AL-FARABI KAZAKH NATIONAL UNIVERSITY FACULTY OF BIOLOGY AND BIOTECHNOLOGY
DEPARTMENT OF BIOTECHNOLOGY

INTRODUCTION TO BIOLOGICAL ENGINEERING

LECTURE 7.
CONCEPTUAL MODELING OF BIOLOGICAL ENGINEERING SYSTEMS: FROM MODERN APPLICATIONS TO ENGINEERING ANALYSIS

PhD., Acting Associate
Professor Doktyrbay Gulina

Lecture Plan

1. Purpose of conceptual models
2. System boundaries
3. Block diagrams
4. Causal maps
5. Stocks and flows
6. Assumptions
7. Model granularity
8. Qualitative dynamics
9. Validation logic
10. Case studies and transition to quantitative models

Learning Outcomes

By the end of this lecture, students should be able to:

LO1 Explain the purpose and limits of conceptual models.

LO2 Construct block, causal and stock-flow representations of biological systems.

LO3 Choose appropriate model boundaries and level of detail.

LO4 State assumptions and identify missing variables.

LO5 Use conceptual models to generate hypotheses and guide measurement.

LO6 Translate a qualitative model into variables and relations for later quantitative analysis.

What is a conceptual model?

A conceptual model is a simplified representation of the important entities, processes and relationships in a system.

A conceptual model is **not yet a full mathematical model**; its first role is to organize knowledge, assumptions and causal relationships in a clear form.

It defines the **system boundary** and specifies which variables, processes and interactions are included or deliberately excluded.

A useful model distinguishes between **inputs, internal states, transformations, outputs, disturbances and controllable variables**.

Conceptual models make assumptions explicit and help identify **missing information, uncertain mechanisms and variables that should be measured experimentally**.

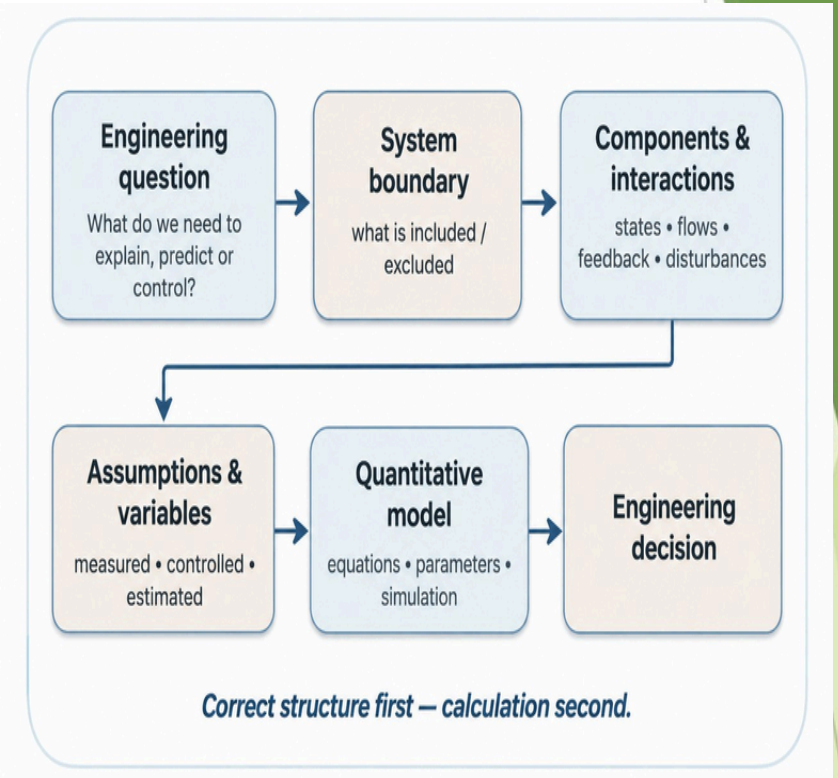
The appropriate level of detail depends on the question being asked: unnecessary complexity can obscure the dominant mechanisms and make interpretation more difficult.

Once the structure of the system is clear, the conceptual model can be translated into **quantitative equations, computational simulations or experimental designs**.

REAL BIOLOGICAL SYSTEM
many variables • interactions • uncertainty
↓
DEFINE THE QUESTION
What do we want to explain, predict or control?
↓
SELECT SYSTEM BOUNDARY
what is included / excluded
↓
IDENTIFY KEY ELEMENTS
inputs • states • processes • outputs
↓
CONCEPTUAL MODEL
causal links • assumptions • measurable variables
↓
TEST / REFINE
experiments • data • model revision

Why model before calculating?

Many analytical errors begin before equations are written: wrong boundary, missing pathway, ambiguous variable or hidden assumption.
Conceptual modelling forces the analyst to define **what belongs inside the system, what remains outside and which exchanges cross the system boundary**.
It identifies the **key components, state variables, interactions, constraints and assumptions** that must be represented before numerical analysis begins.
A conceptual model helps distinguish between **measurable variables, controllable variables, disturbances and quantities that must be estimated**.
Building the system structure first prevents unnecessary complexity and reveals where additional experiments or measurements are required.
Conceptual models also provide a **common language for interdisciplinary teams**, allowing biological, engineering and computational assumptions to be discussed before formal equations are introduced.
Only after the system structure and assumptions are justified should they be translated into **equations, parameters, simulations and quantitative predictions**.



Block diagrams

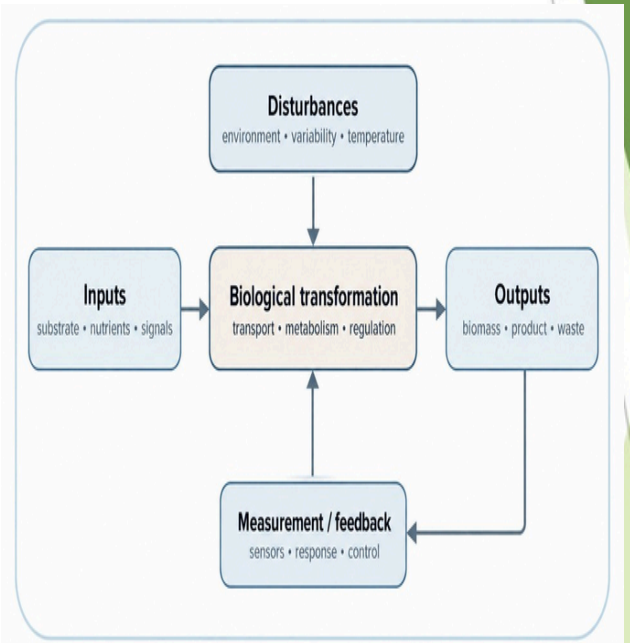
Block diagrams represent major transformations as boxes connected by directed flows.

Each block represents a **process, biological function, unit operation or decision step**, while arrows indicate the direction of material, energy or information transfer. Block diagrams help reveal **system structure, interfaces, dependencies and feedback relationships** before detailed mathematical equations are introduced.

Inputs may include **substrates, nutrients, signals or operating conditions**, whereas outputs can include biomass, products, waste, heat or measured responses. Feedback links show how an output or measured state can influence an upstream process, making block diagrams useful for representing **regulation and control**.

The appropriate level of detail depends on the engineering question: a useful diagram includes only the components and interactions necessary to support the intended analysis.

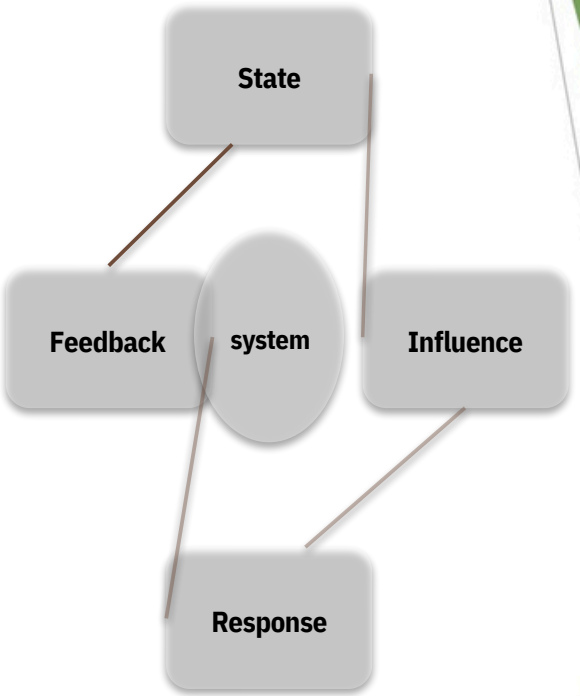
More detailed models can later expand individual blocks into **kinetic models, transport models, control equations or experimental subsystems**.



Causal-loop diagrams

Causal maps focus on influence rather than material flow.

- A positive link means that increasing one variable tends to increase another, all else equal; a negative link indicates the opposite effect.
- A model is a purposeful simplification, not a complete copy of reality.
- Its value depends on whether the chosen states, interactions and assumptions are sufficient for the intended decision.
- Increasing detail is useful only when it changes prediction, interpretation or design choice.
- KEY PRINCIPLE Increasing detail is useful only when it changes prediction, interpretation or design choice.



Stocks and flows

SUBSTRATE stock A model is a purposeful simplification, not a complete copy of reality.

A **stock** is a state variable that represents an accumulated quantity, such as biomass, substrate, product, dissolved oxygen or intracellular metabolite concentration.

Inflows increase the stock, whereas **outflows and consumption processes** decrease it.

In biological systems, flows are often controlled by **reaction rates, transport rates, feeding, secretion, degradation or harvest**.

Stocks describe the current state of the system, while flows determine **how that state changes over time**.

Distinguishing stocks from flows helps prevent common modelling errors, especially when interpreting accumulation, depletion and steady-state behaviour.

A useful stock–flow model includes only those variables and transfer processes that are necessary to answer the engineering question.

INFLOW
feed / production
→
STOCK
Substrate concentration
→
OUTFLOW / CONSUMPTION
uptake • degradation • removal
А ниже второй stock:
BIOMASS STOCK
cell concentration

Boundary choice in modelling

A model of gene expression may include transcription, translation and protein degradation while treating cell growth as an external parameter.

At the **molecular level**, a model may include transcription, translation, protein degradation and regulatory interactions while treating cell growth as an external condition.

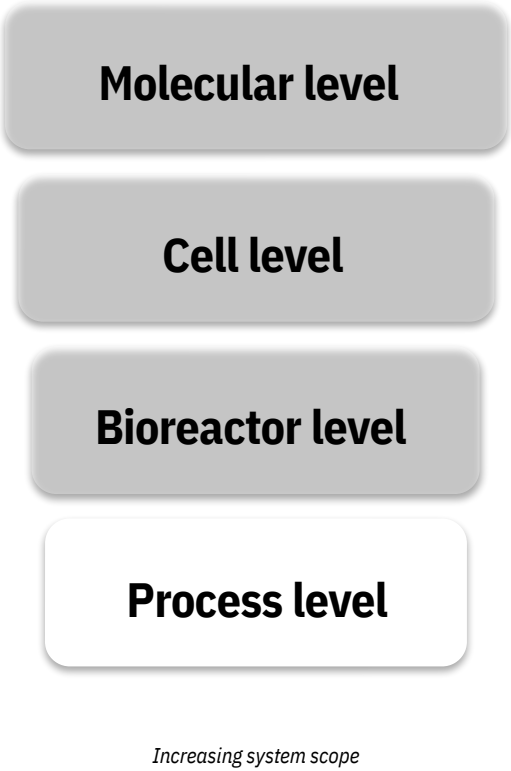
At the **cell level**, growth, metabolism and resource allocation may become explicit state variables, while detailed molecular mechanisms are simplified.

At the **bioreactor level**, the model may focus on biomass, substrate, oxygen, product formation and transport while representing intracellular regulation only through effective kinetic parameters.

At the **process level**, multiple unit operations, recycle streams, purification steps and control strategies may be included, while individual cellular mechanisms remain aggregated.

Changing the system boundary changes which quantities are treated as **states, inputs, disturbances, parameters and outputs**.

The best boundary is therefore the **smallest scope that still captures the mechanisms required for the intended prediction or decision**.



Granularity

Granularity is the level of detail at which the system is represented.

•A **fine-grained model** represents many individual components and mechanisms, such as molecular interactions, regulatory events or intracellular pathways.

A **coarse-grained model** combines details into larger functional units, such as cells, reactors or process stages, reducing complexity while preserving the dominant behaviour.

Excessive detail can create too many parameters, increase uncertainty and make the model difficult to identify, validate and interpret.

Too little detail can hide important mechanisms, producing predictions that are too approximate for the intended application.

A practical modelling strategy is to **start with the simplest representation** and add detail only when experiments, sensitivity analysis or prediction errors show that additional mechanisms are necessary.

Model granularity should therefore be judged by whether additional detail **changes prediction, interpretation or the engineering decision**.

Molecule
binding • transcription
→ **Pathway**
reactions • regulation
→ **Cell**
growth • metabolism
→ **Bioreactor**
transport • mixing • kinetics
→ **Process**
upstream • downstream • control

Assumptions as model components

Assumptions include steady state, perfect mixing, constant temperature, negligible product degradation or homogeneous cell population.

Common assumptions in biological engineering include **steady state, perfect mixing, constant temperature, negligible product degradation and homogeneous cell populations**.

Each assumption reduces model complexity, but it also defines a **condition under which the model may lose accuracy or fail completely**.

Assumptions should therefore be stated explicitly and linked to their **expected effect on predictions, parameter estimates and engineering decisions**.

Whenever possible, assumptions should be tested against **experimental data, sensitivity analysis or alternative model formulations**.

Hidden assumptions are especially dangerous because they can create **false confidence in predictions that appear mathematically precise but are valid only under restricted conditions**.

Assumptions should be revised when new evidence shows that an ignored mechanism or source of variability materially affects system behaviour.

Examples

- steady state
- perfect mixing
- constant temperature
- negligible degradation
- homogeneous population

Inputs, outputs and disturbances

Inputs are controlled or specified quantities entering the model; outputs are predicted or measured responses.

Inputs are specified or manipulated quantities entering the system, such as substrate concentration, feed rate, gas flow, temperature set point or an external biological signal. The **system** transforms these inputs through biological processes such as transport, metabolism, regulation, growth or product formation.

Outputs are measurable or predicted responses, including biomass, product concentration, metabolite levels, fluorescence, heat generation or process performance.

Disturbances affect system behaviour but are not directly manipulated; examples include raw-material variability, environmental fluctuations, biological heterogeneity and unexpected process deviations.

Clearly separating inputs, outputs and disturbances is essential for **experimental design, process control, model validation and interpretation of cause–effect relationships**.

Each variable should therefore be assigned a specific role in the model rather than being included in a single undifferentiated list.

CONTROLLED INPUTS

feed • temperature set point • gas flow • signal →

BIOLOGICAL SYSTEM

transport • metabolism • growth • regulation →

OUTPUTS

biomass • product • metabolites • measured response

DISTURBANCES

environment • variability • contamination • biological noise

Mechanistic versus empirical structure

Mechanistic and empirical models represent biological systems in fundamentally different ways: mechanistic models describe how a system is believed to work, whereas empirical models describe relationships observed in data.

Mechanistic models represent explicit biological or physical processes such as transport, reaction kinetics, regulation, growth or mass transfer.

Their parameters usually have a **physical or biological interpretation**, making the model easier to connect with known mechanisms and engineering constraints.

Empirical models infer relationships directly from experimental data without requiring a detailed representation of the underlying mechanism.

They can be efficient for **prediction, calibration and pattern recognition**, particularly when large datasets are available but system mechanisms are incomplete or difficult to formulate.

Mechanistic models can sometimes support more meaningful extrapolation beyond measured conditions, while empirical models are usually most reliable **within the data domain on which they were trained or fitted**.

Hybrid models combine mechanistic structure with data-driven components, allowing biological knowledge and experimental evidence to complement one another.

Representing uncertainty

Conceptual models should mark uncertain pathways, parameter ranges or alternative hypotheses.

Parameter uncertainty arises when model parameters such as kinetic constants, yields or transport coefficients are not known exactly.

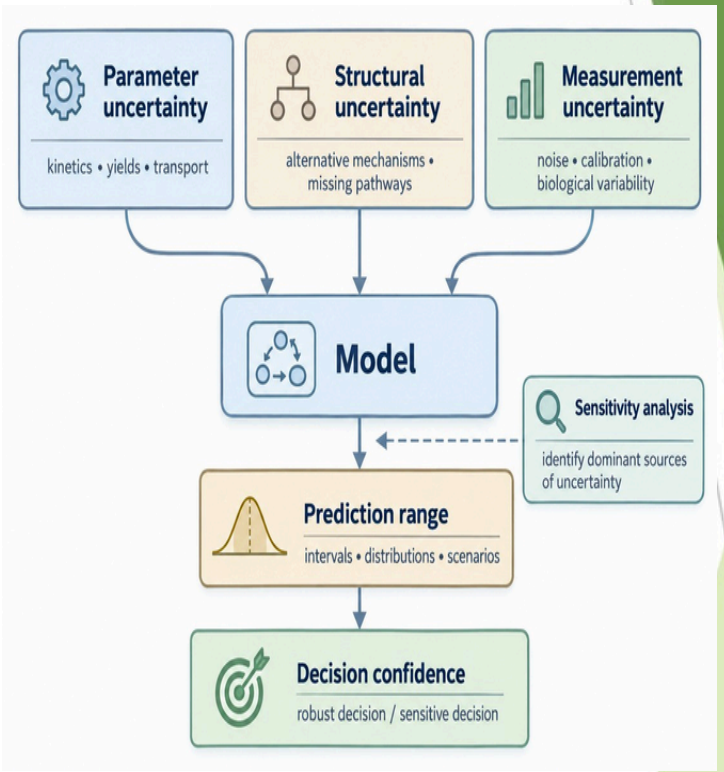
Structural uncertainty occurs when the correct mechanism, pathway or model structure is uncertain or when several alternative explanations are plausible.

Measurement uncertainty reflects experimental error, instrument limitations and biological variability in the data used for model calibration or validation.

Instead of reporting a single predicted value, models should communicate uncertainty using **ranges, confidence intervals, probability distributions or alternative scenarios**.

Sensitivity analysis identifies which uncertain parameters or assumptions have the strongest effect on model outputs and therefore deserve additional measurement or experimentation.

If uncertainty changes the preferred engineering decision, it must be treated as a **design-relevant variable rather than a reporting detail**.



From pathway diagram to engineering model

Biochemical pathway maps often emphasize molecular detail, while engineering models emphasize conserved quantities, rates and constraints.

A biochemical pathway may contain dozens of metabolites, enzymes and regulatory interactions, but not every molecular detail is equally important for an engineering objective.

The first modelling step is to identify the **key engineering states**, such as substrate pools, precursor availability, biomass, cofactors and product concentration.

Individual biochemical reactions can then be grouped into functional processes such as **substrate uptake, precursor formation, target synthesis, competing pathways and by-product formation**.

These processes are represented quantitatively through **rates, fluxes, yields and conservation relationships** rather than by listing every molecular reaction separately.

Constraints such as **enzyme capacity, cofactor availability, thermodynamics, transport and cellular resource limits** define the feasible operating region of the system.

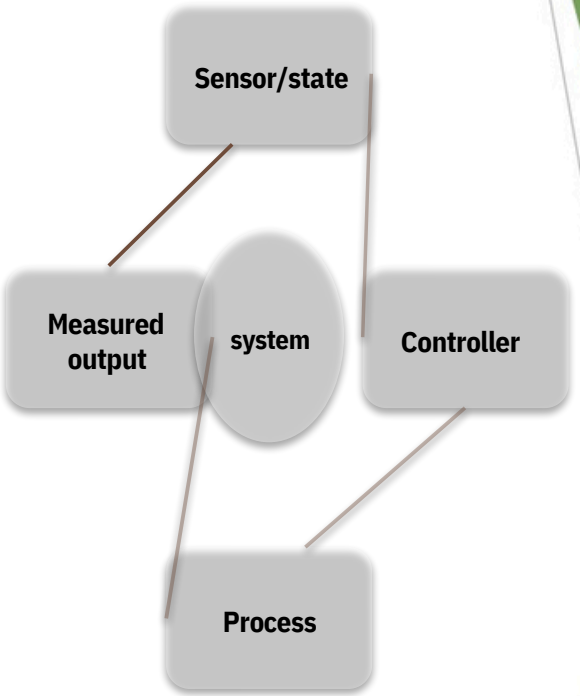
The resulting model links pathway behaviour to engineering outcomes such as **product yield, titer, productivity, selectivity or resource efficiency**.

Representing feedback

Feedback can stabilize a system, create oscillation or generate bistability depending on loop sign and delay.

- In a bioreactor, dissolved oxygen control is an engineered feedback loop; in a cell, transcriptional repression is a biological feedback loop.
- Conceptual models should identify where feedback closes and what is sensed.
- A control structure is only as useful as its measurement quality, response time and ability to manipulate the variable that truly limits performance.
- Effective control requires a measurable variable, a meaningful set point, an actuator with sufficient authority, and response dynamics fast enough to reject disturbances...
- SYSTEM-LEVEL CONSEQUENCE

Feedback links measurement to corrective action.



Representing delays

Biological systems often respond with a time delay because sensing, transport, transcription, translation, protein maturation and population changes occur on different time scales.

Delays separate cause from observed effect. A change in input may occur immediately, while the measurable biological response appears only after several intermediate processes.

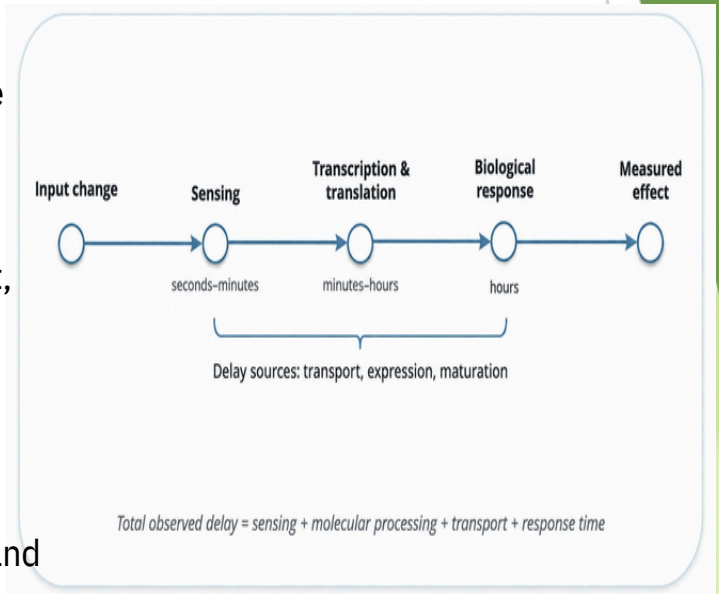
At the molecular level, delays can arise from **gene transcription, protein translation, folding and maturation**; at larger scales they may result from transport, cell growth or population turnover.

Delays can strongly influence **transient behaviour**, producing slow responses, overshoot or oscillations even when the final steady state is stable.

In feedback-controlled biological systems, a long delay between measurement and response can reduce stability and make corrective actions arrive too late.

Conceptual models should therefore identify important delays explicitly and, when possible, associate them with an **approximate characteristic time scale**.

The appropriate representation may range from a simple time-lag annotation to a sequence of intermediate states or a formal dynamic model.



Interface modelling

MaInterfaces are critical regions where two subsystems exchange mass, energy, information or mechanical forces. Many biological engineering failures occur not within the subsystems themselves, but at the boundary between them. Important interfaces include **cell–material, gas–liquid, tissue–device, membrane–fluid and upstream–downstream process boundaries**.

At an interface, system performance is governed by **driving forces, resistance, transfer capacity and compatibility** between the connected subsystems.

Interface models should identify which quantities are exchanged, such as **oxygen, nutrients, metabolites, heat, mechanical stress or electrical signals**.

A local limitation at the interface can dominate the behaviour of the entire system even when both individual subsystems perform well independently.

Relevant engineering variables may include **mass-transfer coefficients, permeability, surface properties, contact area, concentration gradients and interfacial reaction rates**.

Explicit interface modelling is especially important when **transport, adhesion, signal conversion or material compatibility** limits overall performance.

SUBSYSTEM A
cell / gas / upstream unit
→

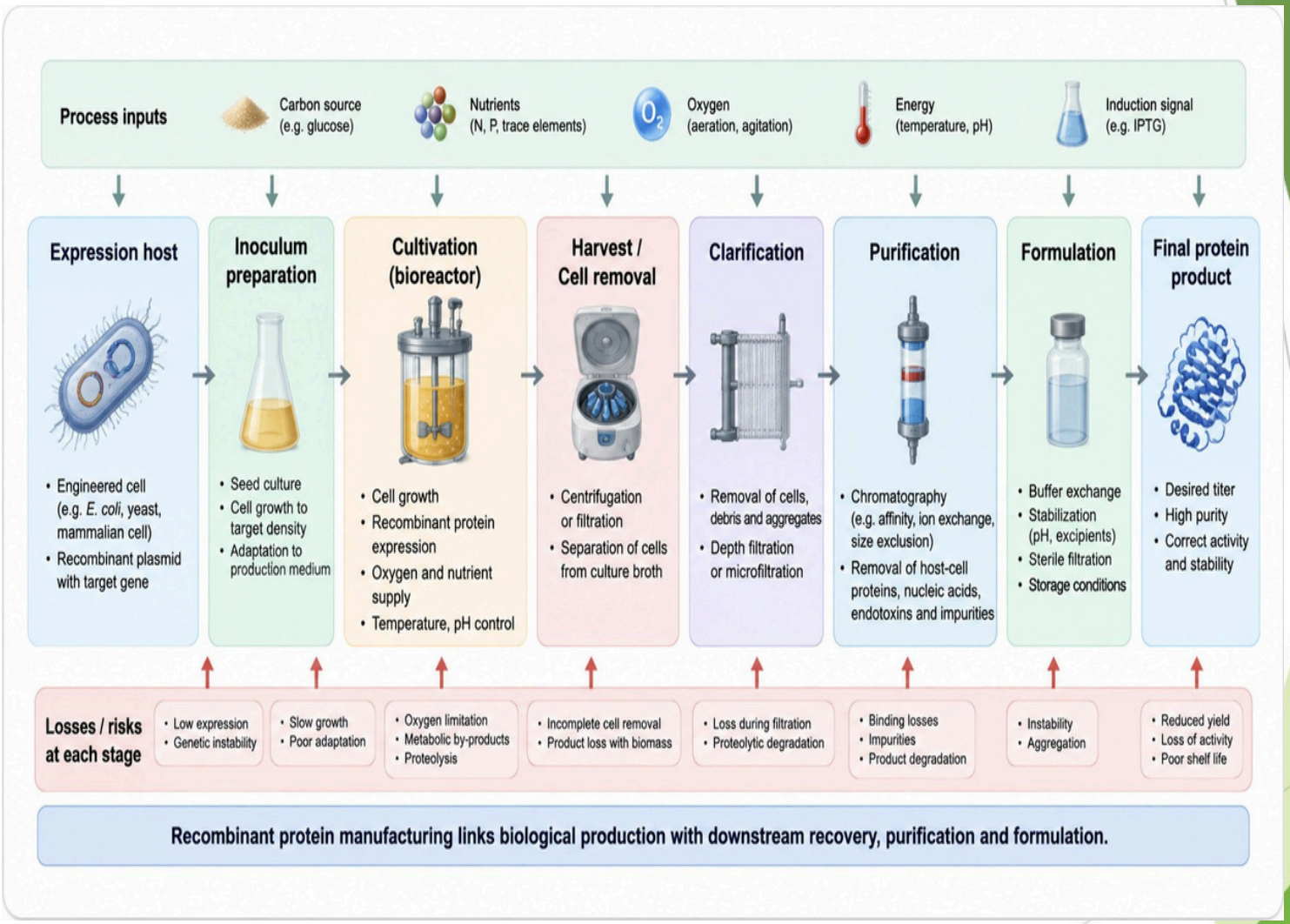
INTERFACE
Driving force
ΔC • ΔP • ΔT • signal
Resistance
m embrane • boundary layer • contact impedance
Capacity
area • permeability • transfer coefficient
→

SUBSYSTEM B
m aterial / liquid / downstream unit
↓

SYSTEM OUTPUT
transport rate • signal quality • product recovery

Можно добавить сверху небольшой блок:
INTERFACE CONDITIONS
surface chemistry • geometry • flow • temperature

Case model: recombinant protein process



Case model: microbial biosensor

A microbial biosensor converts the presence or concentration of a target analyte into a measurable biological signal through a sequence of recognition, intracellular processing and reporter generation.

The process begins with **sample preparation and analyte transport**, which determine whether the target molecule can reach the engineered microbial sensor.

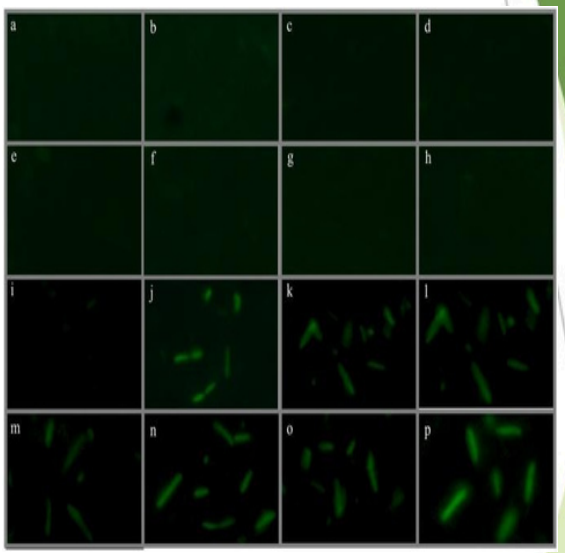
A **recognition element**, such as a receptor, transcription factor or regulatory promoter, detects the analyte and activates the engineered genetic circuit.

The intracellular circuit converts recognition into **reporter expression**, commonly fluorescence, luminescence, pigment formation or another measurable output.

The final readout depends not only on the genetic circuit, but also on **cell viability, analyte uptake, reporter maturation, incubation time and measurement conditions**.

Important disturbances include **sample inhibitors, background fluorescence, temperature variation, pH, competing molecules and biological variability**.

Biosensor performance should be evaluated using **limit of detection, dynamic range, selectivity, response time, reproducibility, false-positive rate and signal stability**.



Case model: tissue-engineered scaffold

A tissue-engineered scaffold acts as a temporary three-dimensional environment that supports cell attachment, nutrient transport and tissue formation while gradually remodeling or degrading as new extracellular matrix is produced.

The system contains four tightly coupled components: **cells, scaffold architecture, biochemical environment and mechanical cues**.

Scaffold properties such as **porosity, pore size, stiffness, surface chemistry and degradation rate** determine how cells attach, migrate and organize within the material.

Oxygen, nutrients and signaling molecules must diffuse through the scaffold, while metabolic waste must be removed; insufficient transport can create poorly viable regions inside thick constructs.

Cells respond to the local environment through **adhesion, proliferation and differentiation**, and simultaneously remodel the scaffold by depositing new extracellular matrix.

Scaffold degradation should be coordinated with **new tissue formation**: degradation that is too rapid can reduce mechanical support, whereas very slow degradation may limit tissue remodeling.

The final engineering outcome is determined by the interaction of **transport, mechanics and cell biology**, rather than by any single material property.

SCAFFOLD DESIGN
porosity • stiffness • degradation • surface chemistry
↓
CELL–MATERIAL INTERACTION
adhesion • migration • proliferation
↔
MASS TRANSPORT
oxygen • nutrients • waste
↓
TISSUE DEVELOPMENT
ECM deposition • differentiation • remodeling
↓
FUNCTIONAL TISSUE
mechanical integrity • biological function

Case model: wastewater treatment

Wastewater treatment can be represented as a coupled biological–process system in which microbial communities convert organic matter and nutrients while reactor operation controls the conditions required for efficient removal.

The main influent variables include **organic carbon, nitrogen, phosphorus, suspended solids and hydraulic flow**, all of which can vary strongly over time.

Different microbial groups perform complementary transformations: **heterotrophs remove organic carbon, nitrifiers oxidize ammonia, and denitrifiers convert nitrate to nitrogen gas**.

Reactor performance is controlled through variables such as **aeration rate, dissolved oxygen, hydraulic retention time, solids retention time and sludge recycle**.

Biological conversion produces **treated water, biomass, carbon dioxide, nitrogen gas and excess sludge**, linking microbial ecology directly to process-scale mass balances.

Important disturbances include **changes in influent composition, toxic compounds, temperature shifts and hydraulic overload**, which can disrupt microbial activity and treatment efficiency.

The model is evaluated using outputs such as **COD/BOD removal, ammonia and nitrate concentrations, phosphorus removal, sludge production and effluent quality**.

INFLUENT WASTEWATER *COD • NH_4^+ • PO_4^{3-} • solids*

→

BIOLOGICAL REACTOR *carbon oxidation • nitrification • denitrification*

↓

PROCESS CONTROL *aeration • DO • SRT • recycle* →

SOLID–LIQUID SEPARATION *clarifier / membrane*

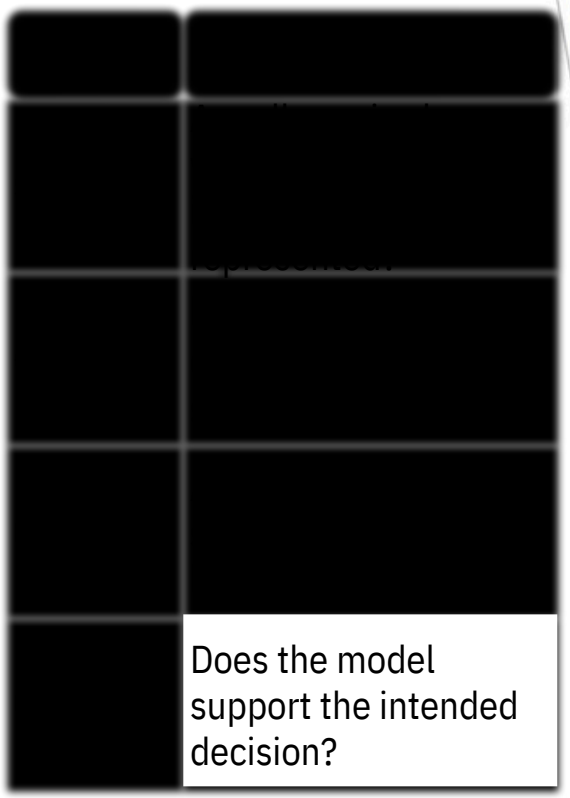
→

TREATED EFFLUENT *low COD • low N • low P*

Validation at conceptual level

A conceptual model is validated by checking whether it contains the mechanisms needed to explain observed behavior.

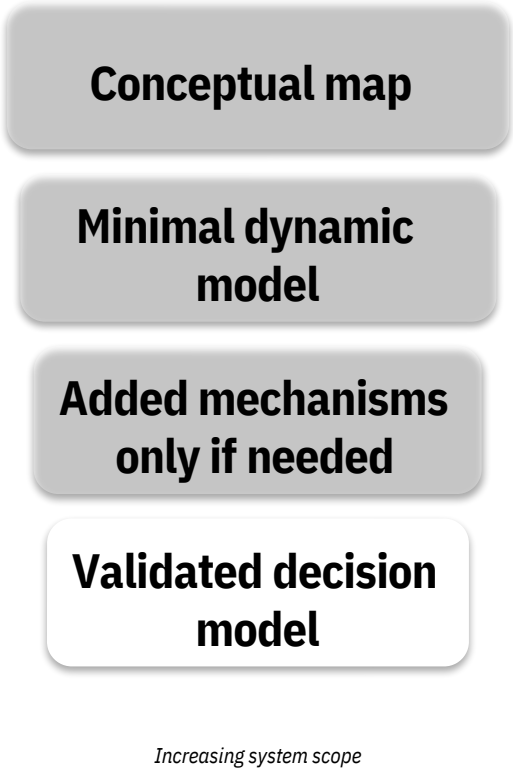
- Ask whether important outputs can change in the model when known interventions are applied.
- Compare the map with literature, expert knowledge and process observations.
- Conceptual validation precedes parameter fitting.
- Mechanistic reasoning, small-scale experiments, replicated data and field-scale validation answer different questions; engineering decisions should state which evidence level...
- TEACHING POINT Confidence should match the strength of evidence.



Model parsimony

Parsimony means using no more structure than necessary to explain and predict the behavior of interest.

- A model is a purposeful simplification, not a complete copy of reality.
- Redundant complexity increases parameter demand and can hide causal interpretation.
- A parsimonious model can be extended later when new evidence shows missing mechanisms.
- This is especially important in biological systems with limited data.
- Its value depends on whether the chosen states, interactions and assumptions are sufficient for the intended decision.



Transition to equations

A conceptual model becomes quantitative when system states, flows, constraints and assumptions are translated into mathematical relationships that describe how the system changes over time.

Each **stock or accumulated quantity** becomes a state variable, such as biomass, substrate concentration, product concentration or dissolved oxygen.

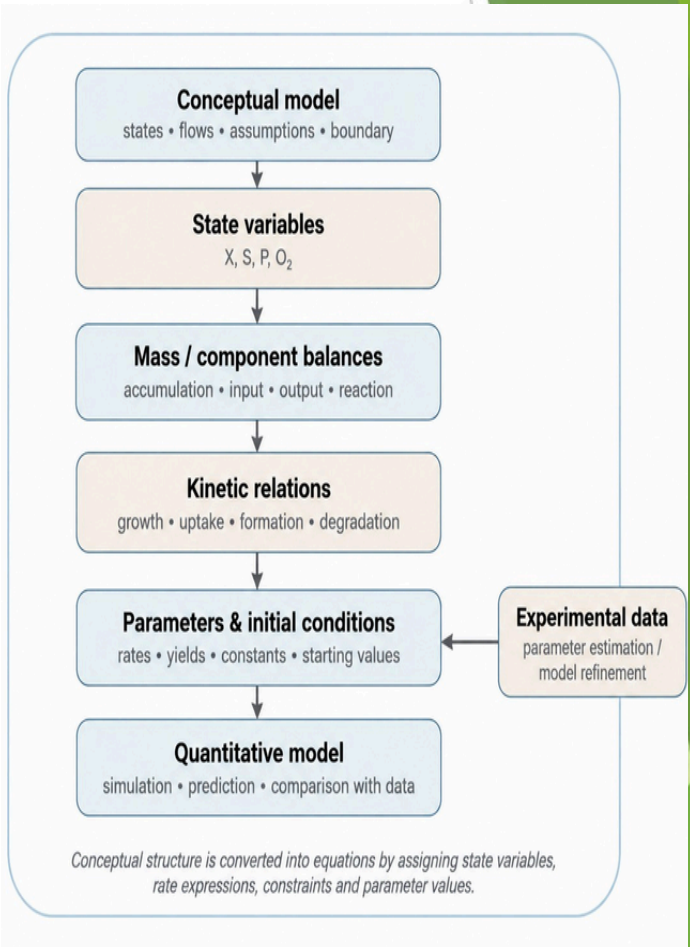
Flows and transformations become rate terms describing production, consumption, transport, degradation or removal.

Conservation balances define the basic mathematical structure by linking accumulation to inputs, outputs, generation and consumption.

Kinetic relationships describe how process rates depend on state variables, operating conditions and model parameters.

Algebraic constraints can represent **equilibria, capacity limits, operating rules or relationships between variables** that do not require separate dynamic equations.

Parameter values are then estimated from **experimental measurements, literature data or model calibration**, allowing the equations to generate quantitative predictions.



Transition to experiments

A conceptual model becomes experimentally useful when its uncertain mechanisms, unmeasured states and competing hypotheses are translated into specific measurements and perturbations.

Основной текст слева

Each uncertain relationship in the model represents a **testable hypothesis** that can be challenged with targeted experiments. Unmeasured or poorly constrained state variables identify **measurement needs**, such as biomass, metabolite concentration, reporter activity or transport rate.

Experimental perturbations should target variables that can **distinguish between competing mechanisms**, rather than simply generate more data.

Time-course measurements are particularly valuable for dynamic models because they reveal delays, transient responses and feedback that endpoint measurements may miss.

Experimental observations are compared with model predictions to identify **parameter errors, missing mechanisms or incorrect assumptions**.

The resulting evidence is then used to **update the model and design the next experiment**, creating an iterative model–experiment cycle.

MODEL PREDICTION
expected behaviour • uncertainty
↓
EXPERIMENT DESIGN
perturbation • controls • measurements
↓
EXPERIMENTAL DATA
time course • phenotype • response
↓
MODEL–DATA COMPARISON
agreement • residuals • unexpected behaviour
↓
MODEL UPDATE
parameters • assumptions • mechanisms

Communication value

Conceptual models provide a shared visual language that allows biologists, engineers, data scientists and decision-makers to discuss the same system even when they use different technical terminology or mathematical tools.

A clear conceptual model makes the **system boundary, variables, interactions and assumptions** visible to everyone involved in the project.

Visual representations help different specialists identify where their interpretations of the same biological system **agree, differ or remain uncertain**.

Explicit notation improves **traceability**, linking verbal assumptions to measurements, equations, experiments and engineering decisions.

Standardized symbols, arrows and labels reduce ambiguity and make models easier to **review, reproduce and modify**.

Conceptual diagrams are particularly valuable during interdisciplinary work because they allow biological mechanisms, process constraints and experimental evidence to be discussed within a **single system representation**.

As the project develops, the same conceptual structure can serve as a bridge between **qualitative reasoning, quantitative modelling and experimental design**.

Take-home modelling workflow

A useful engineering model is built iteratively: define the purpose, choose the system boundary, identify the important states and interactions, make assumptions explicit, and validate the model against evidence.

1. Define the purpose. Start with the engineering question: what should the model explain, predict, compare or help decide?

2. Choose the system boundary. Specify what is included in the model, what is treated as external context and which exchanges cross the boundary.

3. Identify key states and flows. Select the quantities that accumulate, change or determine system performance.

4. Represent interactions and assumptions. Make causal links, feedback, delays, disturbances and simplifications explicit.

5. Check consistency. Use conservation relationships and qualitative reasoning to verify that the proposed structure can explain known system behaviour.

6. Add quantitative detail only when needed. Introduce equations, kinetics and parameters when they improve prediction, interpretation or the engineering decision.

7. Validate and refine. Compare predictions with experimental data, identify mismatches and update the model iteratively.

Student Tasks and Discussion

1. Define one system boundary relevant to this lecture and explain what is deliberately excluded from the analysis.
2. Identify the three most important measurable variables for judging performance; state their units and why they matter.
3. Draw one conceptual diagram (process, causal loop, balance, or decision map) using the notation introduced in the lecture.
4. Describe one realistic failure mode or limitation and propose an engineering control, design change or monitoring strategy.
5. Locate one peer-reviewed article from the last five years that extends the lecture topic and summarize its engineering contribution, evidence and remaining limitation.

Selected References and Further Reading

1. Le Novère, N. et al. The Systems Biology Graphical Notation. Nat Biotechnol 27, 735-741 (2009).

2. Le Novère, N. et al. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Biotechnol 23, 1509-1515 (2005).

3. Hucka, M. et al. The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524-531 (2003).

4. Waltemath, D. et al. Reproducible computational biology experiments with SED-ML-The Simulation Experiment Description Markup Language. BMC Syst Biol 5, 198 (2011).

5. Bergmann, F.T. et al. COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. BMC Bioinformatics 15, 369 (2014).

6. Lloyd, C.M., Halstead, M.D.B. & Nielsen, P.F. CellML: its future, present and past. Prog Biophys Mol Biol 85, 433-450 (2004).

7. Li, C. et al. BioModels Database: an enhanced, curated and annotated resource for published quantitative kinetic models. BMC Syst Biol 4, 92 (2010).

8. Choi, K. et al. Tellurium: an extensible Python-based modeling environment for systems and synthetic biology. BioSystems 171 74-79 (2018).

9. Medley, J.K. et al. Tellurium notebooks-an environment for reproducible dynamical modeling in systems biology. PLoS Comput Biol 14, e1006220 (2018).

10. Alon, U. Network motifs: theory and experimental approaches. Nat Rev Genet 8, 450-461 (2007).