočet proběhl v pořádku), případně „error“ (v případě, že během výpočtu došlo k chybám).

## 5.2.2 Workers

Výpočetní agenti (*workers*) zajišťují výpočetní část celého systému. Samotné výpočty je možné realizovat nezávisle ve dvou prostředích – Python a MATLAB.

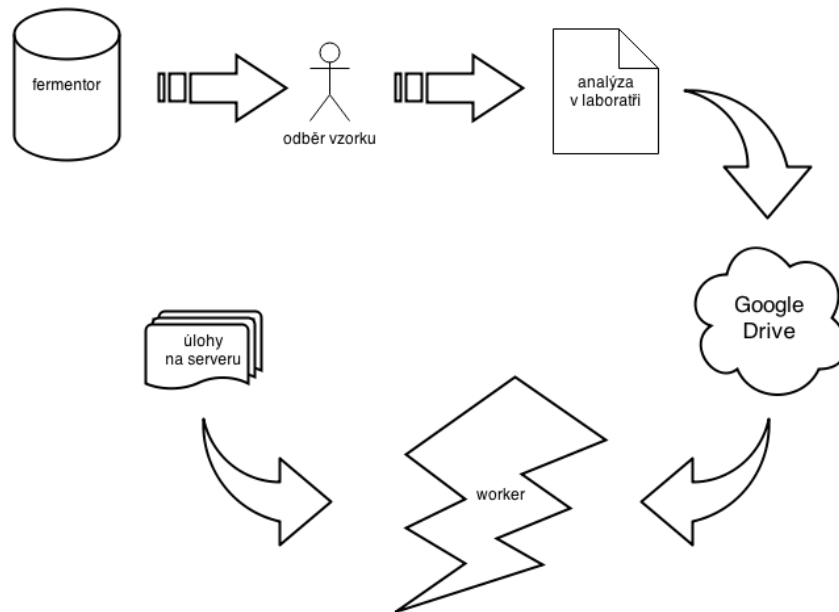
Schema práce workera je vidět na obrázku 5.5. Worker se v pravidelných intervalech dotazuje na server, zda fronta úloh neobsahuje dosud nezpracované zadání (status úlohy „wait“). V případě, že se ve frontě na serveru taková úloha nachází, worker ji označí jako „busy“, stáhne si ji a začne na ní pracovat.

Součástí zadání mohou být i online data. Offline (laboratorní) data worker využívá ze synchronizované složky (viz kap. 5.2.2.2).

Po skončení výpočtu worker řešení úlohy nahraje do fronty úloh a stav úlohy změní v případě úspěšného výpočtu na „done“, v případě výskytu chyb během výpočtu na „error“.

### 5.2.2.1 Realizace přenosu dat

Samotný přenos dat procesu je zajišťován dvěma nezávislými kanály – jako parametry v JSON těle požadavku (při zadávání úlohy) a dále synchronizací dat z laboratoře pomocí služby Google Drive.



Obrázek 5.6: Proces laboratorního měření a jeho napojení na výpočetní systém.

Zdroj: autor

### 5.2.2.2 Synchronizace dat z laboratoře

Laboratorní data jsou zapisována do sešitů programu Excel s definovanou strukturou. Ta umožňuje jejich automatické načítání do MATLABu a Pythonu. Soubor obsahuje kromě standardních polí pro laboratorní měření dle vzorkovacího plánu i mimořádná měření.

Proces laboratorního měření včetně samotného přenosu dat z laboratoře k výpočetnímu serveru je znázorněn na obrázku 5.6. Pracovník obsluhy nejdříve musí z fermentoru odebrat za přísných podmínek, zabraňujících vzniku jakékoli kontaminace fermentoru, vzorek kultivační směsi. Následně je vzorek přenesen do laboratoře, kde je analyzován. Analýza může trvat řádově i hodiny. Výsledek analýzy je poté zapsán do XLS sešitu a pomocí Google Drive synchronizován na linuxový server výpočetních nástrojů.

Pro synchronizaci dat z laboratoře byly proto vytvořeny dva Google účty – writer (s právy zapisovat; použit pro upload souborů z laboratoře) a reader (s právy pouze číst; pro download dat na výpočetní server). V laboratoři je pro synchronizaci nainstalován Google Drive klient pro Windows. Na straně serveru je pro tento účel použit linuxový klient Grive [18] (zároveň je na serveru spuštěn skript, který v pravidelných intervalech kontroluje změny v Google Drive úložišti; automatickou synchronizace Grive totiž nepodporuje). Tímto způsobem jsou načítána data ze sešitů programu Excel ve

chvíli, kdy laboranti vyplní nový záznam laboratorního měření.

### 5.2.2.3 Python worker

Worker běží ve více vláknech, jejichž počet je možné měnit v nastavení. Jedno vlákno kontroluje frontu, další vlákna slouží k výpočtům.

Výpočetní funkce se nacházejí v souboru *functions.py*:

**testDataPython(...)** – testovací funkce pro ověření komunikace se systémem SCADA.

Příklad vstupních dat – viz příloha A.2.

**multiple(x,y)** – funkce pro výpočet vektorového součinu. Vstupem jsou vektory čísel.

### 5.2.2.4 MATLAB worker

Worker využívá jen jedno vlákno. Pro náročnější výpočty je tak výhodnější využívat Python workera.

Rozhraní pro komunikaci workera doplňují knihovny *JSONLab*[10] (knihovna pro práci s formátem JSON) a *urlread2*[4] (rozšíření podpory standardní funkce *urlread*).

Výpočetní funkce se nacházejí v souboru *estimates.m*:

**testDataMatlab(...)** – testovací funkce pro ověření komunikace se systémem SCADA.

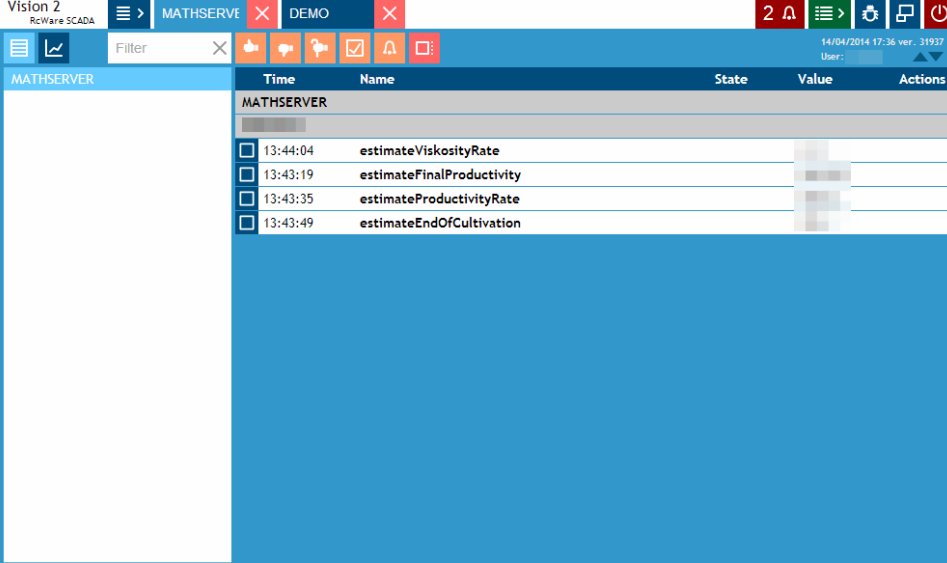
Příklad vstupních dat – viz příloha A.2.

**estimateViscosityRate** – odhad rychlosti poklesu viskozity. V průběhu viskozity vyhledá maximum a zbytek dat proloží lineární funkcí. Získaný odhad směrnice váhuje s průměrnou hodnotou viskozity (získanou z historických dat) tak, aby se při malém počtu vstupních dat kompenzovala chyba odhadu.

**estimateEndOfCultivation** – odhad konce kultivace. Proložením měření viskozity lineární funkcí a následnou extrapolací s pomocí fixně nastaveného prahu vypočítá přibližný odhad konce kultivace.

**estimateProductivityRate** – odhad rychlosti produkce. Průběh měření Nystatinu proloží lineární funkcí pro získání směrnice poklesu.

**estimateFinalProductivity** – odhad množství koncového produktu. S využitím odhadu délky kultivace extrapoluje průběh měření Nystatinu, čímž získá přibližný odhad množství Nystatinu na konci kultivace.



The screenshot shows the Vision 2 SCADA interface. At the top, there are window titles 'MATHSERVE' and 'DEMO'. Below the title bar is a toolbar with various icons. The main area displays a table with the following data:

Time	Name	State	Value	Actions
MATHSERVER				
<input type="checkbox"/>	13:44:04	estimateViscosityRate		
<input type="checkbox"/>	13:43:19	estimateFinalProductivity		
<input type="checkbox"/>	13:43:35	estimateProductivityRate		
<input type="checkbox"/>	13:43:49	estimateEndOfCultivation		

Obrázek 5.7: Nové GUI pro zobrazování údajů z výpočtů. Skutečné hodnoty rozostřeny.

Zdroj: Energocentrum Plus s. r. o.

### 5.2.3 SCADA

Současné grafické rozhraní SCADA je na obrázku 5.7. Ve své prvotní verzi funguje v textovém režimu, v budoucnu by mělo přibýt bohatší a přehlednější grafické rozhraní.

Rozhraní SCADA má sloužit na velině provozu VUAB Pharma a. s. jako vizualizace implementovaných výpočetních nástrojů.

## 5.3 API

Následující text shrnuje uživatelské API (Application Programming Interface) systému (komunikaci SCADA – server a worker – server). Syntaxe vstupu (data) a výstupu (result) je definována podle specifikace zadané funkce.

Přehled HTTP požadavků vnější a vnitřní komunikace je shrnut v tabulce 5.3. Tabulka 5.2 obsahuje seznam předávaných parametrů při zadávání úlohy.

Tabulka 5.2: Předávané parametry úlohy.

parametr	popis
user	ověření uživatele
version	verze požadované výpočetní funkce
function	název funkce pro výpočet
data	JSON struktura se vstupními daty
result	výstupní JSON struktura s výsledky výpočtů

Tabulka 5.3: Přehled HTTP požadavků.

Metoda	URL	content-type	Popis
POST	/assignments	application/json	zadání úlohy, požadované parametry viz tabulka 5.2
POST	/resolution	application/json	vyzvednutí řešení úlohy
GET	/assignment/0	application/x-www-form-urlencoded	vyzvednutí zadání úlohy (worker)
PUT	/assignment/ID	application/x-www-form-urlencoded	nastavení statusu „busy“ u právě počítané úlohy
PUT	/resolution/ID	application/x-www-form-urlencoded	uložení výsledku výpočtu

### 5.3.1 Formát dat

Pro datovou výměnu byl zvolen formát JSON [24], zejména kvůli lepší kompatibilitě se systémem SCADA. Oproti dříve hojně používanému XML má odlehčenou strukturu a snazší syntaxi i čitelnost pro člověka. Použit je pro zadávání úloh (odesílán v těle HTTP požadavku) a stejně tak v návratových hodnotách.

### 5.3.2 Zadání úlohy

Úloha je zadána systémem SCADA podle požadavků na zobrazení uživateli. Přes HTTP je zaslána JSON struktura, obsahující požadovaná pole (viz tabulka 5.2). Záznam je následně uložen do fronty (databáze) a zadavateli je navraceno ID úlohy.



Příklad těla zadání:

```
POST, /assignments:
{
  "user": "xxxxxxx",
  "function": "estimateViscosityRate",
  "version": "1.0"
}
```

Odpověď serveru:

```
{
  "status": "ok",
  "task_id": X
}
```

### 5.3.3 Výpočet úlohy

Worker se v pravidelných dotazuje na frontu. Pokud v ní existuje úloha se statusem „wait“, označí ji worker jako „busy“, stáhne její zadání a začne pracovat na výpočtu. Příklad stažení zadání:

```
GET, /assignment/0
{
  "worker": "matlab",
  "user": "xxxxxxx"
}
```

Odpověď serveru:

```
{
  "status": "ok"
  "function": "estimateViscosityRate"
  ...
}
```

Označení úlohy stavem „busy“:

```
PUT, /assignment/ID
{
  "status": "busy",
  "user": "xxxxxxx"
}
```

Po skončení výpočtu je do pole „result“ uložen výsledek a úloze je status změněn v případě úspěšného výpočtu na „done“, v případě výskytu chyb na „error“. Příklad:

```
PUT, /resolution/ID
{
  "status": "done",
  "result": 100,
  "user": "xxxxxxx"
}
```

### 5.3.4 Výsledek úlohy

Výsledek úlohy je vrácen proti požadavku s definovaným ID úlohy.

Příklad požadavku o výsledek:

```
POST, /resolution:
{
  "user": "xxxxxxx",
  "task_id": 15
}
```

Odpověď serveru:

```
{
  "status": "done",
  "result": 130
}
```

# Kapitola 6

## Vyhodnocení vlivu a přínosů

Přínos nového komunikačního rozhraní je nesporný – podařilo se vyřešit propojení datových kanálů do jednoho systému, který umožňuje další rozšiřování. Při požadavku zobrazovat další parametry, které by byly žádoucí k optimalizaci výtěžnosti výroby, máme nyní v rukou snadnou možnost jak požadované výpočty nad daty prakticky v reálném čase přenést a zobrazit přímo na velině výroby.

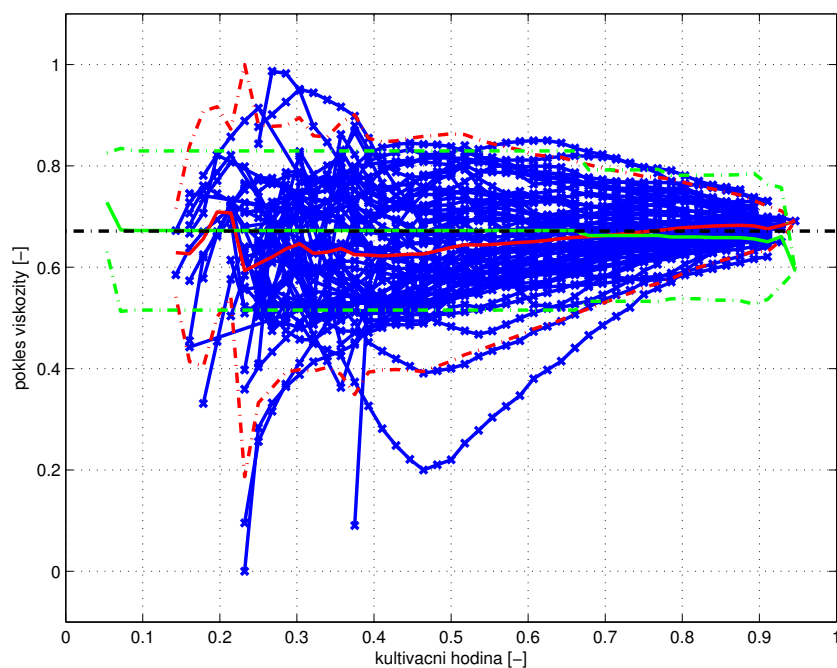
### 6.1 Přesnost odhadů

Získali jsme první online odhady parametrů, které lze kvantifikovat. Proti samotné zkušenosti obsluhy sice nedosahují výraznější přesnosti, přesto lze považovat za úspěch i tento první malý krok, který online monitoring rozšiřuje.

Většina hodnot je použitelná až od druhé poloviny kultivace, což je již poměrně pozdě, nicméně vychází to z použité metodiky – před přibližně 0.3 délky kultivace nejsou k dispozici potřebná laboratorní měření.

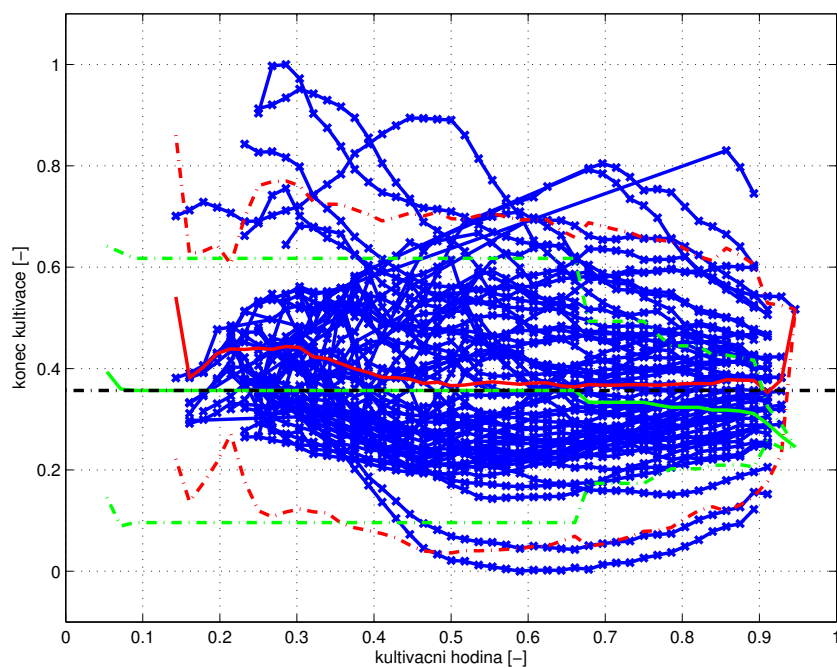
Prioritou odhadů bylo co nejvíce podchytit odchylky v šaržích – kritérium tedy bylo, dostat co možná nejmenší rozptyl odhadu. Na základě přesného odhadu by bylo možné například zkrátit dobu kultivace neúspěšně probíhající kultivace nebo naopak v rámci možností mírně prodloužit kultivaci šarže úspěšnější.

Ukázky byly simulovány na historických datech z roku 2012 a 2013. Data pro otestování přesnosti odhadů byla rozdělena do dvou skupin. Trénovací data, ze kterých byly vypočítány průměrné hodnoty ukazatelů a testovací data, která sloužila k vyhodnocení chyb daných přístupů.



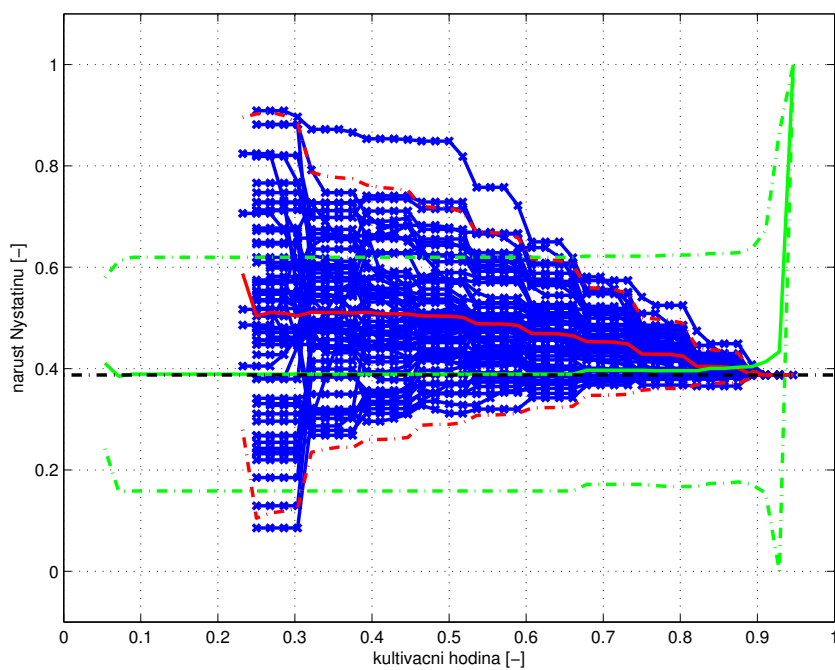
Obrázek 6.1: Přesnost odhadu ukazatele rychlosti poklesu viskozity.

Zdroj: autor



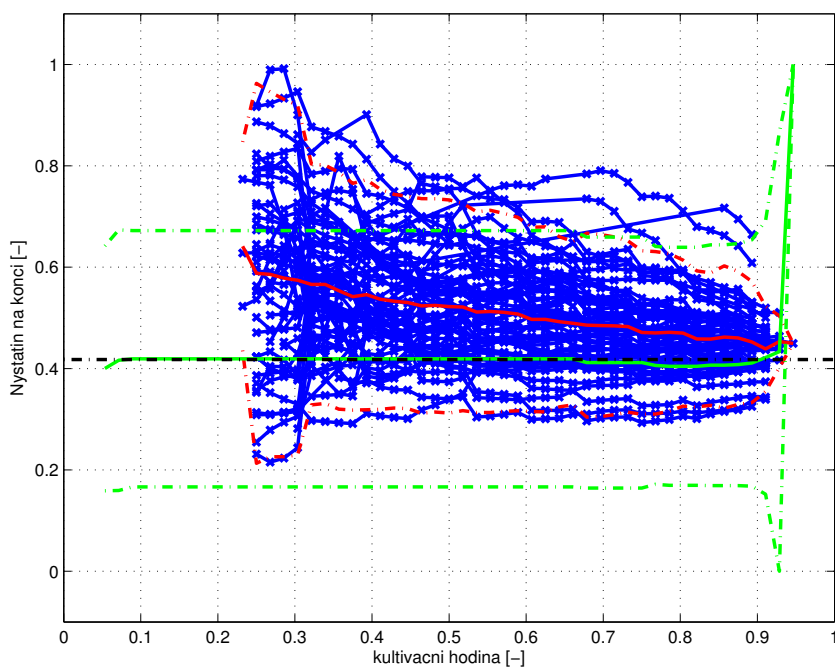
Obrázek 6.2: Přesnost ukazatele konce kultivace.

Zdroj: autor



Obrázek 6.3: Přesnost ukazatele rychlosti produkce.

Zdroj: autor



Obrázek 6.4: Přesnost ukazatele koncové produkce.

Zdroj: autor

Přesnosti odhadů níže jsou ilustrovány na vzorku sudých šarží přibližně za jeden rok. Liché šarže byly použity jako trénovací (toto rozlišení bylo použito kvůli zamezení vlivům ročních období; vliv například venkovních teplot na průběh kultivace není totiž zanedbatelný).

Na obrázcích 6.1 až 6.4 je význam čar vždy stejný. Černá čárkovaná čára značí nulovou chybu odhadu. Červená plná čára značí průběh chyby odhadu při použití funkcí, popsaných v kapitole 4.3. Zelená plná čára potom pro porovnání představuje chybu odhadu čistě na základě průměrné hodnoty. Modré čáry s křížky značí průběhy odhadů pro jednotlivé šarže. Zelená, respektive červená čerchovaná čára značí interval, kam spadá přibližně 95 % všech dat vygenerovaných danou metodikou odhadu.

Na obrázku 6.1 je přesnost odhadu ukazatele rychlosti poklesu viskozity. Lze vidět, že přibližně od 0.7 délky kultivační hodiny je již rozptyl odhadu menší než při použití průměrné hodnoty.

Na obrázku 6.2 lze vidět přesnost odhadu ukazatele délky kultivace. Protože je navázán na ostatní odhady, není jeho hodnota příliš přesná. Kvůli velkému rozptylu délek kultivace má tento ukazatel také velký rozptyl. V praxi tedy není příliš použitelný. Ukazatel by však mohl být velmi praktický například pro optimalizaci výroby, proto se jeho zpřesňováním budeme zabývat v navazující práci.

Na obrázku 6.3 je vývoj chyby odhadu v případě rychlosti produkce (nárůst Nystatinu). Všimněme si, že již od 0.6 délky kultivace má odhad menší rozptyl než odhady na základě střední hodnoty a velmi pěkně konverguje ke koncové hodnotě.

Obrázek 6.4 potom zobrazuje vývoj chyby v případě odhadu koncové produkce. Již od 0.6 délky kultivace odhad také konverguje ke koncové hodnotě a zachycení variability šarže je tak mnohem přesnější, než v případě odhadu na základě střední hodnoty. Na druhou stranu odhad ukazatel nadhodnocuje (střední hodnota odhadu na konci není nulová).

Shrneme-li obrázky 6.1 až 6.4, je vidět, že přesnost klíčových ukazatelů je sice uspokojivá až ke konci kultivace, ale vzhledem k tomu, že dosud jsme neměli odhad žádný (resp. číselný odhad byl možný až po skončení kultivace), dá se i tak hovořit o úspěchu.

Některé odhadované parametry sice ještě nemají požadovanou přesnost (odhad délky kultivace a odhad koncové produkce), ale jiné pro monitoring poskytují online cennou informaci o aktuálním stavu parametrů (rychlost poklesu viskozity a rychlost produkce). Obsluze to tak umožní pohodlnější monitorování procesu.

### 6.1.1 Zkušenosti obsluhy

Z konzultace s pracovníky výroby, kteří se starají o průběh kultivace, vyplynulo, že jejich zkušenost umožňuje odhad parametrů s větším předstihem. Potíž ovšem je, že tyto odhady se velmi špatně kvantifikují. Vychází totiž z dlouholeté zkušenosti pracovníků a jakéhosi citu pro věc.

Na vytěžování znalostí pracovníků o průběhu procesu se dále pracuje, nicméně lze předpokládat, že kvantifikace těchto informací bude z inženýrského hlediska nadále problematická.

# Kapitola 7

## Závěr

K analýze klíčových závislostí studovaného fermentačního procesu byla použita PCA analýza, která popsala a odhalila některé zajímavé vztahy vnitřních proměnných. Jako klíčové ukazatele byly vytipovány délka kultivace, rychlost poklesu viskozity, rychlost produkce a velikost produkce na konci kultivace.

Pro snazší monitorování procesu kultivace byly odhady klíčových proměnných implementovány do výpočetních nástrojů. Odhady sice zatím nejsou zcela přesné, poskytují ale cennou informaci o vývoji proměnných během kultivace online.

Implementované komunikační rozhraní přináší velký posun v možnostech monitorování a výzkumu studovaného procesu. Nyní je již možné v reálném čase pouze s ohledem na časovou náročnost laboratorních měření získat přístup k datům, která mohou být využita pro monitoring probíhající fermentace. Snadno lze též provádět nad daty výpočty prakticky v reálném čase a jejich výsledky zobrazovat přímo obsluze.

Zobrazování online indikátorů nemá sice přímý vliv na produktivitu fermentačního procesu, ale může ji ovlivnit nepřímo. Hlavní rozhodovací autoritou je zde totiž obsluha výroby, která díky novým ukazatelům bude mít více informací k činění rozhodnutí. Po patřičném zpřesnění ukazatelů pak mohou být ukazatele využity například pro zkrácení špatně se vyvíjejících šarží, což by na celkovou produkci výroby mohlo mít velmi pozitivní vliv.

Celý systém je nyní připraven pro implementaci dalších funkcí, které vyplynou z vývoje projektu, a které budou obsluze poskytovat další informace.

Cíl práce, implementace rozhraní pro komunikaci a zobrazování online výpočtů, byl tedy splněn. Implementace výpočetní strany je snadno rozšiřitelná a i když v současné době nenaplnuje svůj funkční potenciál, v praxi se jí podařilo otestovat a poskytuje modulární prostředí pro další rozšiřování funkčnosti.



Nahlédnutí do problematiky reálné výroby a reálných dat i situací s tím spjatých bylo zajímavé. Na rozdíl od ideálů akademického prostředí reálná data trpí celou řadou problémů a ne vždy jsou řešení těchto potíží přímočará. Aplikace v průmyslovém prostředí vyžadují svůj čas a mimo samotného technického řešení často více času stojí jej vykomunikovat a prosadit.

## 7.1 Další práce

V současné době se ve spolupráci s kolegy s VŠCHT pracuje na vytvoření modelu procesu, který by měl být v budoucnu použit například k predikci limitace kyslíkem a dalším odhadům, které budou moci být jako další výpočetní moduly připojeny do stávajícího rozhraní.

Zpřesnění stávajících odhadů by bylo možné změnou metodiky – například zahrnutím více vztahů do odhadu či využitím připravovaného modelu. Bylo by také praktické ještě k zobrazovaným informacím přidat intervaly spolehlivosti.

Zároveň se bude pracovat na zpřesnění ukazatele konce kultivace, který by mohl být použit pro optimalizaci výroby. Další práce zahrnují implementaci MICA metody pro komplexní monitoring více proměnných online.

Na závěr bych všem zúčastněným stranám ještě jednou rád poděkoval.

# Literatura

- [1] Sarolta Albert and Robert D. Kinley. Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. *Trends in Biotechnology*, 19(2):53–62.
- [2] Arne Bluma, Tim Höpfner, Patrick Lindner, Christoph Rehbock, Sascha Beutel, Daniel Riechers, Bernd Hitzmann, and Thomas Scheper. In-situ imaging sensors for bioprocess monitoring: state of the art. *Analytical and bioanalytical chemistry*, 398(6):2429–38.
- [3] Leo H. Chiang, Riccardo Leardi, Randy J. Pell, and Mary Beth Seasholtz. Industrial experiences with multivariate statistical analysis of batch process data. *Chemometrics and Intelligent Laboratory Systems*, 81(2):109–119.
- [4] Jim Hokanson. urlread2. <http://www.mathworks.com/matlabcentral/fileexchange/35693-urlread2>. Dostupné online 28.4.2014.
- [5] Kunlun Hu and Jingqi Yuan. Multivariate statistical process control based on multiway locality preserving projections. *Journal of Process Control*, 18(7):797–807.
- [6] Jong-Min Lee, Chang Kyoo Yoo, and In-Beum Lee. Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis. *Journal of Biotechnology*, 110(2):119–136.
- [7] B. Lennox, G. A. Montague, H. G. Hiden, G. Kornfeld, and P. R. Goulding. Process monitoring of an industrial fed-batch fermentation. *Biotechnology and Bioengineering*, 74(2):125–135.
- [8] Martin Malý. Rest architektura pro webové api. <http://www.zdrojak.cz/clanky/rest-architektura-pro-webove-api/>. Dostupné online 28.4.2014.
- [9] Karel Melzoch, Kateřina Demnerová, and Pavel Dostálek. *Modelování bioprocusů a jednotkových operací, analýza rizik*. Fondu rozvoje vysokých škol, 1999.

- [10] Qianqian Fang, Qianqian Fang, Qianqian Fang, and Qianqian Fang. JSONlab. <http://www.mathworks.com/matlabcentral/fileexchange/33381-jsonlab--a-toolbox-to-encode-decode-json-files-in-matlab-octave>. Dostupné online 28.4.2014.
- [11] RNDr. Danka Haruštiaková, Ph.D., RNDr. Jiří Jarkovský, Ph.D., RNDr. Simona Littnerová, doc. RNDr. and Ladislav Dušek, Ph.D. *Vícerozměrné statistické metody v biologii*. AKADEMICKÉ NAKLADATELSTVÍ CERM, s. r. o. Brno.
- [12] Varanon Uraikul, Christine W. Chan, and Paitoon Tontiwachwuthikul. Artificial intelligence for monitoring and supervisory control of process systems. *Engineering Applications of Artificial Intelligence*, 20(2):115–131.
- [13] Chang Kyoo Yoo, Jong-Min Lee, Peter A. Vanrolleghem, and In-Beum Lee. On-line monitoring of batch processes using multiway independent component analysis. *Chemometrics and Intelligent Laboratory Systems*, 71(2):151–163.
- [14] Cenk Ündey, Duncan Low, Jose C. Menezes, and Mel Koch. *PAT Applied in Biopharmaceutical Process Development And Manufacturing: An Enabling Tool for Quality-by-Design*. CRC Press, Boca Raton, FL, 1 edition edition, December 2011.
- [15] About SQLite. <https://sqlite.org/about.html>. Dostupné online 28.4.2014.
- [16] Apache mod WSGI. [http://flask.pocoo.org/docs/deploying/mod\\_wsgi/](http://flask.pocoo.org/docs/deploying/mod_wsgi/). Dostupné online 28.4.2014.
- [17] FlaskRESTful user's guide. <http://flask-restful.readthedocs.org/en/latest/>. Dostupné online 28.4.2014.
- [18] Grive project. <http://www.lbreda.com/grive/>. Dostupné online 28.4.2014.
- [19] MathWorks production server. <http://www.mathworks.com/products/matlab-production-server/>. Dostupné online 28.4.2014.
- [20] RcWare. <http://rcware.eu/public/rcware>. Dostupné online 28.4.2014.
- [21] SQLAlchemy. <http://www.sqlalchemy.org/>. Dostupné online 28.4.2014.
- [22] TAČR - Znalostní řízení biotechnologické výroby antibiotika nystatin. <http://www.isvav.cz/projectDetail.do?rowId=TA03010165>. Dostupné online 28.4.2014.
- [23] VUAB pharma a. s., historie společnosti. <http://www.vuab.cz/index/clanky/kategorie/historie>. Dostupné online 28.4.2014.
- [24] Úvod do JSON. <http://www.json.org/json-cz.html>. Dostupné online 28.4.2014.

# Příloha A

## A.1 Anotace

Tabulka A.1: Přehled anotací.

Popis anotace	Zkratka používaná v grafech
Celkové množství přidaného oleje proti pění	AntiFoamOilTotal
Doba kultivace	CultLength
Poměr počáteční DO (dissolved oxygen) vůči ustálené hodnotě na konci kultivace	DOratio
Ustálená hodnota obsahu rozpuštěného kyslíku na konci kultivace	DOSTeadyState
Klasifikace průběhů DO (dissolved oxygen)	DOTgroup
Čas prvního odtahu	Draw1Time
Doba mezi prvním a druhým odtahem	DrawTimeDiff
Počet příkrmů	FeedingNumber
Celkový objem příkrmu	FeedingVolume
Celkový objem média včetně odtahů při sklizni	HarvestVolume
Odhad růstové rychlosti biomasy	Mi
Mia na konci fermentace	MiaCultEnd
Množství přidaného NTF	NTFAmount
Čas přidání NTF	NTFInsertionTime
Doba limitace kyslíkem	OxygenLimitationLength
pH konec kultivace	pHCultEnd
pH při seedování ve FT	pHInoculation
Počet jednotek antibiotika v době prvního odtahu	ProductivityDraw1
Počet jednotek antibiotika v době druhého odtahu	ProductivityDraw2

Produktivita v okamžiku poslední analýzy nystatinu	ProductivityFinal
Rychlost produkce nystatinu	ProductivityRate
Rychlost spotřeby cukrů	SachConsumRate
Redukující látky na konci kultivace	SachCultEnd
Počáteční koncentrace cukrů	SachInitVal
Objem vsádky před zaočkováním	StartingVolume
čas (v kultivačních hodinách), kdy se poprvé přidá olej do fermentoru	TimeOfFirstOil
Rychlost poklesu viskozity v produkční fázi	ViscosityDecreaseRate
Maximální hodnota viskozity	ViscosityMaxValue
Výtěžnost YP/S (množství produktu na substrát)	Yield
Čas sklizně (celková doba kultivace)	HarvestTime
Rozdíl nystatinu na začátku kultivace a na konci	NystDiff
Mia na konci fermentace, udhadnutá pomocí vývoje Nystatinu ve ferm. tanku (zahrnuje i odtahy)	MiaCultEndFit
Koncentrace Nystatinu ve 40. kultivační hodině	Nyst40
Počátek fermentace	CultStartDate
Nystatin z poslední analýzy	NystLast
Čas poslední analýzy Nystatinu	NystLastTime
Množství cukrů (redukujících látek) před koncem kultivace (prumer pres 120. az posledni hodinu)	SachEndAvg
Počet příkrmů křídou	ChalkNumber
Čas prvního přidání křídou	ChalkFirstTime

Tabulka A.2: Přehled skupin anotací pro PCA analýzu.

Název	ovlivni obsluha (primo)
	znamo uz na zacatku
	znamo az v prubehu
	znamo az na konci
	souvisi s cukry
	souvisi s kyslikem
	souvisi s pH, viskozitou
	souvisi s nystatinem
	souvisi s predockovacim tankem
	korelace 07

AntiFoamOilTotal	1	0	1	0	0	0	0	0	0	0
CultLength	1	0	0	0	0	0	0	0	0	1
DORatio	0	0	0	1	0	1	0	0	0	1
DOSTeadyState	0	0	0	1	0	1	0	0	0	0
DOTgroup	0	0	0	1	0	1	0	0	0	0
Draw1Time	1	0	1	0	0	0	0	0	0	0
DrawTimeDiff	1	0	1	0	0	0	0	0	0	0
FeedingNumber	1	0	0	1	1	0	0	0	0	1
FeedingVolume	1	0	0	1	1	0	0	0	0	1
HarvestTime	1	0	0	1	0	0	0	0	0	0
HarvestVolume	0	0	0	1	0	0	0	0	0	0
MiaCultEnd	0	0	0	1	0	0	0	1	0	1
MiaCultEndFit	1	1	1	1	1	1	1	1	1	1
MilletPreservesAge	1	1	0	0	0	0	0	0	1	0
NTFAmount	1	0	1	0	0	0	0	0	0	0
NTFInsertionTime	1	0	1	0	0	0	0	0	0	0
Nyst40	0	1	0	0	0	0	0	1	0	0
NystDiff	0	0	0	1	0	0	0	1	0	1
NystLast	0	0	0	1	0	0	0	1	0	1
NystLastTime	0	0	0	1	0	0	0	1	0	1
pHCultEnd	0	0	0	1	0	0	1	0	0	0
pHInoculation	0	1	0	0	0	0	1	0	0	0
PreInoculationTankName	0	1	0	0	0	0	0	0	1	0
ProductivityDraw1	0	0	1	0	0	0	0	1	0	0
ProductivityDraw2	0	0	1	0	0	0	0	1	0	0
ProductivityFinal	0	0	0	1	0	0	0	1	0	1
ProductivityRate	0	0	1	0	0	0	0	1	0	1
SachConsumRate	0	1	0	0	1	0	0	0	0	0
SachCultEnd	0	0	0	1	1	0	0	0	0	0
SachInitVal	0	1	0	0	1	0	0	0	0	0
StartingVolume	0	1	0	0	0	0	0	0	1	0
TimeOfFirstOil	1	1	0	0	1	0	0	0	0	0
ViscosityDecreaseRate	0	0	0	0	0	0	1	0	0	0
ViscosityMaxValue	0	0	0	0	0	0	1	0	0	0
Yield	0	0	0	1	1	0	0	0	0	1

## A.2 Příklad testovacích dat

Testovací vstup pro funkce *testDataPython* a *testDataMatlab*.

```
{
  'online_data':
  {
    'time': ['2014-03-17T13:40:07Z', '2014-03-17T13:42:07Z',
            '2014-03-17T13:44:07Z', '2014-03-17T13:46:07Z'],
    'volume': [22, 23, 24, 24],
    'batch_name': ['X_105_2013', 'X_106_2013', 'X_106_2013', 'X_106_2013']
  }
  'meteo_data':
  {
    'time': [0,5,10,15],
    't': [22, 23.2, 24, 24.5],
    'p': [10, 1024, 1223, 1220]}
  }
}
```

# Příloha B

## Obsah přiloženého CD

K této práci je přiloženo CD, na kterém je uložena elektronická verze práce a zdrojové kódy.

Struktura CD je následující:

*kohouj14\_DP.pdf* – elektronická verze práce

*computingserver/* – zdrojové kódy serveru a workerů