

Putting the *FLOW* in workflow: Using plume prediction AI to quantify groundwater risk and liability

Nico Higgs B.Sc., M.Sc.
Steven Mamet B.Sc., M.Sc., Ph.D.
Steven Siciliano, B.Sc., Ph.D.
Environmental Material Science Inc.

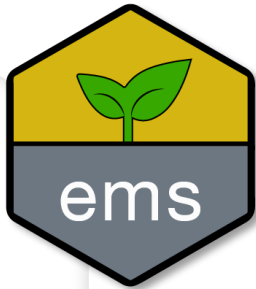
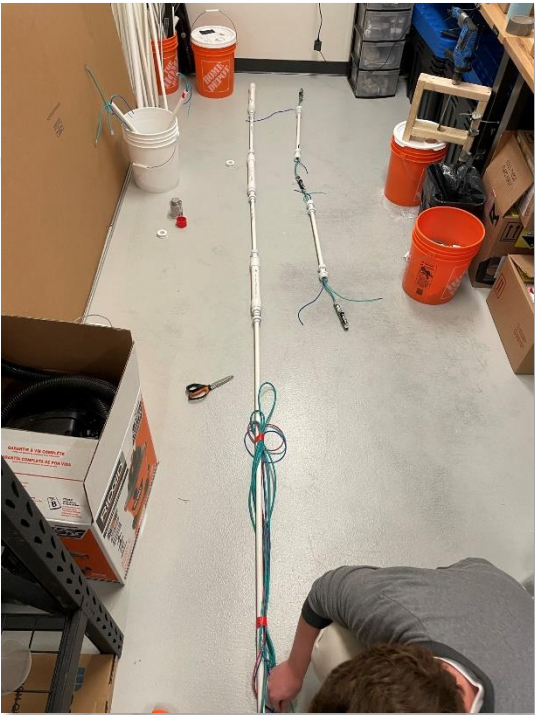
www.ems-inc.ca



ems

Environmental
Material Science

Who is EMS?



COMPANY FOUNDING

Early Adopters
FCL, Suncor, Transgas



RELEASE OF SOFTWARE-ENABLED HARDWARE

PlumeFutures

2000

2019

2022

2023

2024

R&D

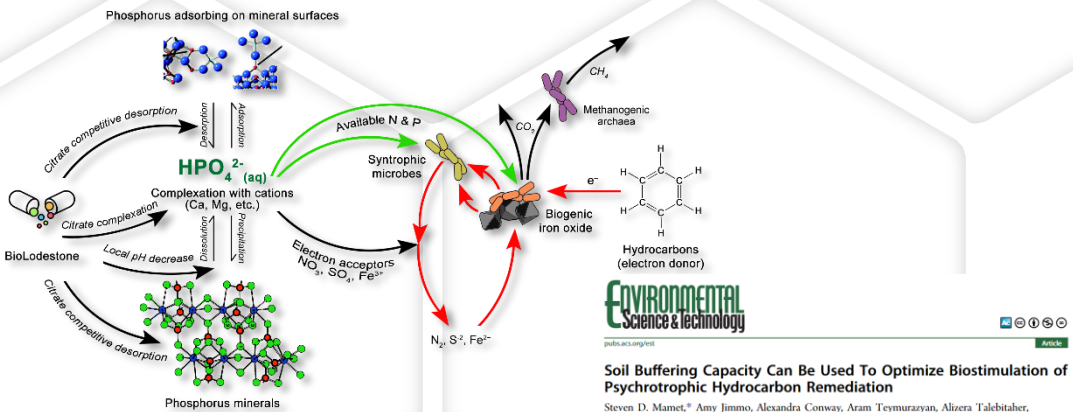
\$10M of Sponsored Research by
Siciliano Group into In Situ
Remediation Technologies

FUNDING

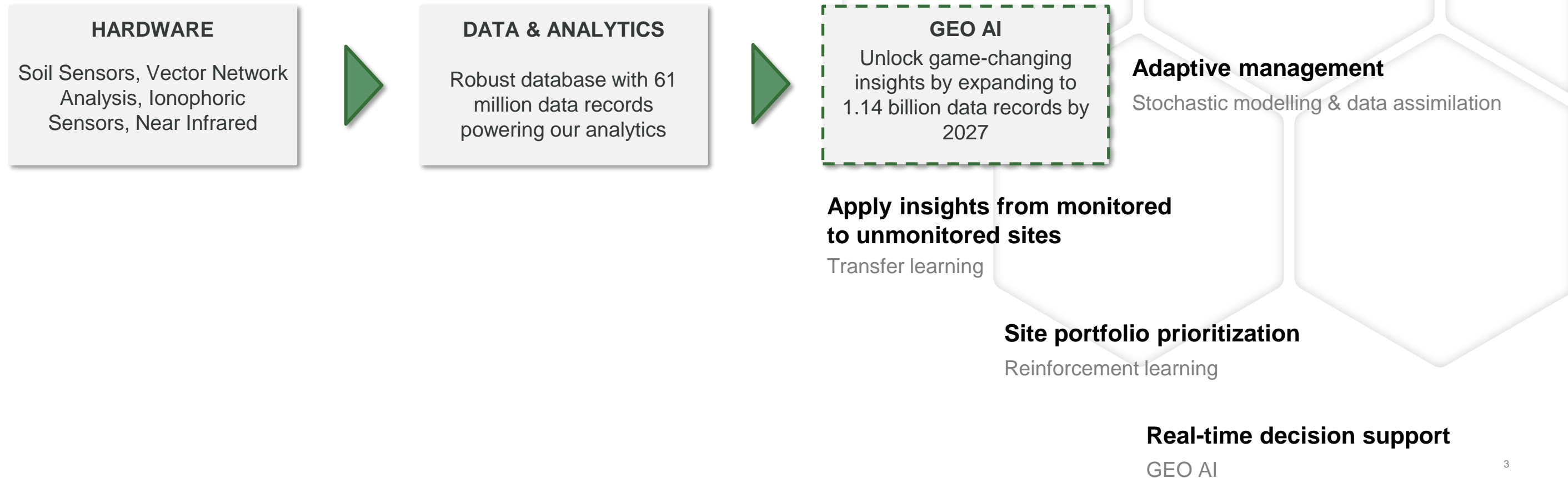
\$1.6M Raised from
Family Offices, Angels,
and Syndicates

EXPANSION & SCALING

US, EU, and Australian Sales



What if we created the world's largest water/soil contaminant database?





WaterSense Operation

Gas permeable membrane keeps the sensor **water-proof**, while allowing gases to diffuse to the **NDIR** sensor.

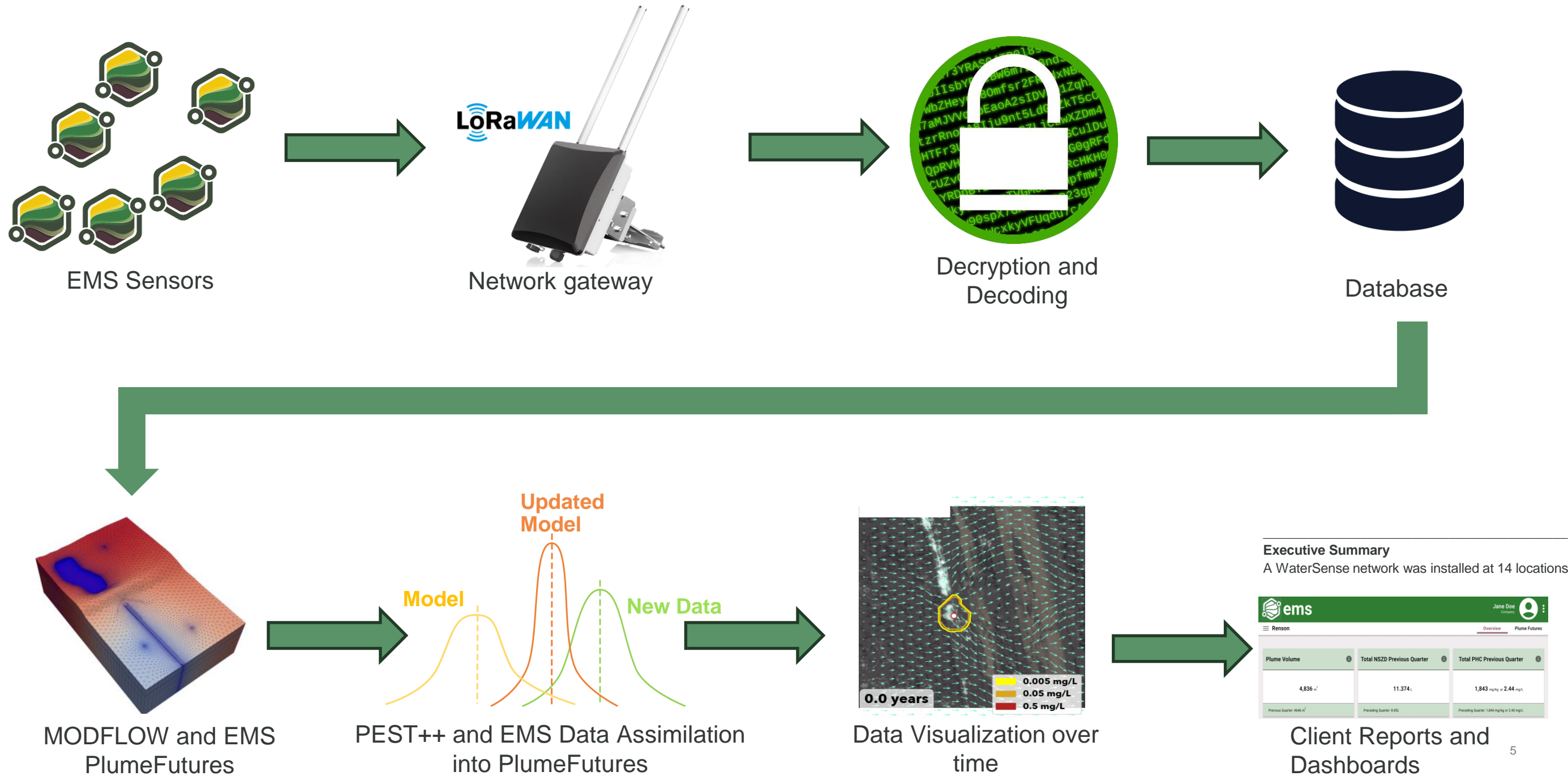
Measures **CO₂**, **CH₄**, & **PHC vapours** – calculate dissolved concentrations. Measures **temperature** & **pressure** directly.

Line or battery powered. **Multi-year battery life**. Solar panel add-on for recharging batteries.

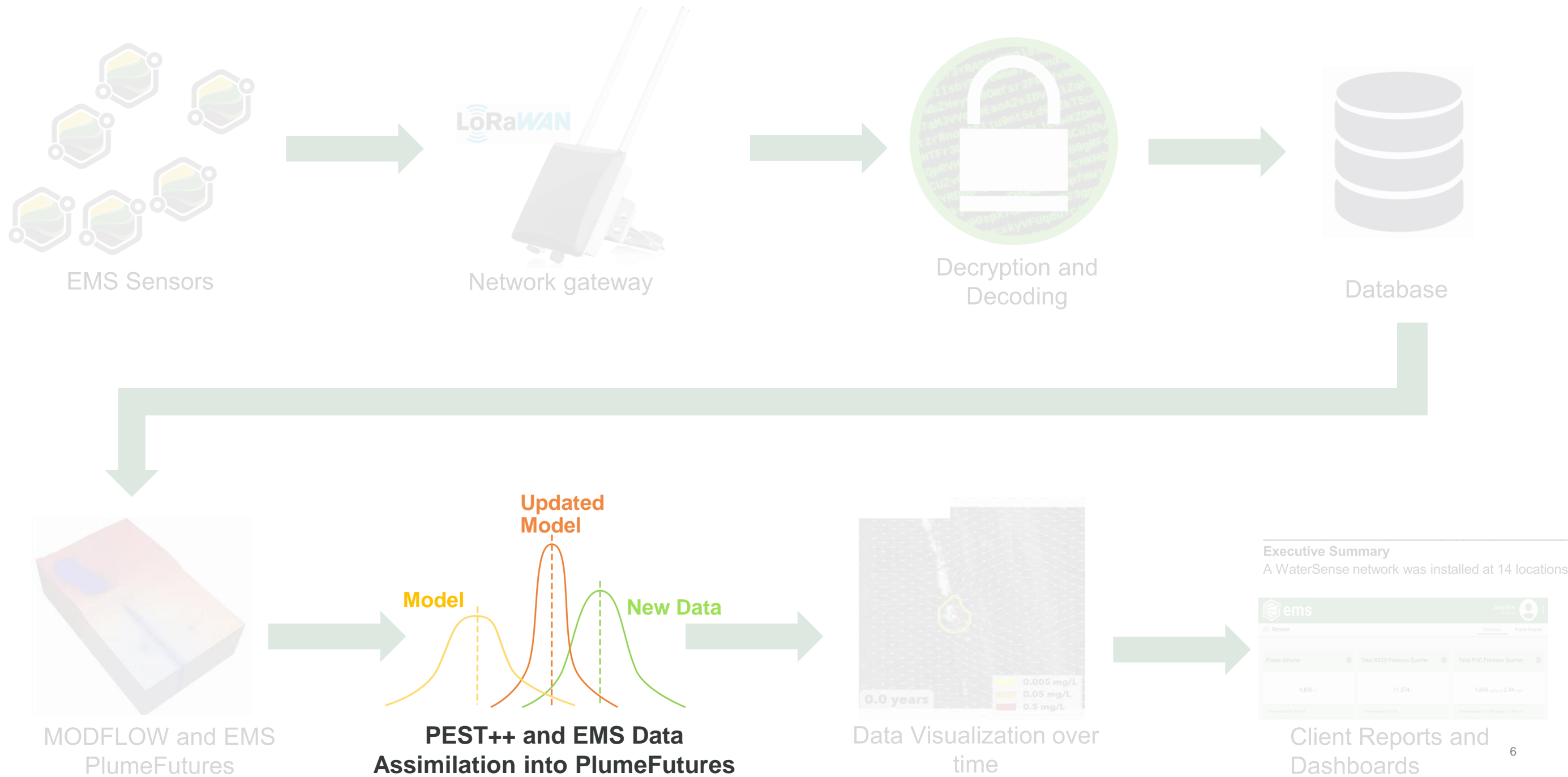
Flexible sample times from minutes to days. Typically, every **30 – 60 minutes**.

Automated data communication and storage. Reduced site trips. Viewable trends on-line.

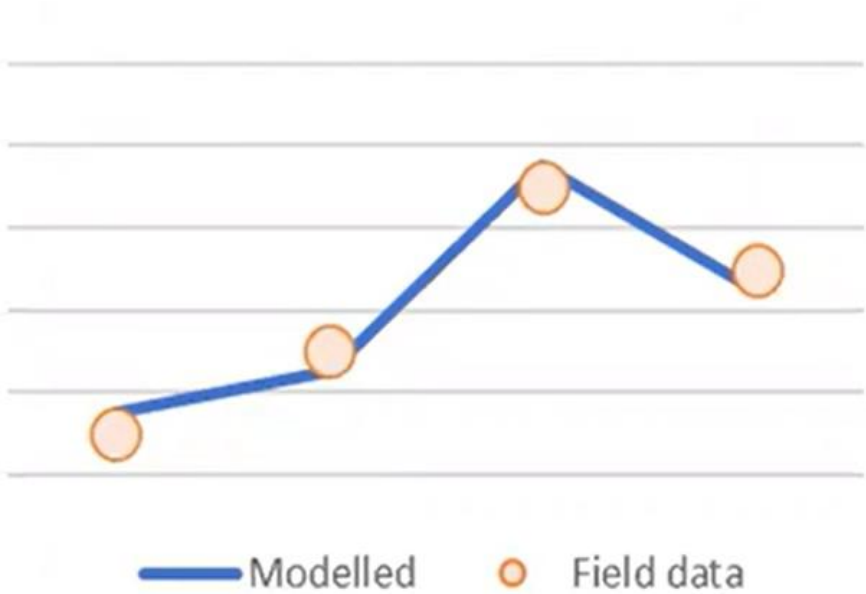
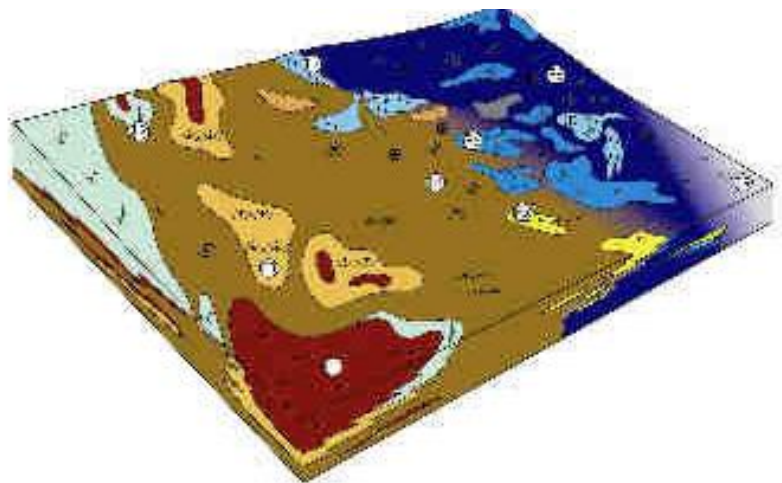
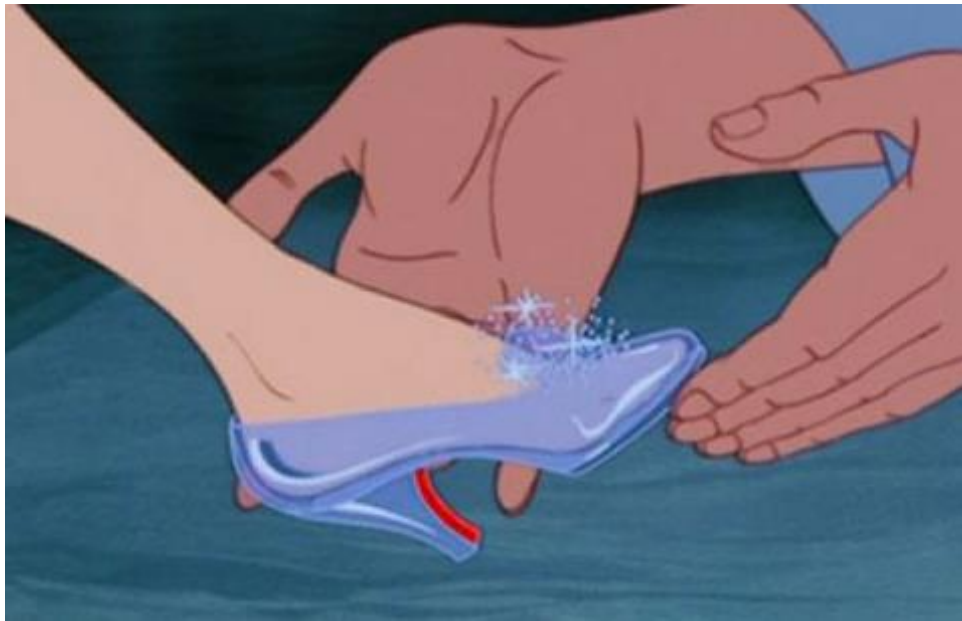
Data Flow



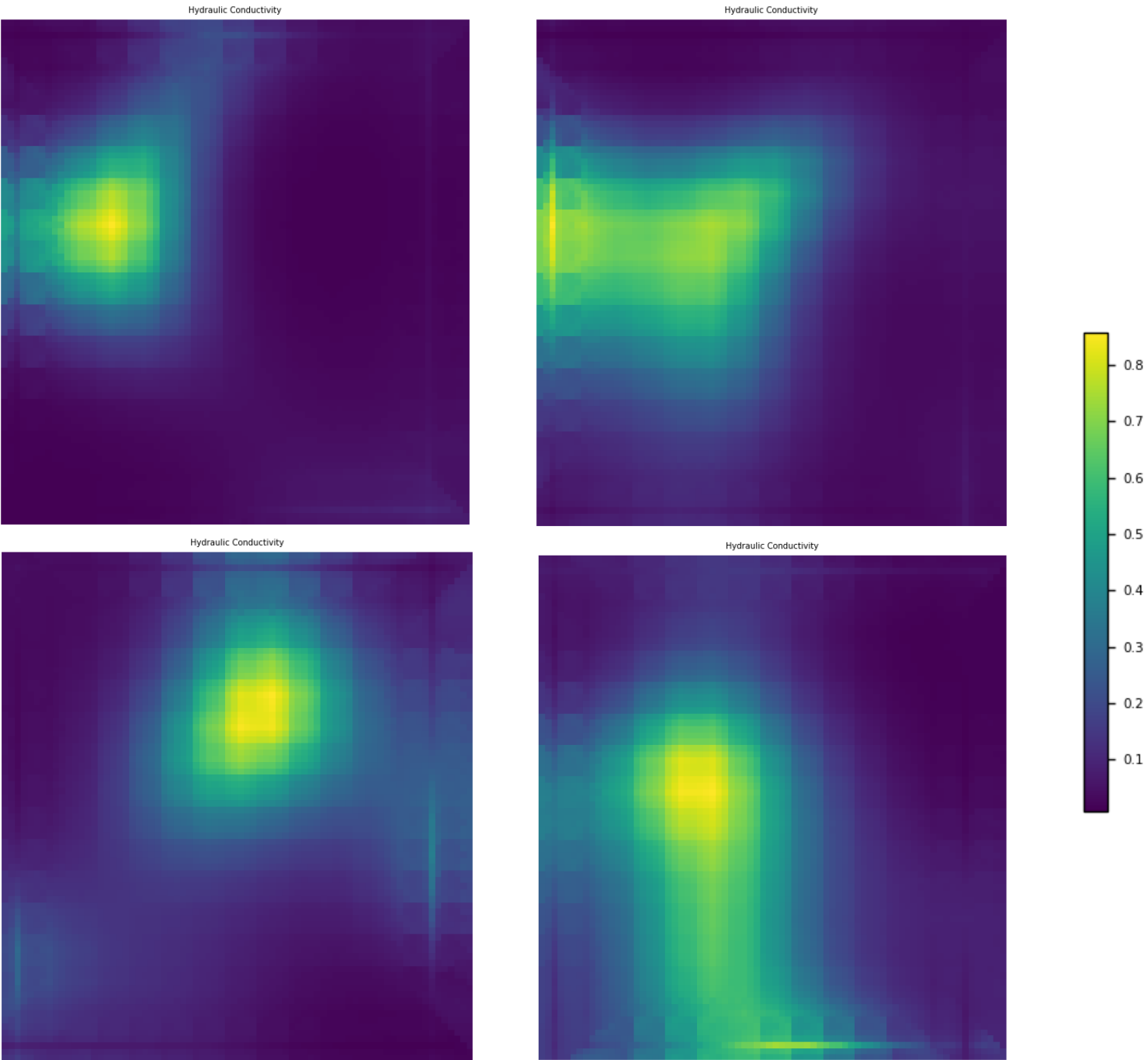
Data Flow



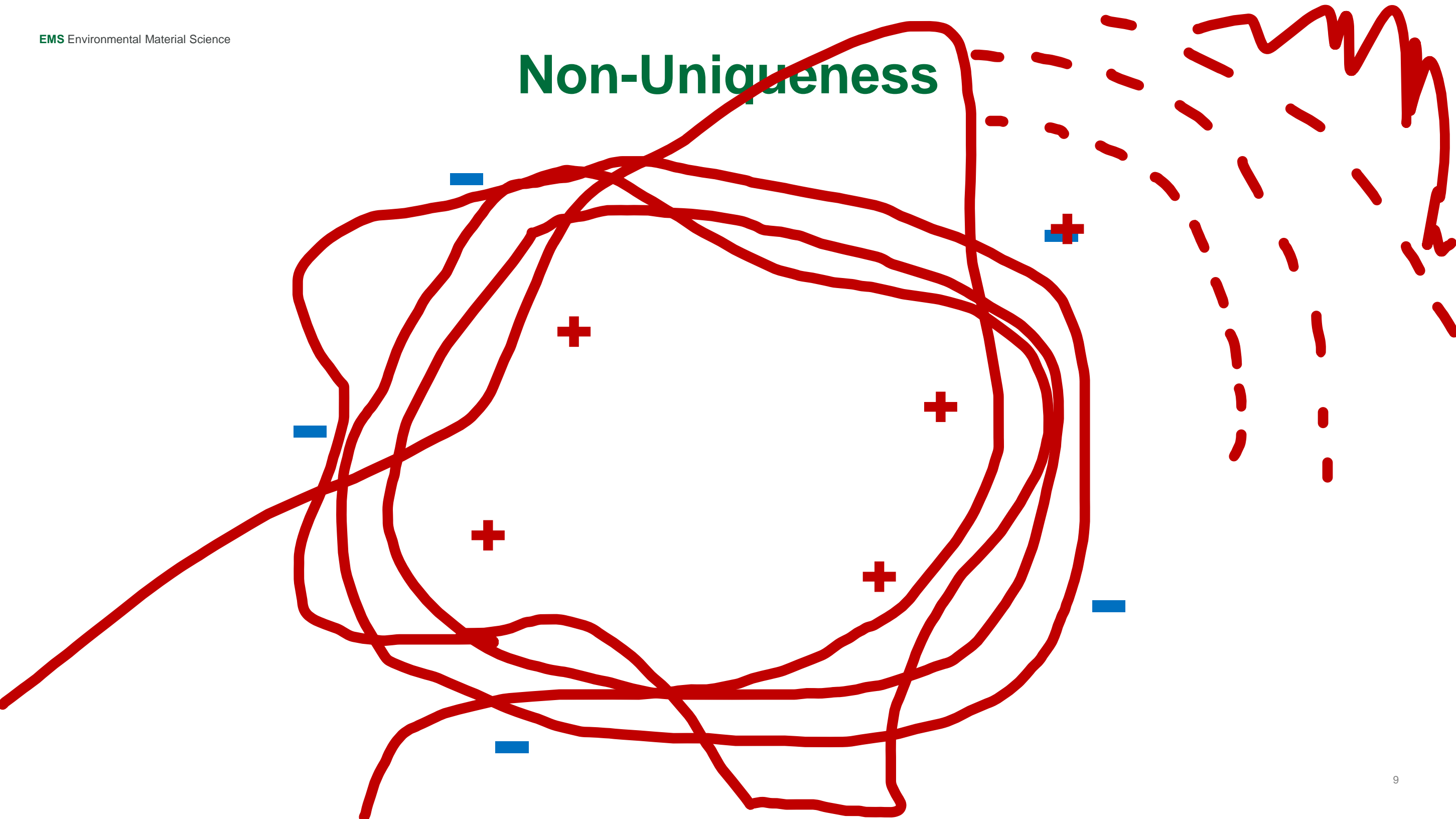
Cinderella Syndrome



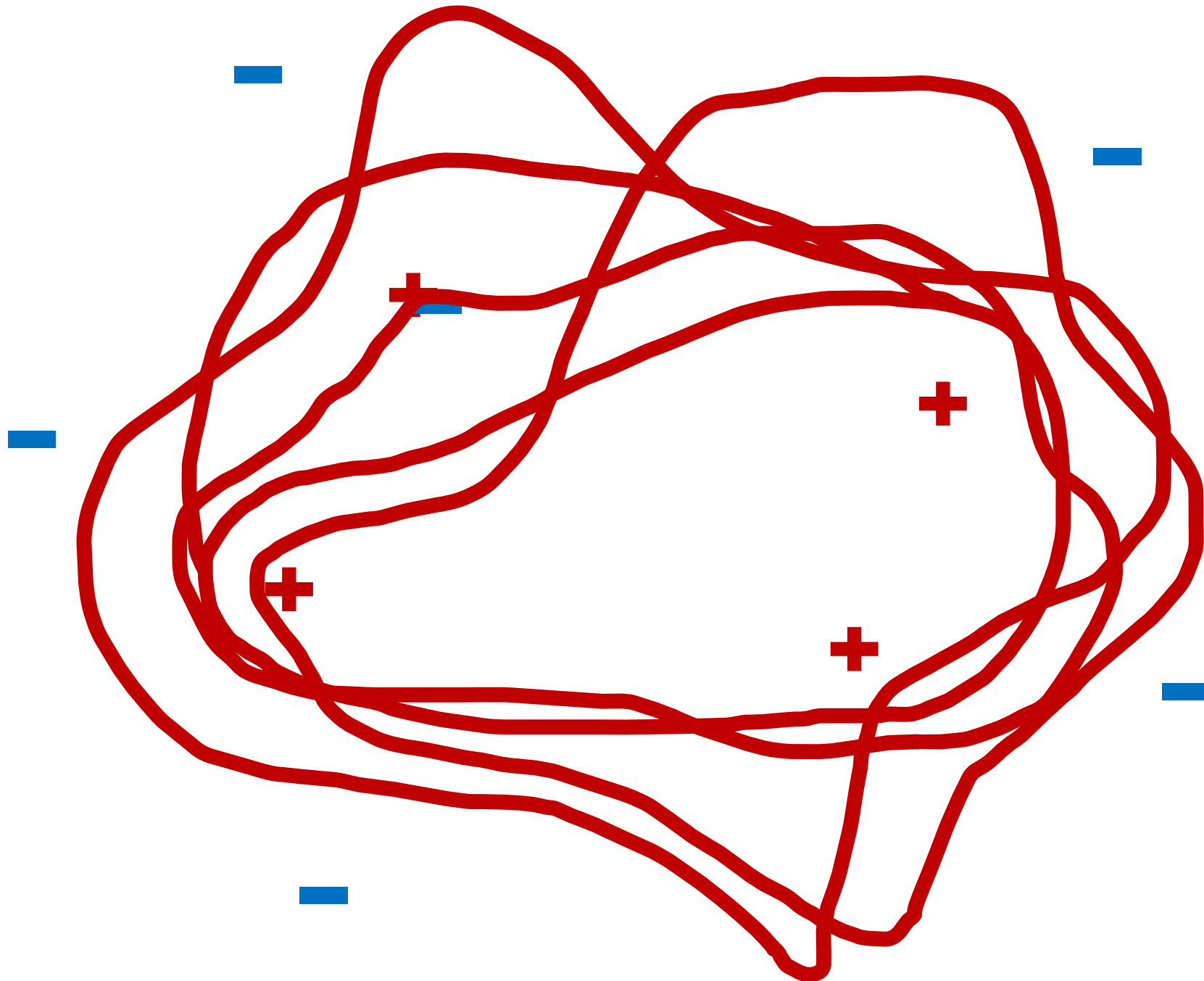
Non-Uniqueness



Non-Uniqueness

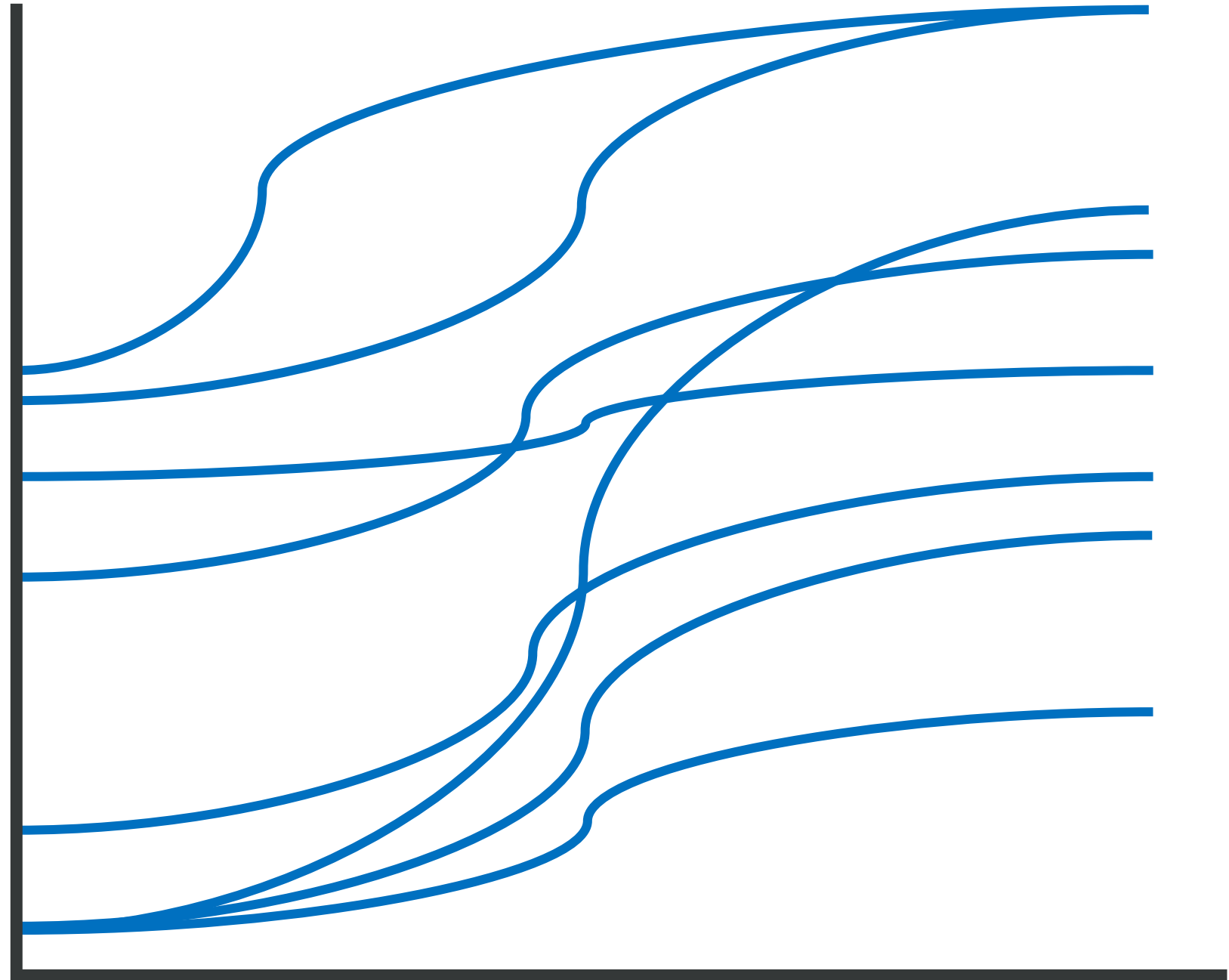


Non-Uniqueness



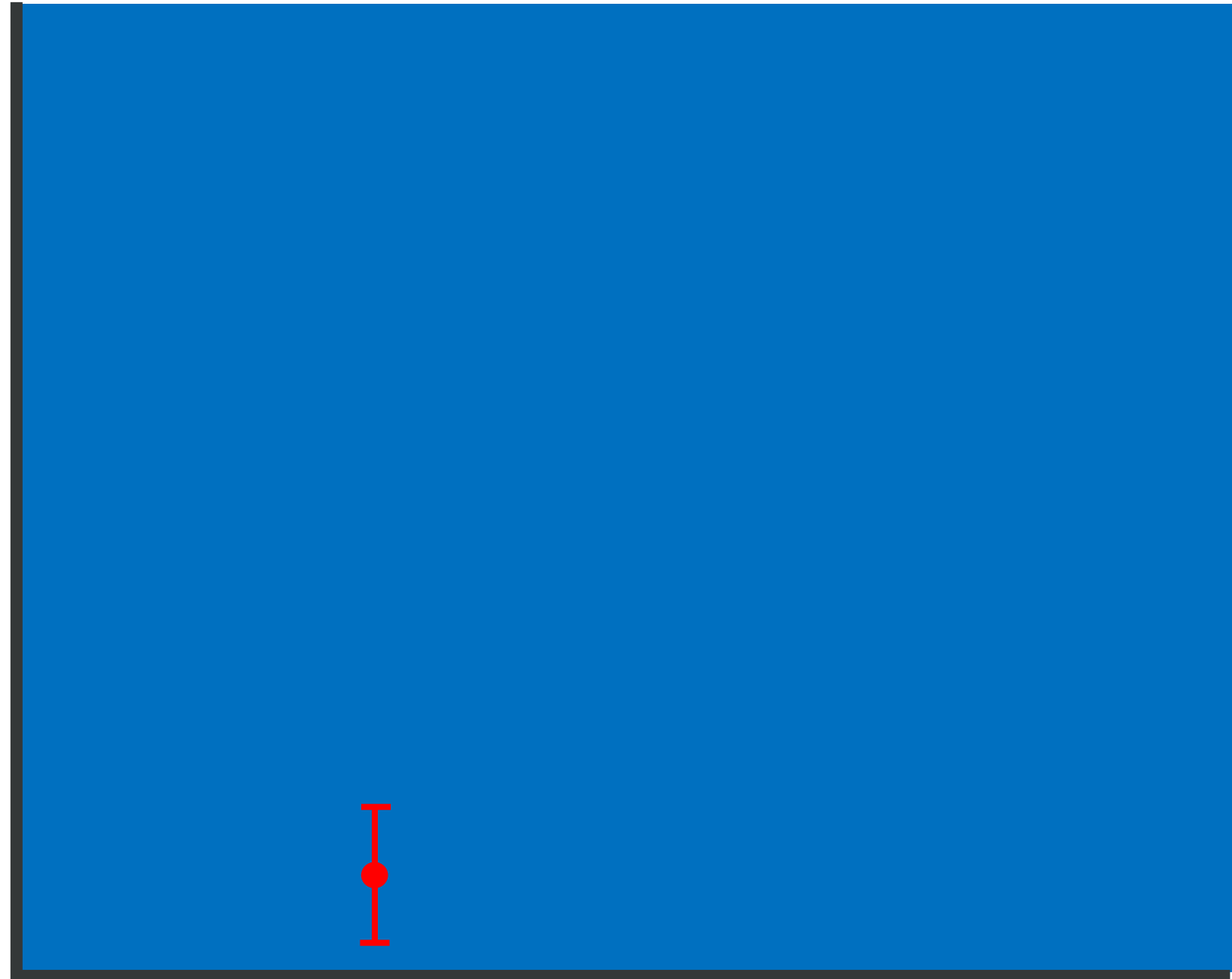
Machine Learning

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



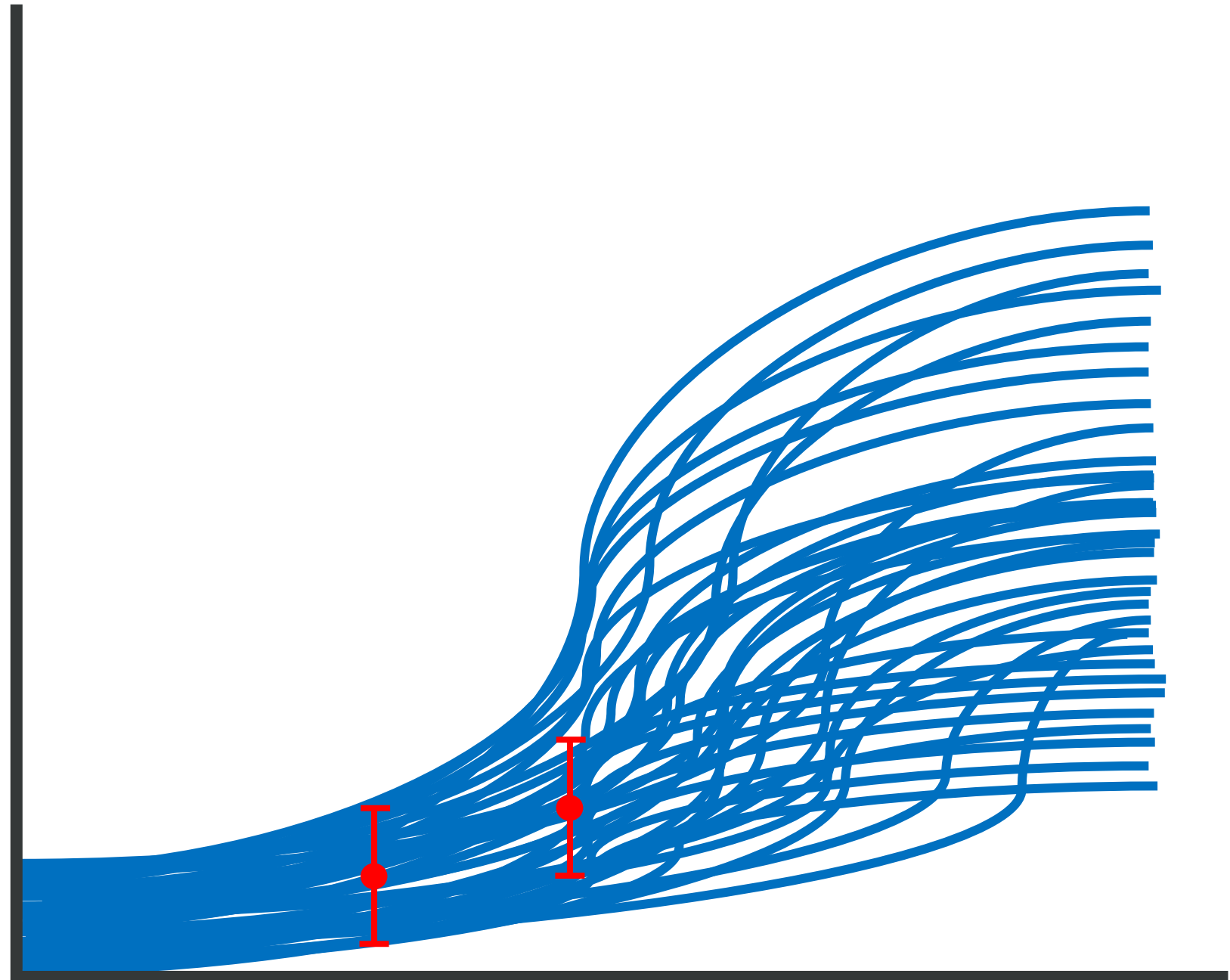
Machine Learning

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



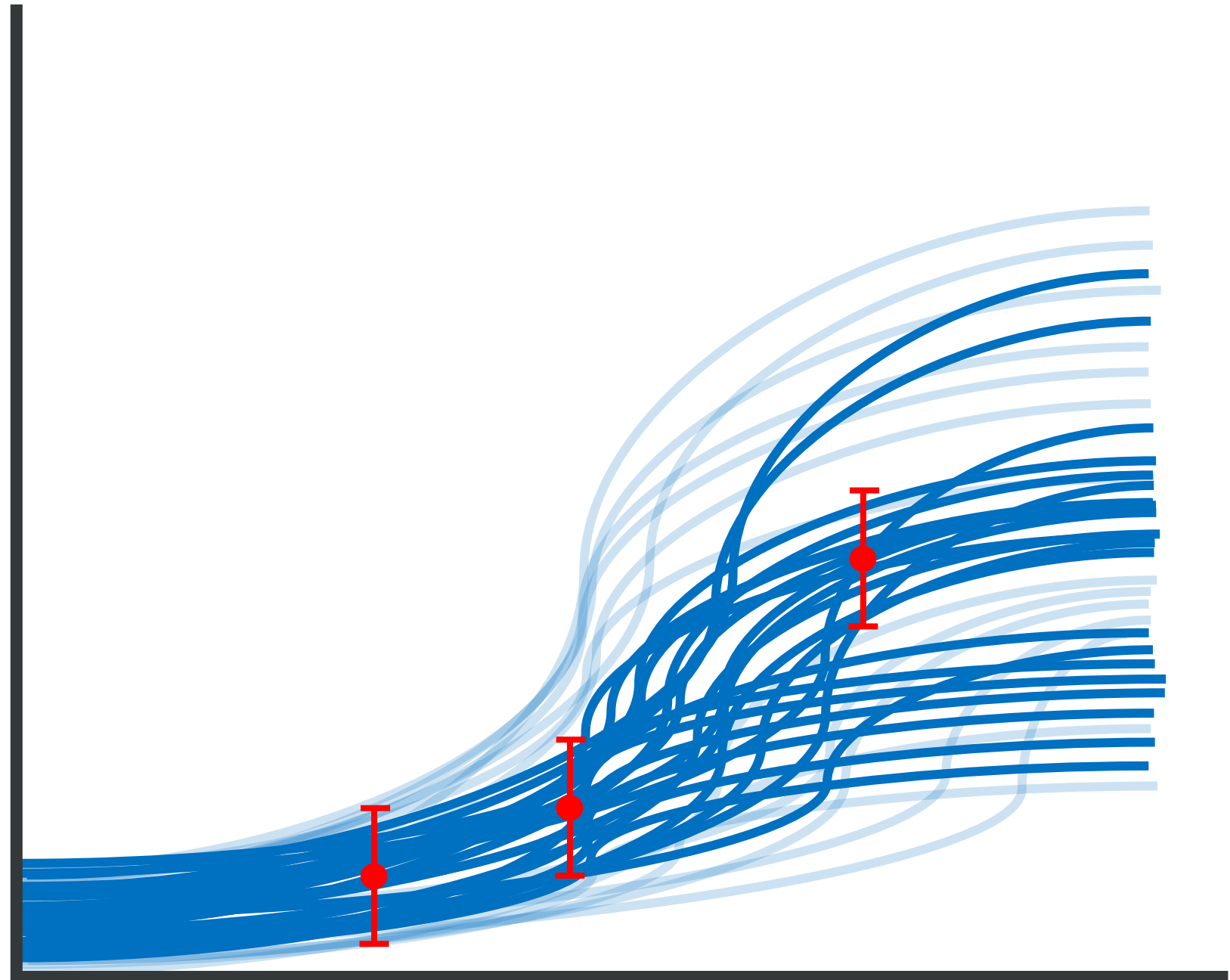
Machine Learning

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



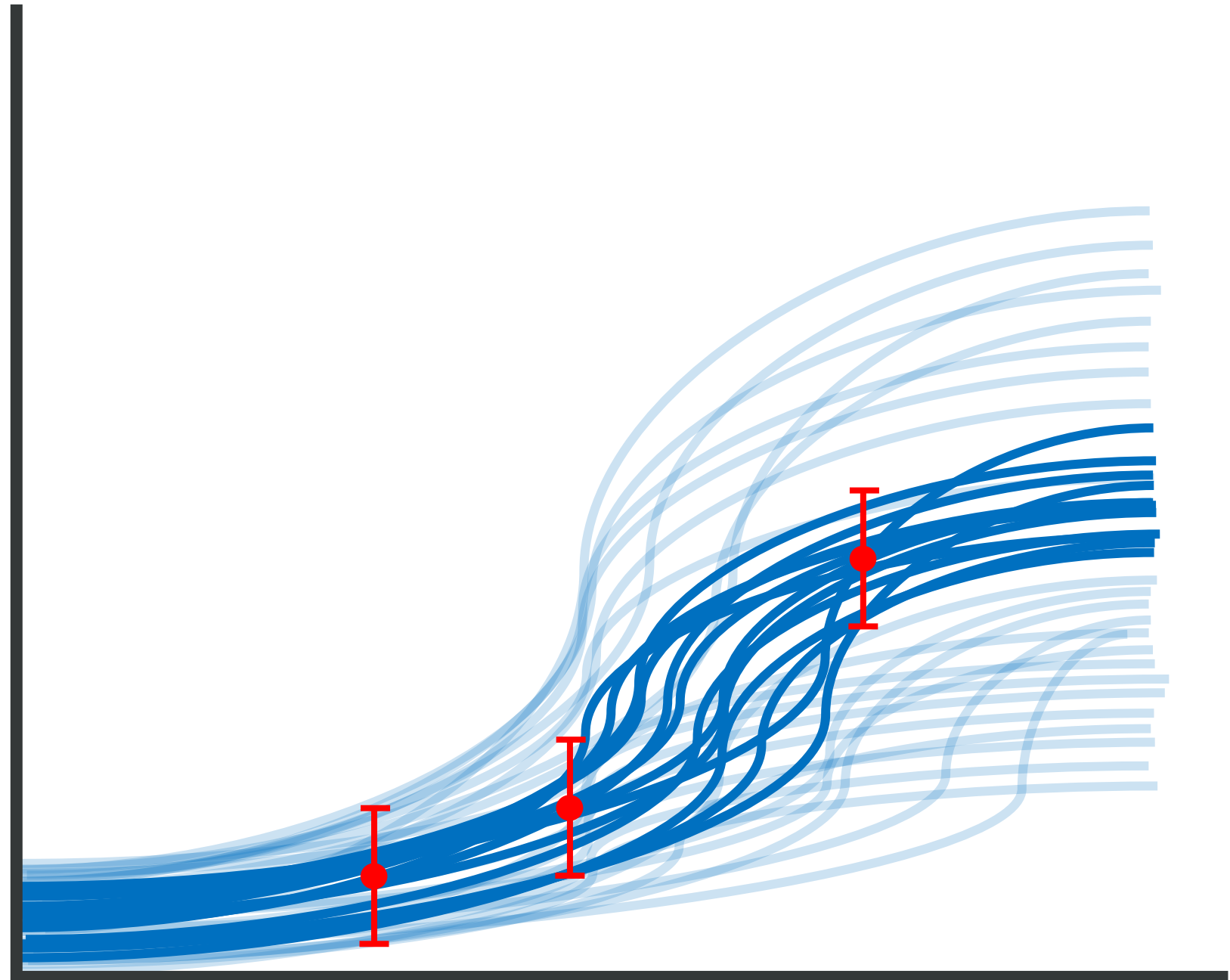
Machine Learning

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



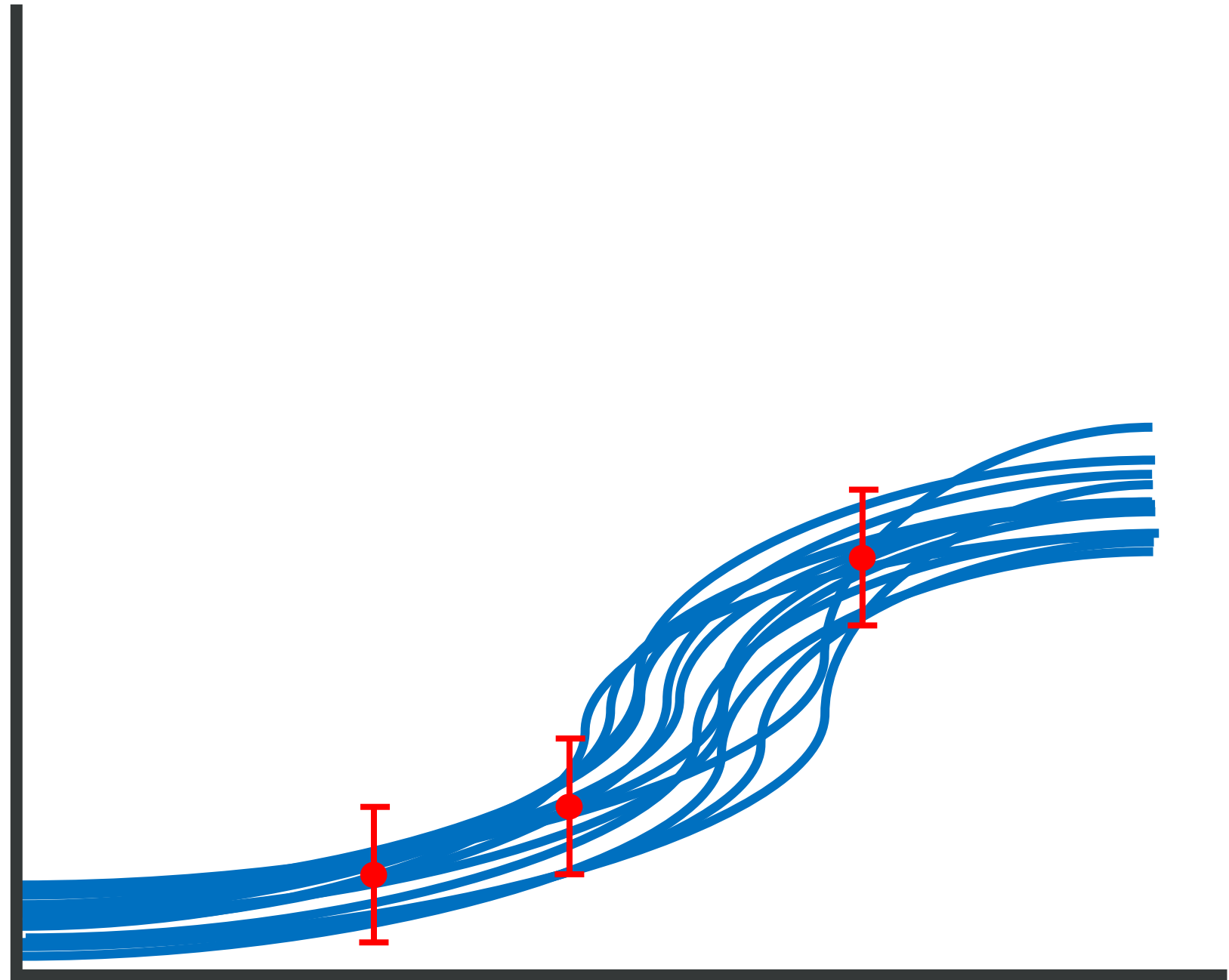
Machine Learning

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



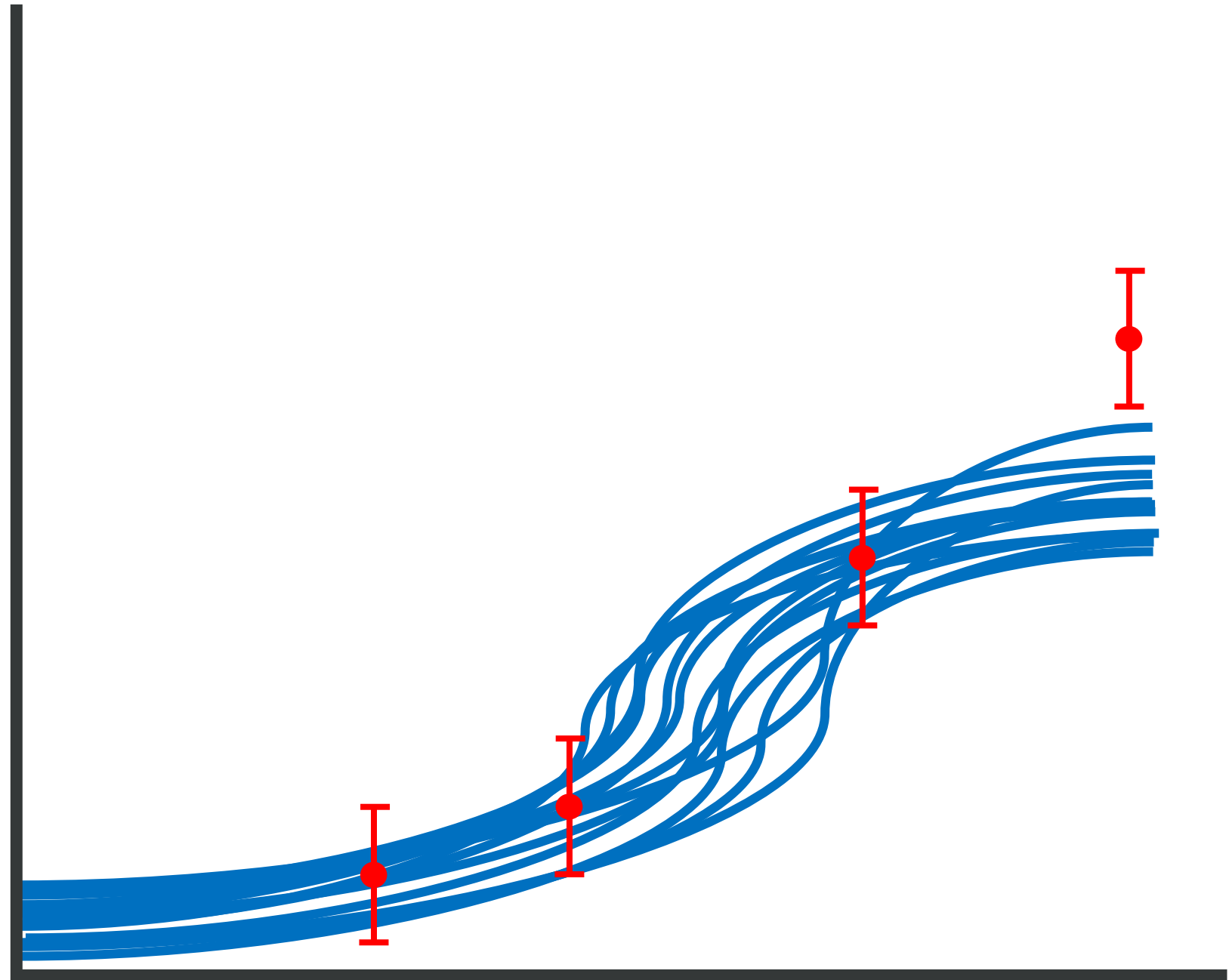
Machine Learning

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



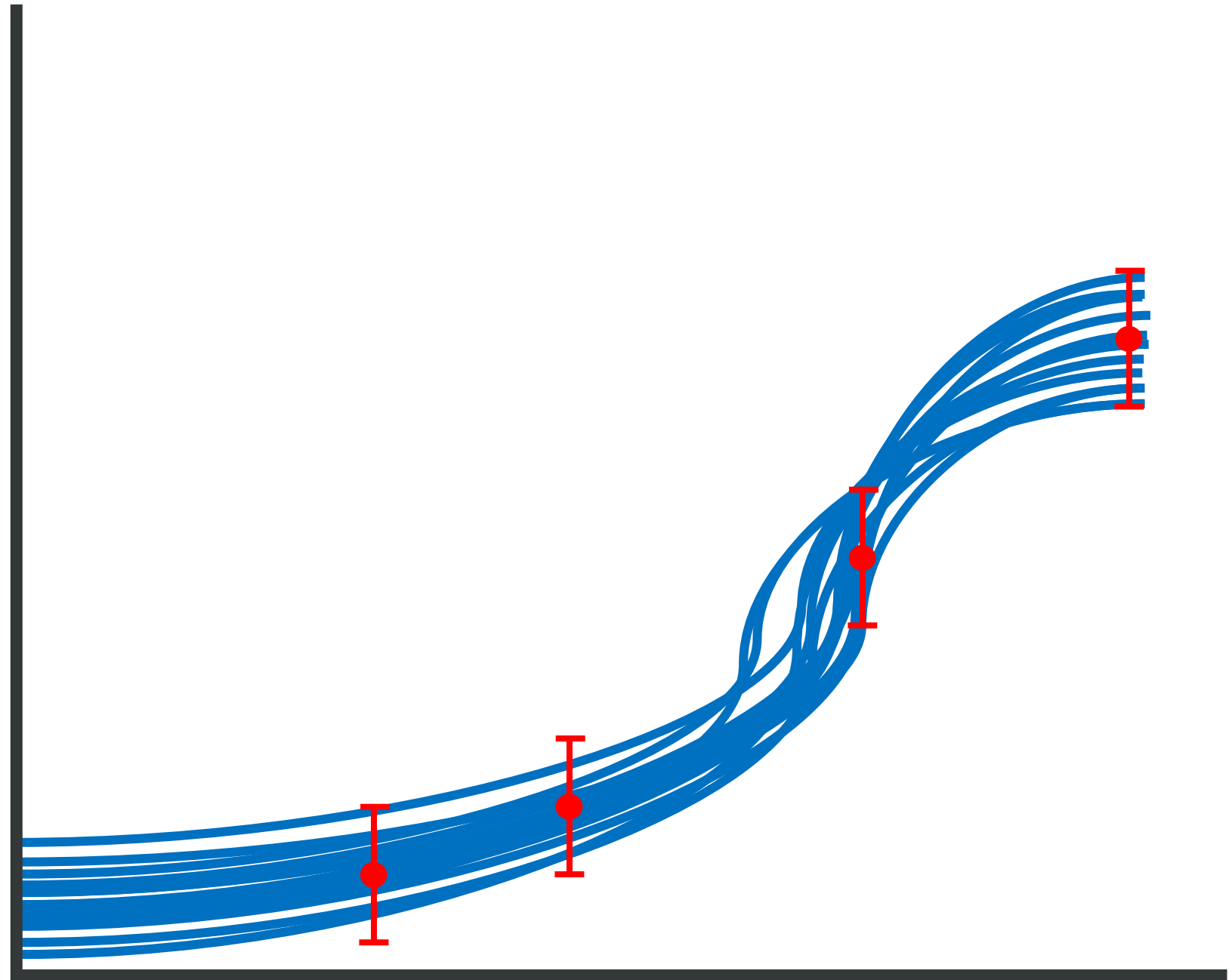
Machine Learning

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

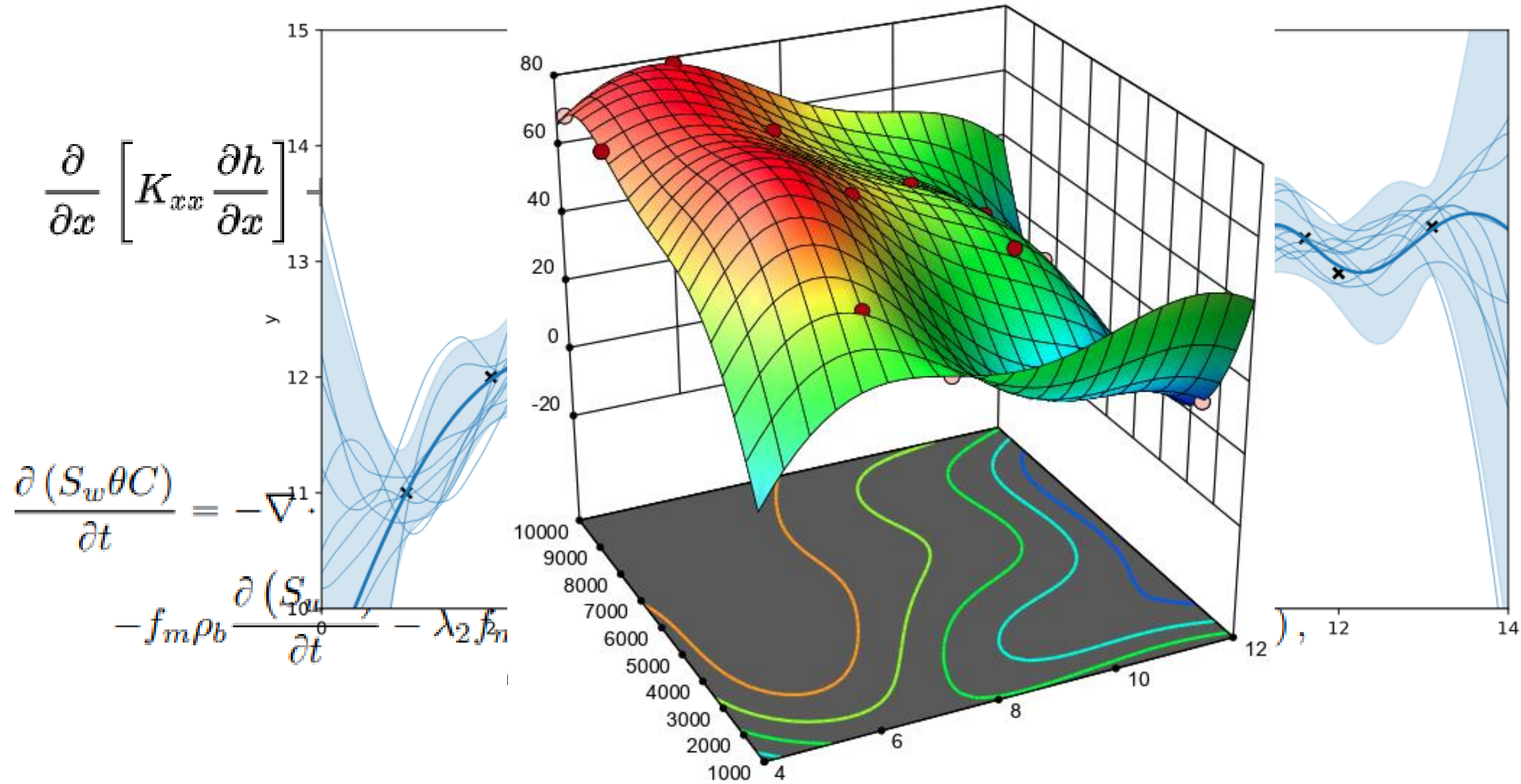


Machine Learning

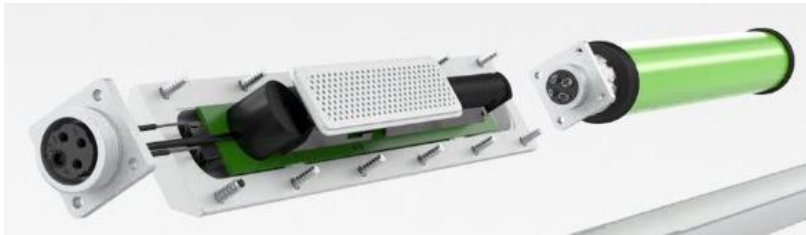
$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



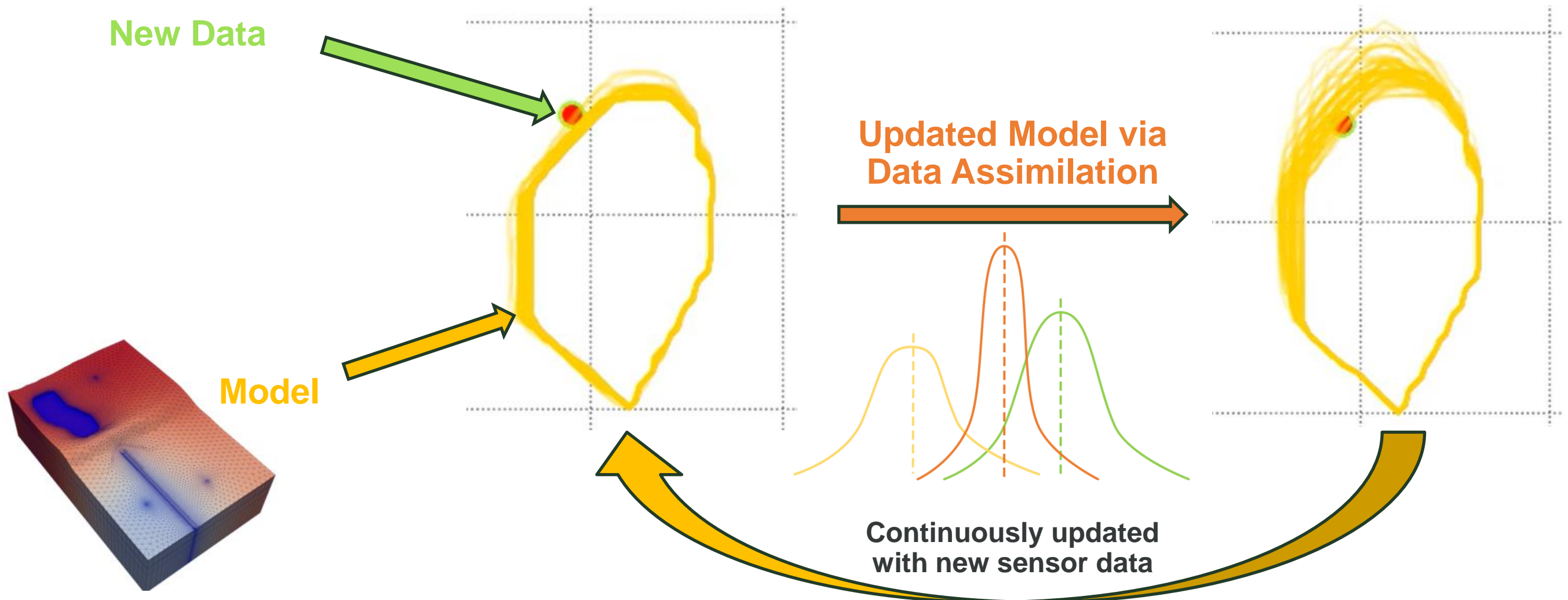
Data Assimilation



What is the value of PlumeFutures and continuous data?



Updated Views of Contamination Risk



Data Assimilation

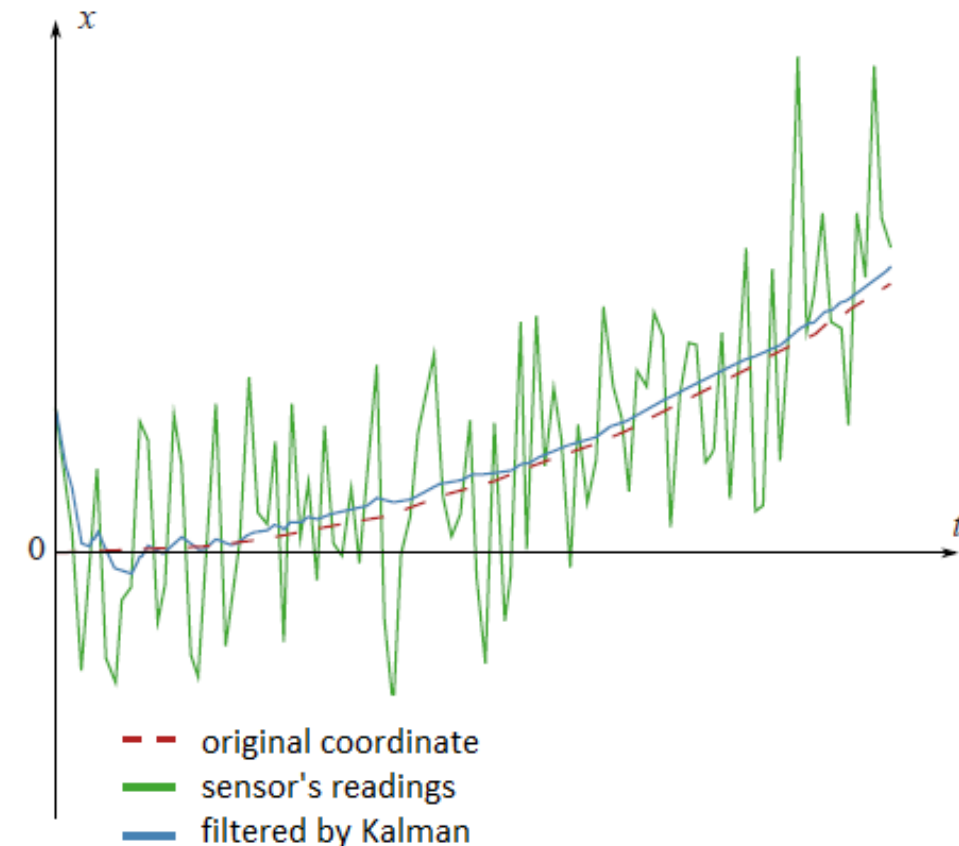
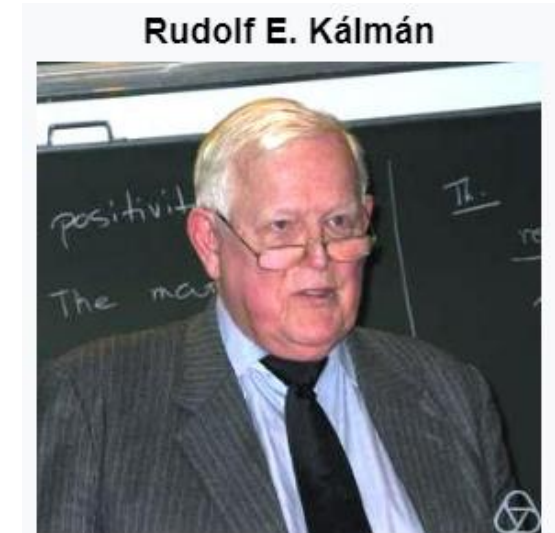


Kalman Filter

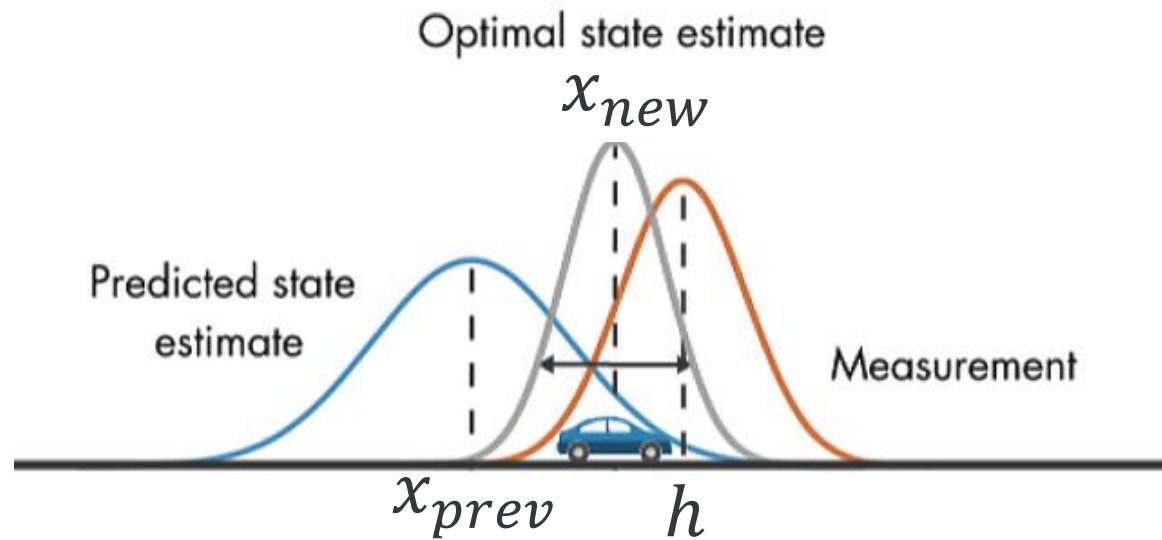
The optimal* way to combine noisy estimates to produce a better and less noisy estimate

Seemingly endless list of applications:

- Target tracking (e.g. missiles, kamikaze drones)
- Unmanned vehicles and robots (e.g. self-driving cars, drones, submarines etc.)
- Numerical Weather Prediction
- GPS devices (e.g. google maps)
- Signal Processing
- Econometrics
- Apollo program
- NASA Space Shuttle
- Navy submarines



Kalman Filter



$$K = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_h^2}$$

↑
“Kalman gain”

$$x_{new} = x_{prev} + Kr$$

measurement model

↓ ↓

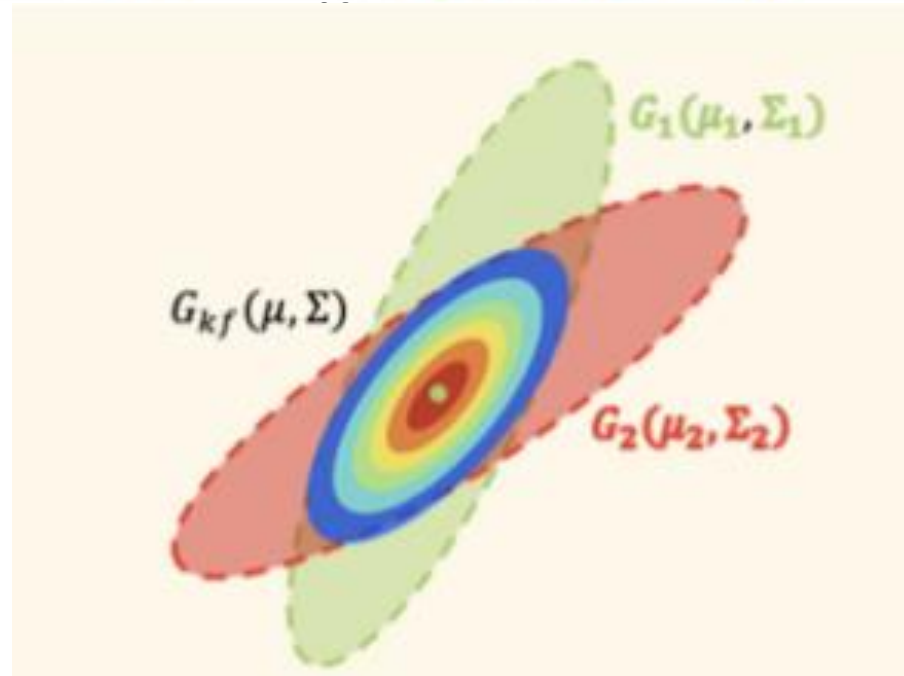
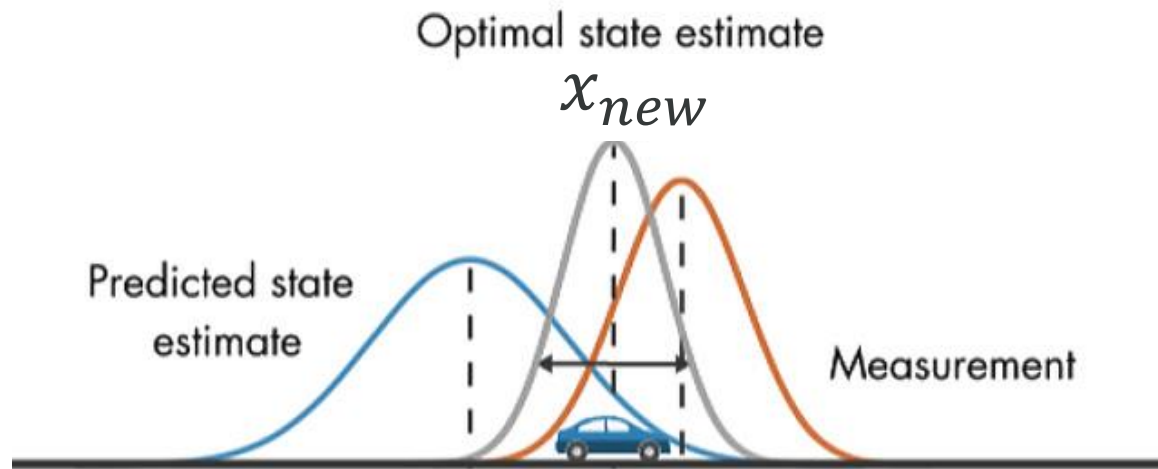
$$r = h - Zx$$

Size of update determined by the relative difference between the model uncertainty (σ_z^2) and the measurement uncertainty (σ_h^2).

If the model uncertainty is lower ($\sigma_z^2 < \sigma_h^2$) then we have a smaller update.

If the model uncertainty is higher ($\sigma_z^2 > \sigma_h^2$) then we have a larger update.

Kalman Filter



$$x_{new} = x_{prev} + Kr$$

$$r = h - Zx$$

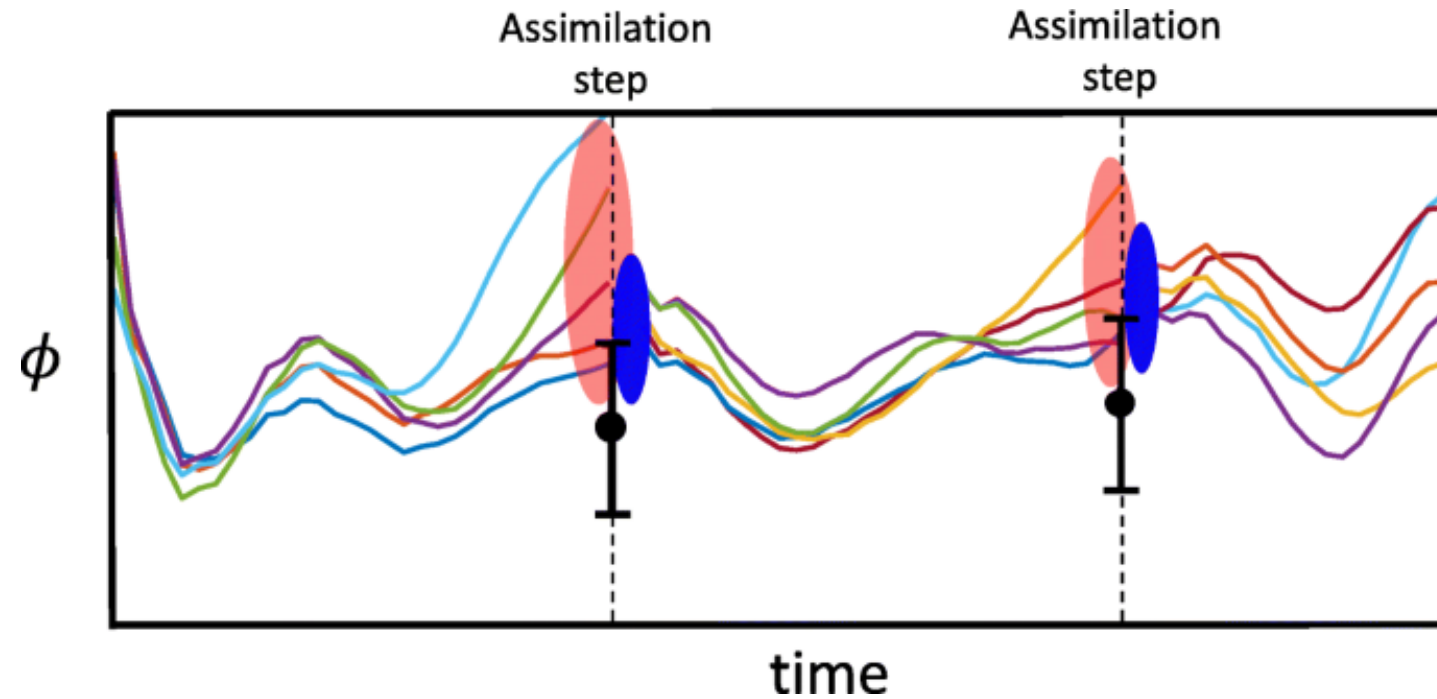
$$K = C(x)Z^T [ZC(x)Z^T + C(\varepsilon)]^{-1}$$

↑
"Kalman gain"

$$K = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_h^2}$$

$$K = \frac{C(x)Z^T}{[ZC(x)Z^T + C(\varepsilon)]}$$

Ensemble Kalman Filter



$$x_{new} = x_{prev} + Kr \quad r = h - Z(x)$$

$$K = C(x)Z^T[ZC(x)Z^T + C(\varepsilon)]^{-1}$$

Approximated with the statistics of the ensembles


Ensemble Kalman Filter

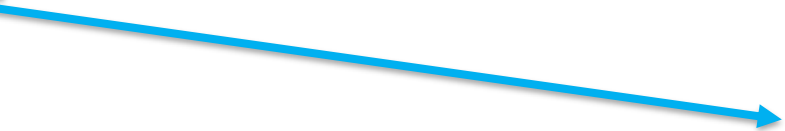
$$x_{new} = x_{prev} + Kr$$

$$r = h - Z(x)$$

$$r = h - o$$

$$K = \underbrace{C(x)Z^T}_{\text{green box}} [\underbrace{ZC(x)Z^T}_{\text{blue box}} + C(\varepsilon)]^{-1}$$


$$C(x)Z^T \approx \frac{1}{N-1} (x - \bar{x})(o - \bar{o})^T$$


$$ZC(x)Z^T \approx \frac{1}{N-1} (o - \bar{o})(o - \bar{o})^T$$

$$x_{new} = x_{prev} + C(x, o)[C(o) + C(\varepsilon)]^{-1}r$$

Iterative Ensemble Smoother

$$\begin{aligned}x_1 &= x_{prev} + \Delta x_1 \\x_2 &= x_1 + \Delta x_2 \\&\dots \\x_{new} &= x_i + \Delta x_i\end{aligned}$$

$$\begin{aligned}r &= h - Z(x) \\r &= h - o\end{aligned}$$


$$\Delta x_i = -((J_e^T C(\varepsilon)^{-1} J_e) + (1 + \lambda) C(x)^{-1})^{-1} (C(\varepsilon)^{-1} (x_i - x_{i-1}) + J_e^T r)$$

$$J_e = C(\varepsilon)^{\frac{1}{2}} \frac{\Delta_{sim}}{\Delta_{par}} C(\varepsilon)^{-\frac{1}{2}}$$



$$\Delta_{par} \approx \frac{1}{N-1} (x - \bar{x})(o - \bar{o})^T$$

$$\Delta_{sim} \approx \frac{1}{N-1} (o - \bar{o})(o - \bar{o})^T$$

Data Assimilation Resources



[About](#) [Engagement](#) [Worked Examples](#) [Research](#) [Education](#) [Software](#) [Resources](#)



About GMDSI

Groundwater Modelling Decision Support Initiative (GMDSI) is an industry-funded and aligned project focused on improving the role that groundwater modelling plays in environmental management and decision making.

[ABOUT](#)

Lead collaborating organisations



pyEMU



PlumeFutures

Machine learning AI has enabled:

- Model prep time reduced from **weeks to 1-3 hours**.
- Model updating fully **automated**.
- Model predictions targeted at **what clients care about (risk and liability)**, not what ~~needs~~ modelers care about (parameters).
- **Revolutionizing scale**: from manual modeling of select sites to analyzing **thousands of locations**, processing **millions of data points**.

Case Study: Deliberate Delineation

Contours of the groundwater Benzene threshold from 200 models that equally fit the available data.

Identified potential plume instability.

AI informed borehole locations to delineate risk.





ems

Environmental
Material Science

Site insights for a
better world.

Nico Higgs
nicohiggs@ems-inc.ca

Steven Mamet
stevemamet@ems-inc.ca

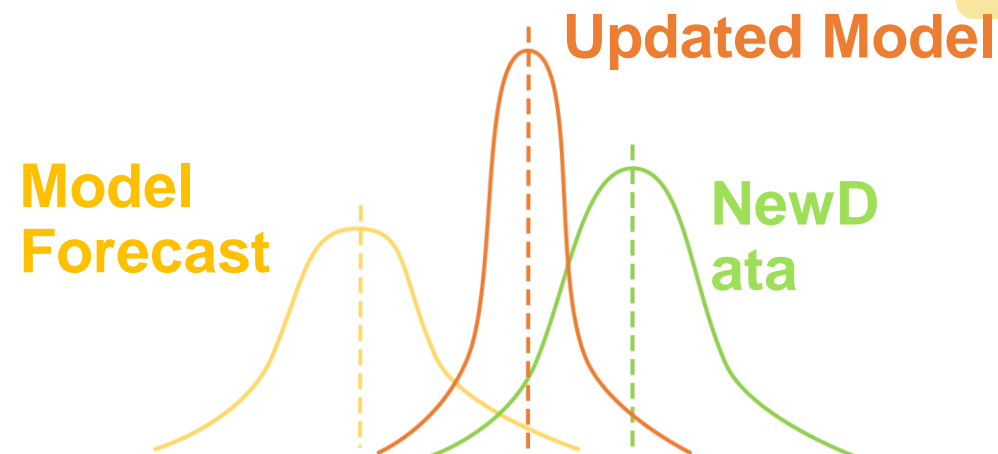
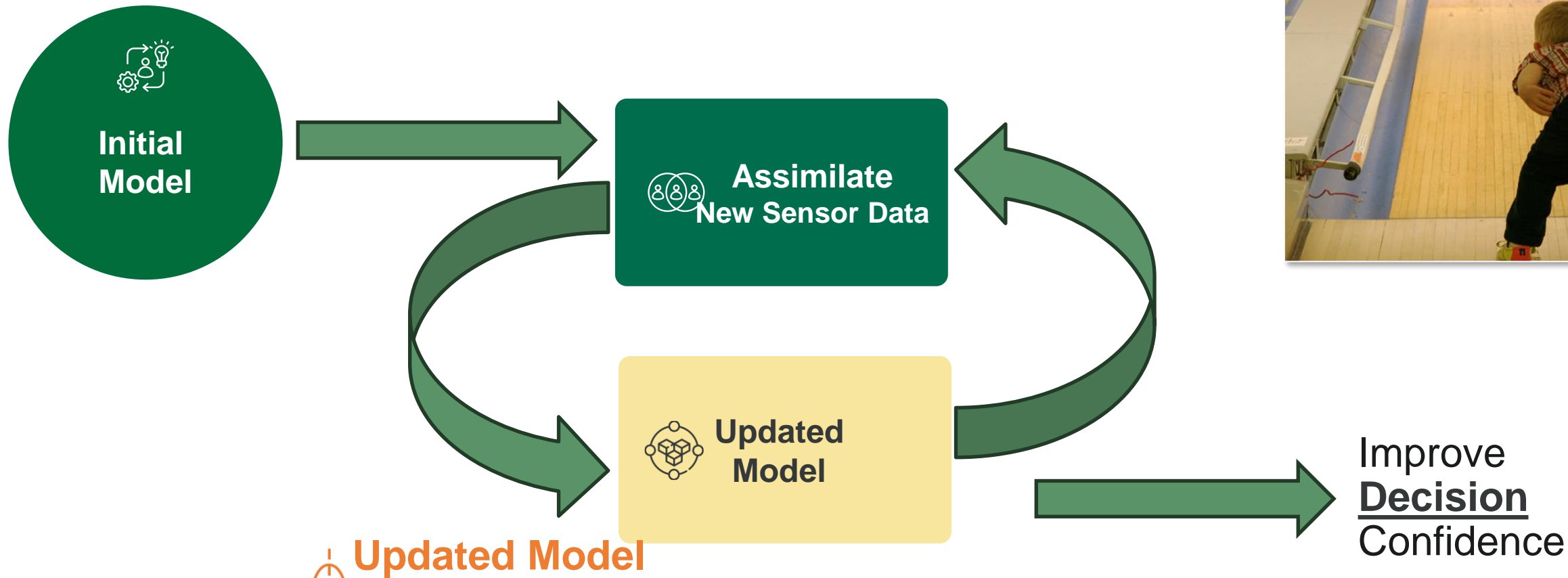
Steven Siciliano
stevesiciliano@ems-inc.ca

www.ems-inc.ca



Why Autonomous Sensors?

Continuously Monitor and Update Decision Critical Predictions



Risk/Liability continuously updated:

- ✓ Plume Stability
- ✓ Natural depletion (NSZD)
- ✓ Is applied remediation effective?

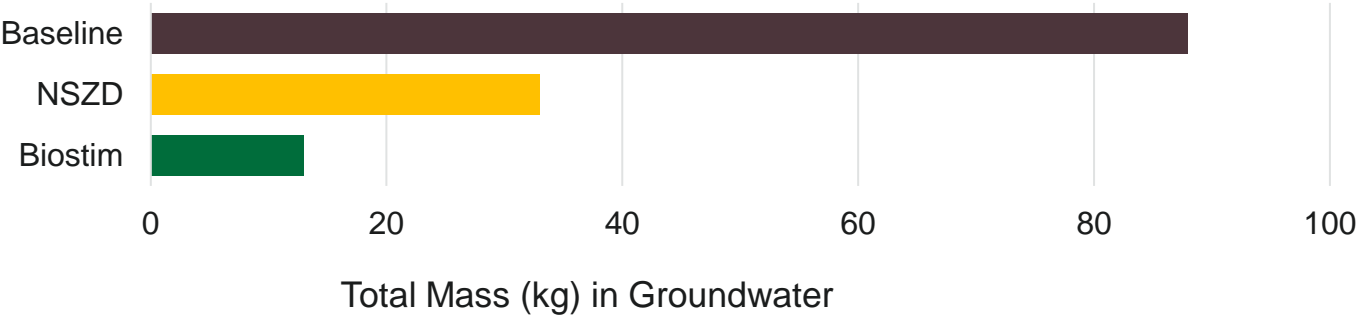
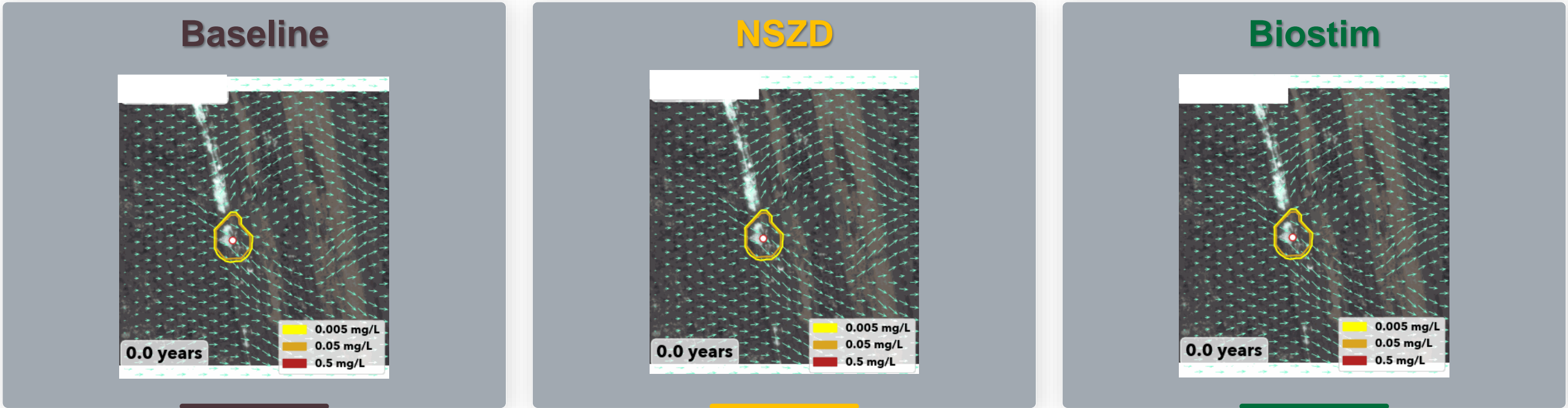
Why EMS?

Increased Data at Reduced Cost

| | Traditional Annual Sampling | EMS Autonomous Sensor Direct Measurement |
|---|---|--|
| # Monitoring Wells | 6 | 6 |
| Sample Frequency | Twice a year or Annually (depending on Regulations) | Every 30 minutes |
| Field Work | Pull sample, send to lab, wait for report | None after sensor install into existing monitoring wells/easy removal |
| Analysis | Single point in time sample 10 mL of 250 mL / 1L water One report | Continuously updated, online visual trends predict risks, fast decisions, assess more sites |
| Data Points per well each sampling event (e.g., annual) | 1-2 | 17,520 |
| Time to Risk Understanding | ~ 3 Years | ~ 3 to 6 months |
| Cost | \$30,000 - \$60,000 | \$14,500 or \$29,000 / year |
| Cost/data point | \$5,000.00 - \$10,000.00 each | \$0.80 - \$1.60 each |

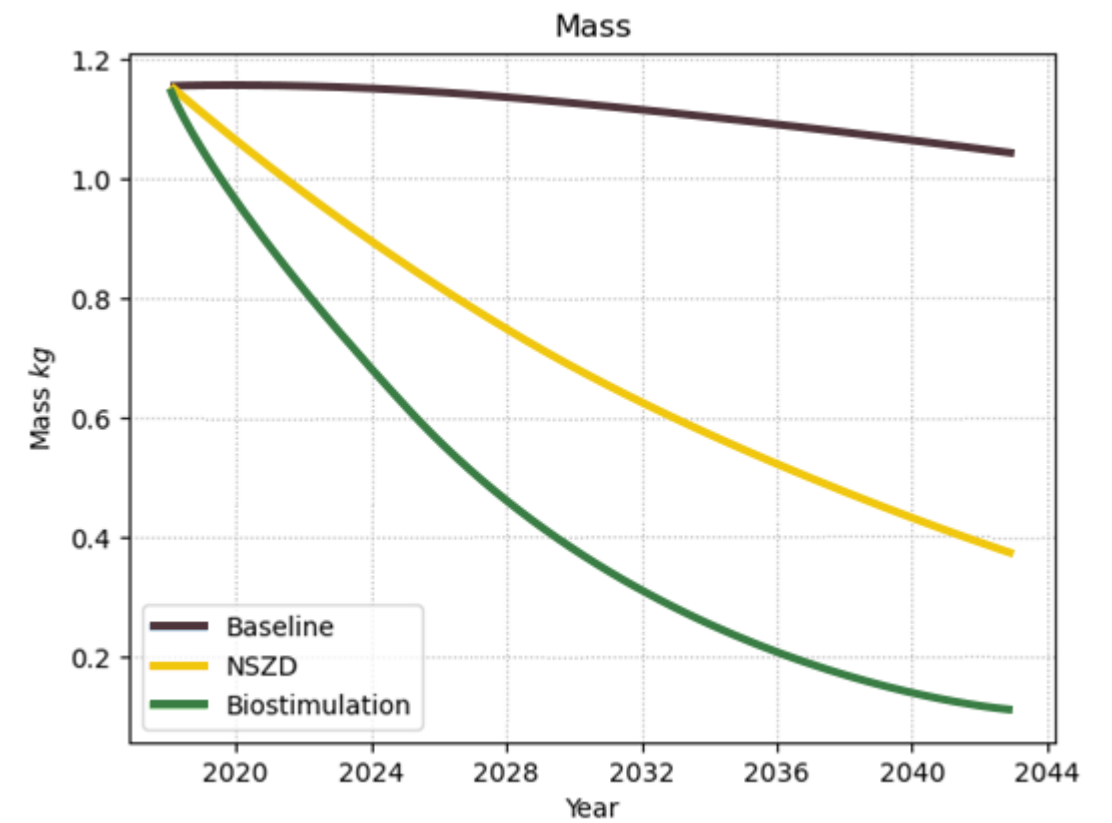
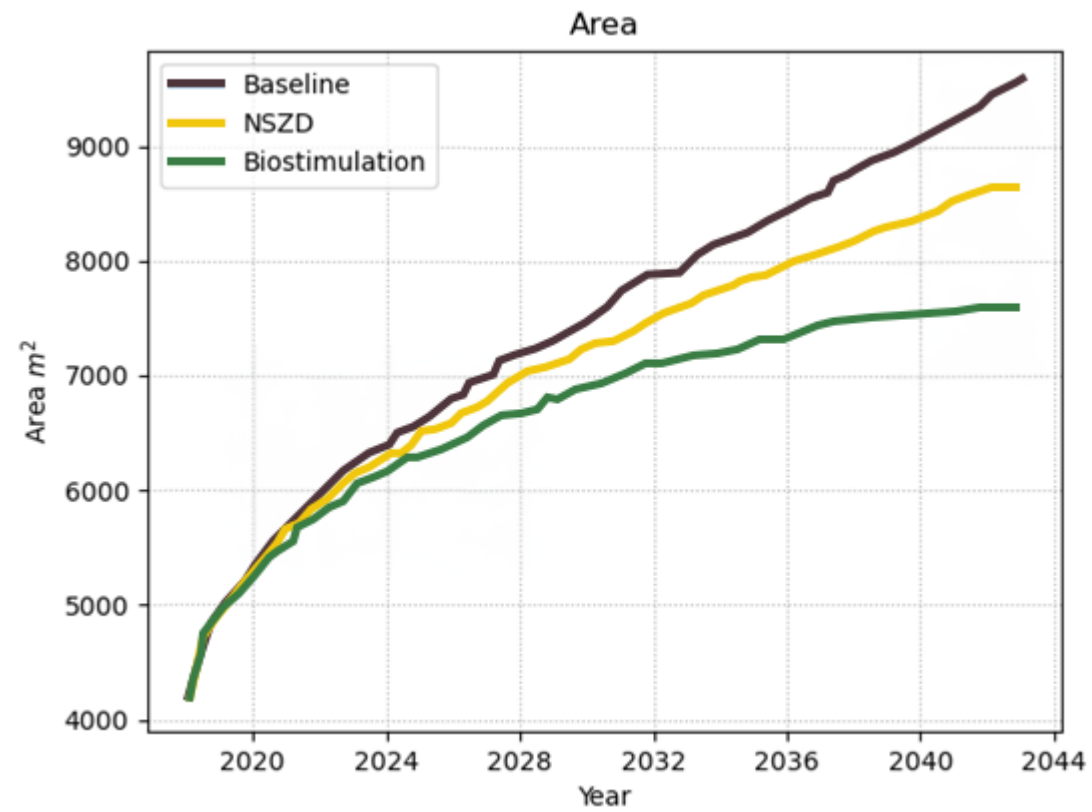
Why PlumeFutures?

Scenario Planning



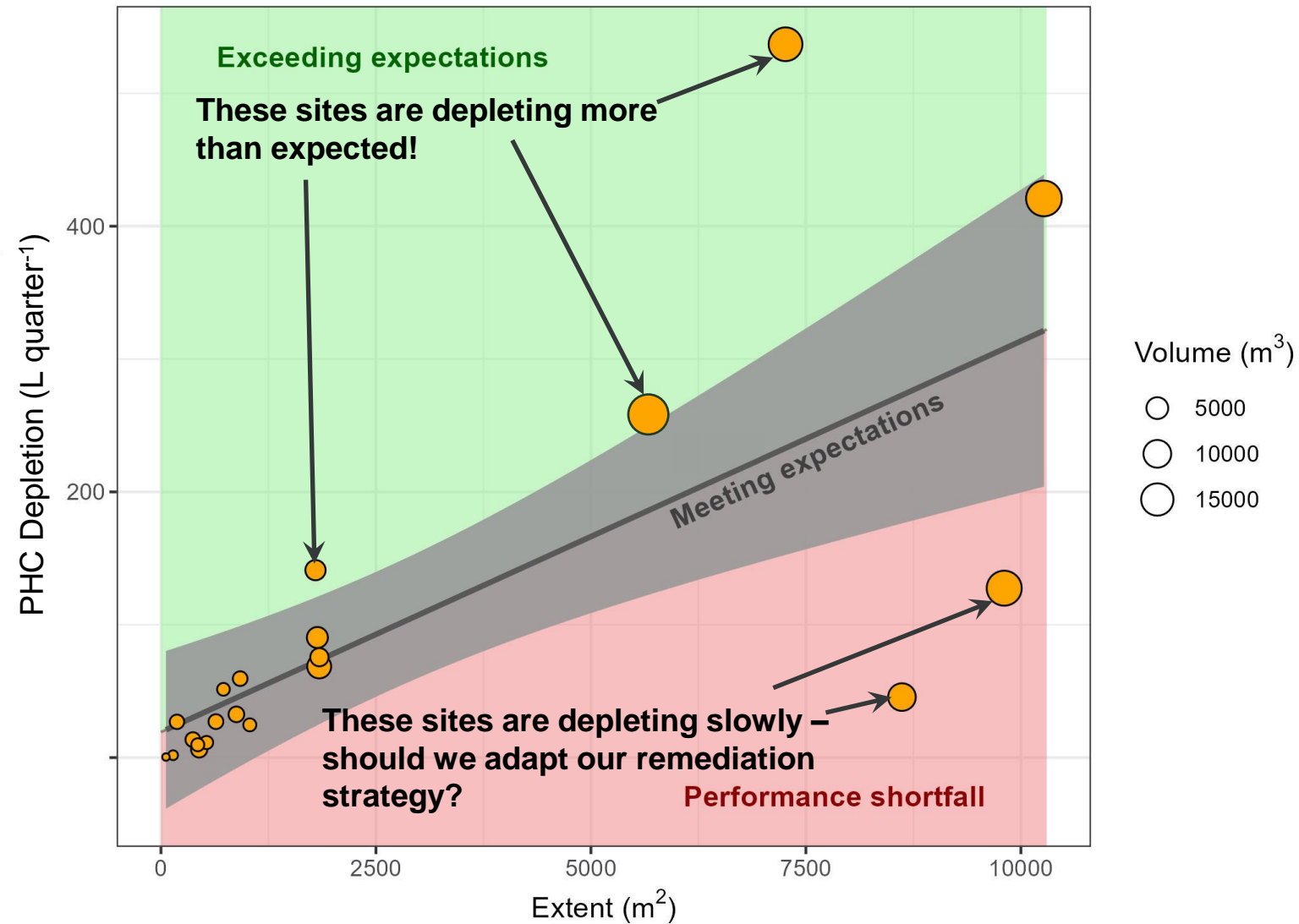
Why PlumeFutures?

Scenario Planning



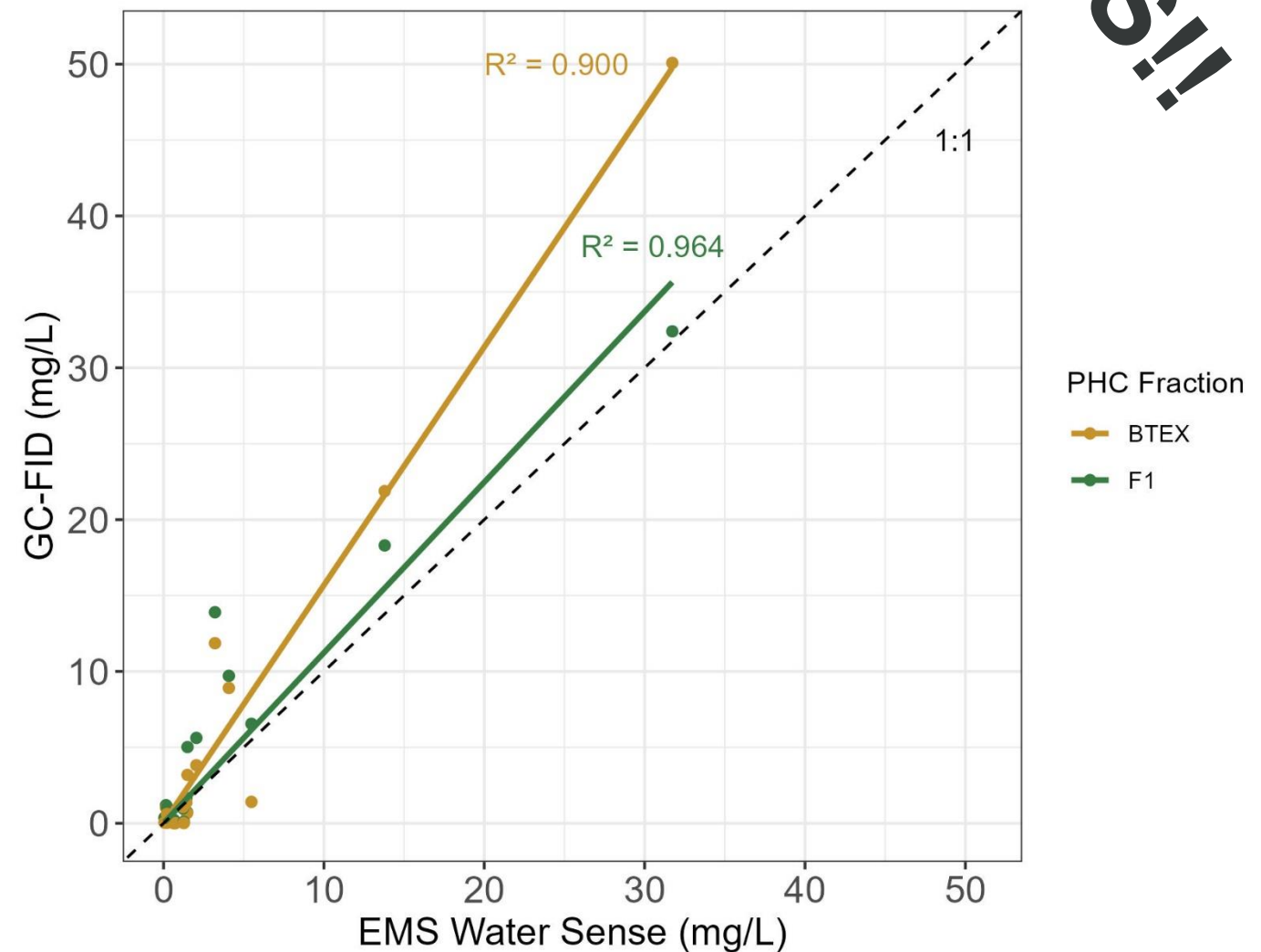
Why PlumeFutures?

Site Triage





But can WaterSense really estimate GW PHC concentrations?



YES!!



Specifications

| Component | Range | Resolution | Accuracy (+/-) | Precision | Drift (yr ⁻¹) |
|-------------------------------|-------------|------------|----------------|-----------|---------------------------|
| O2 (%) ¹ | 0 – 25 | 0.01 | 0.6 | 0.3 | <5 |
| CO2 (%) | 0 – 10 | 0.002 | 0.002 | 0.03 | 0.05 |
| CH4 (%) | 0 – 50 | 0.0004 | 0.006 | 0.17 | 0.05 |
| PHC (%) | 0 – 2.1 | 0.001 | 0.003 | 0.0007 | 0.025 |
| Temperature (°C) ² | -40 – 85 | 0.01 | 1.0 | 0.1 | 0.02 |
| RH (%) ² | 0 – 100 | 0.008 | 3.0 | 0.02 | 0.5 |
| Pressure (mb) ² | 300 – 1,100 | 0.0018 | 1.0 | 0.013 | 1.0 |

¹SGX Industrial Oxygen [Sensor](#) (SGX-4Ox-EL)

²BOSCH Combined Temperature, Humidity, and Pressure [Sensor](#) (BME280)

Sensor performance is reported according to the following definitions:

- **Range** – Reported in absolute units as the minimum and maximum values of the range the sensor detects.
- **Accuracy*** – Reported in absolute units as the Root Mean Squared Error (RMSE).
- **Precision*** – Reported in absolute units as the standard deviation.
- **Resolution** – Reported in absolute units as the smallest detectable change.
- **Drift** – Reported as the difference between RMSE at 1 year after calibration and RMSE at calibration.

PlumeFutures

MODFLOW is a widely used groundwater modeling software developed by the United States Geological Survey (USGS). Here's what it does:

1. **Groundwater Flow Simulation:** MODFLOW simulates the flow of groundwater through aquifers. It can model how water moves underground, which is essential for understanding and managing water resources.
2. **Grid-Based Model:** The area being studied is divided into a grid of cells. Each cell represents a specific volume of the aquifer. The model calculates the water flow between these cells over time.
3. **Input Parameters:** Users provide various inputs like hydraulic conductivity, recharge rates (**HYDRUS**), pumping rates, and boundary conditions. These parameters define how water moves through the aquifer.
4. **Applications:** MODFLOW is used for water supply planning, contamination assessment, and managing groundwater resources. It helps in predicting how groundwater levels change in response to natural and human activities.

PEST++ is an advanced version of the PEST (Parameter ESTimation) software, which is used for model calibration and uncertainty analysis. Here's what it does:

1. **Model Calibration:** PEST++ adjusts the input parameters of a model (like MODFLOW) to match observed data (e.g., water levels, flow rates). This process ensures that the model accurately represents the real-world system.
2. **Optimization:** PEST++ uses optimization algorithms to find the best set of parameters that minimize the difference between observed data and model predictions.
3. **Uncertainty Analysis:** It also quantifies the uncertainty in model predictions. This means it helps in understanding how certain or uncertain the model outputs are, given the possible variations in input parameters.
4. **Automation and Efficiency:** PEST++ is designed to be more efficient and capable of handling large and complex models compared to the original PEST. It can automate many of the calibration tasks, making it faster and more reliable.

How They Work Together

When used together, MODFLOW and PEST++ provide a powerful toolkit for groundwater modeling:

- **MODFLOW** creates a detailed simulation of groundwater flow.
- **PEST++** calibrates the MODFLOW model to ensure it accurately reflects observed data and assesses the uncertainty in the model predictions.

This combination is essential for making informed decisions about groundwater management and ensuring that models are reliable and useful for planning and analysis.

PlumeFutures

Workflow Example

1. **Initial Setup:** Develop HYDRUS and MODFLOW models for your study area. Define the soil properties, boundary conditions, and initial conditions for HYDRUS. Define the aquifer properties and boundary conditions for MODFLOW.
2. **HYDRUS Simulation:** Run HYDRUS to simulate soil moisture dynamics and calculate recharge rates based on precipitation, evapotranspiration, and soil properties.
3. **MODFLOW Simulation:** Use the recharge rates from HYDRUS as input for MODFLOW. Run MODFLOW to simulate groundwater flow and levels.
4. **Calibration with PEST++:** Use observed data (e.g., soil moisture content, groundwater levels) to calibrate both models. PEST++ will adjust the parameters in HYDRUS and MODFLOW to minimize the differences between observed and simulated values.
5. **Iterate and Refine:** Iterate the process, running HYDRUS and MODFLOW simulations, and recalibrating with PEST++ until the models are well-calibrated and accurately reflect the observed data.
6. **Uncertainty Analysis:** Use PEST++ to perform uncertainty analysis, providing insights into the reliability of the model predictions.

By integrating HYDRUS with MODFLOW and using PEST++ for calibration, you can create a robust and accurate modeling framework that covers the entire hydrological cycle from the unsaturated zone to the groundwater system. This integrated approach is valuable for various applications, including water resource management, agricultural planning, and contaminant transport studies.

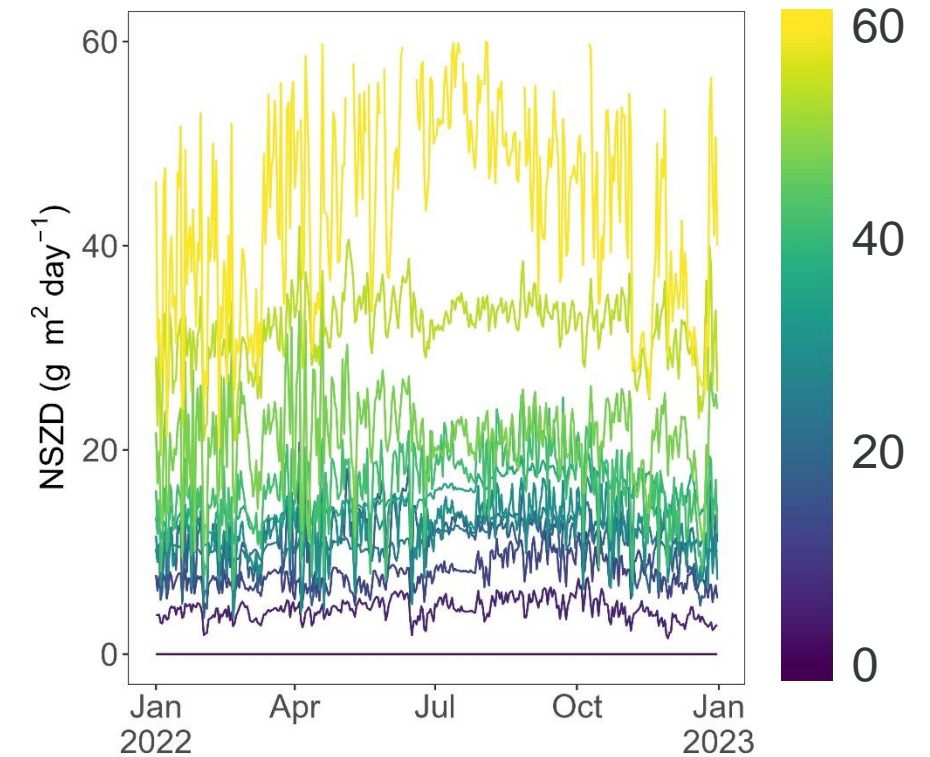
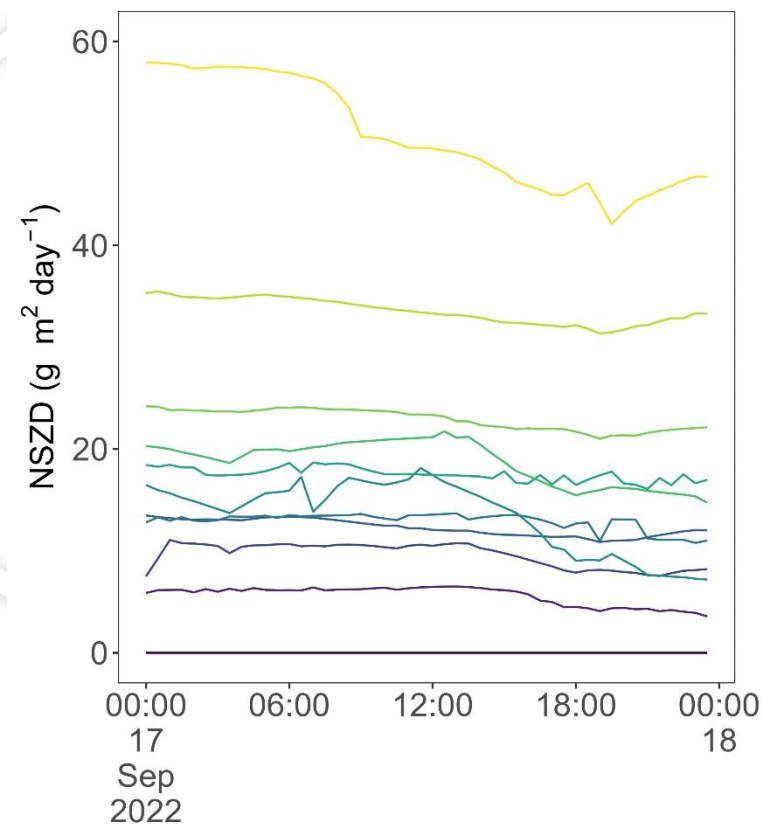
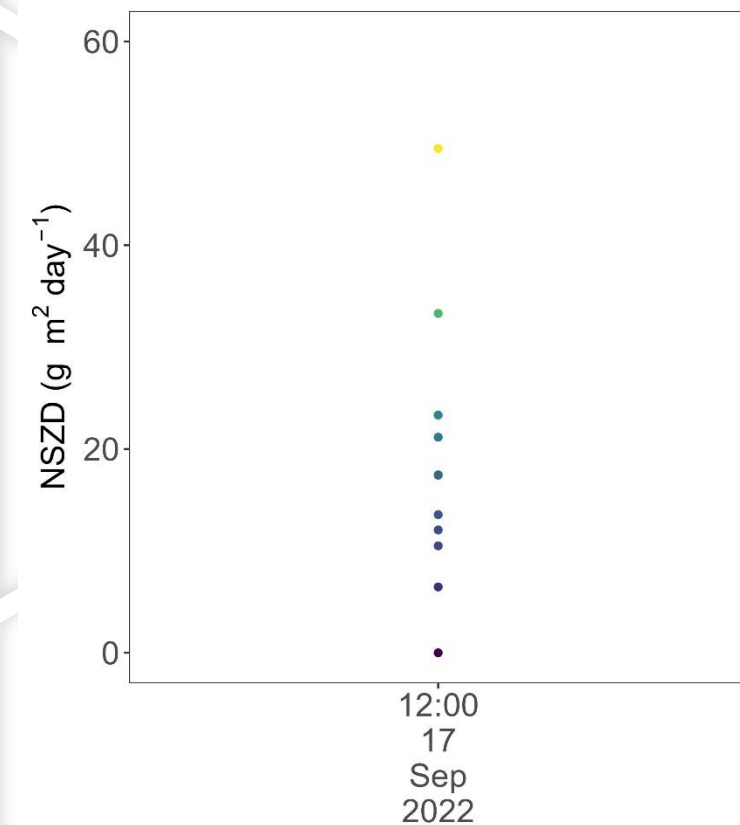
The Challenge of Point-in-time Sampling to Assess Risks

Traditional: various samples at a site

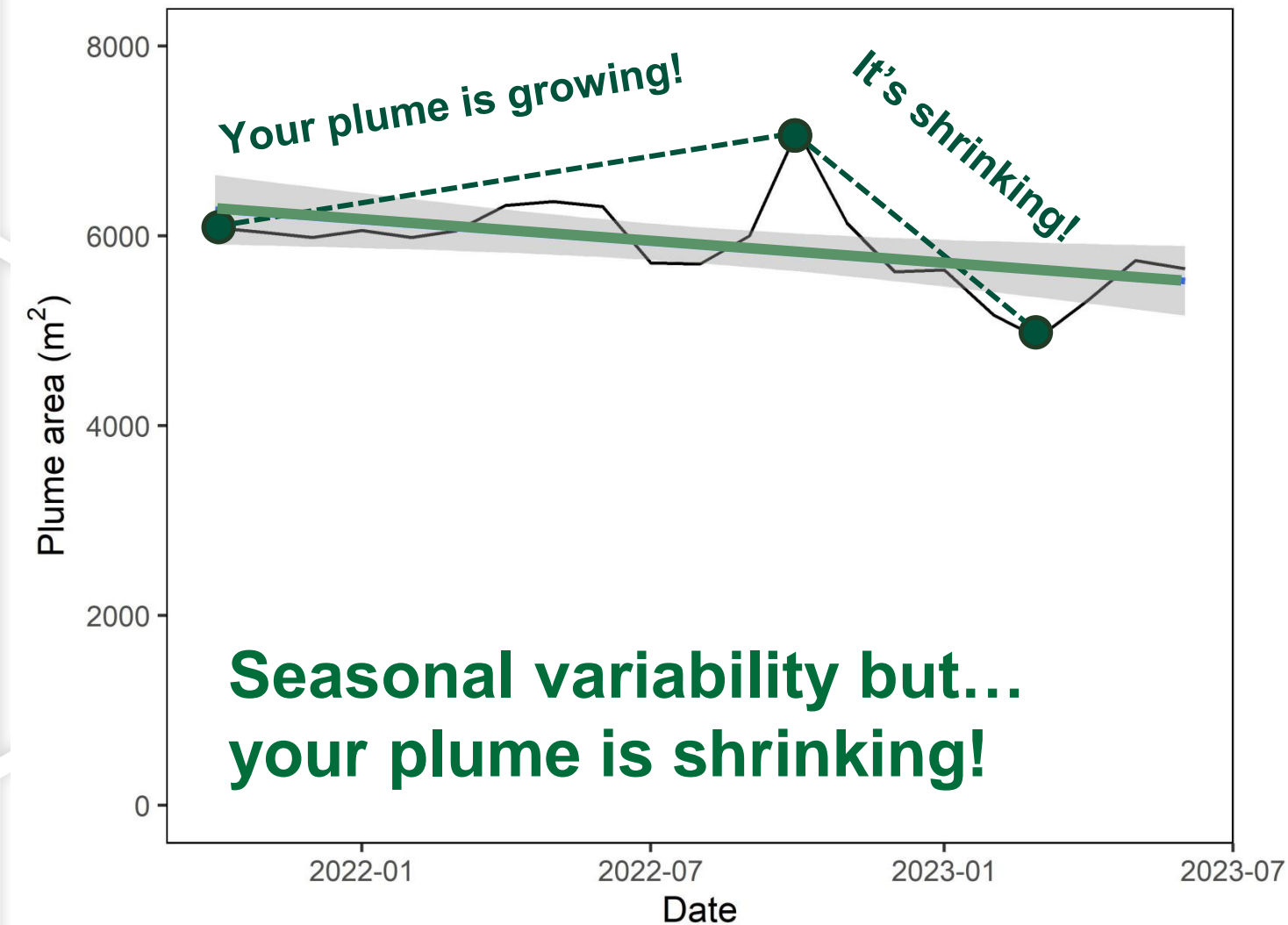
48 measurements show sample variability over 1 day

17,520 measurements show variability over 1 year!

NSZD
($\text{g m}^{-2} \text{ day}^{-1}$)



The Challenge of Point-in-time Sampling to Assess Risks



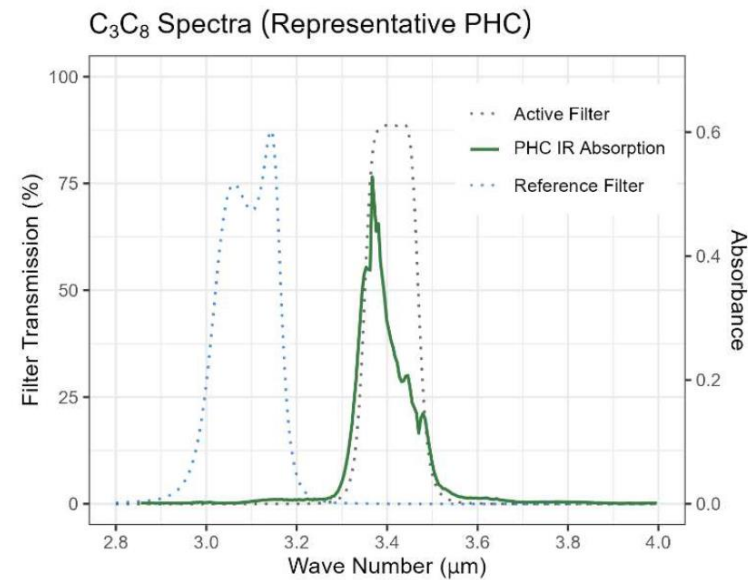
- Typically, 5-10 sensors/site
- 30 min resolution.

— Plume area through time
 — Trend line

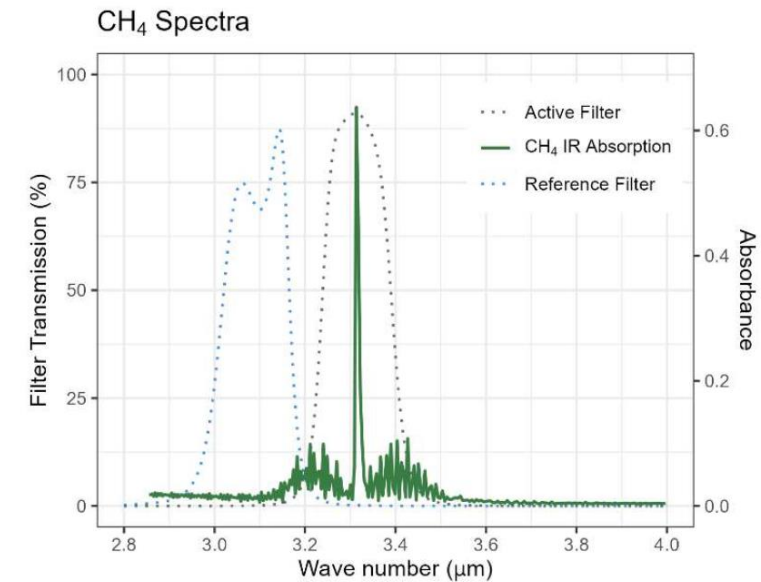
NDIR

Gas Absorption: Different gases absorb infrared light at specific wavelengths. For PHCs, CH_4 , and CO_2 , this absorption can be detected and measured. Here are the details about the 4-channel NDIR detector:

- Channel 1 (reference): $3.1\ \mu\text{m}$
- Channel 2 (CH_4): $3.3\ \mu\text{m}$
- Channel 3 (PHC): $3.4\ \mu\text{m}$
- Channel 4 (CO_2): $4.2\ \mu\text{m}$



EMS 4-channel NDIR sensor measure IR absorbances from 3.2 to $3.5\ \mu\text{m}$ (C-H stretching) and 4.2 - $4.3\ \mu\text{m}$ (O=C=O stretching). Above highlights how PHCs (propane (C_3C_8)) is used as a representative PHC) are determined relative to a reference channel.



Above highlights how CH_4 is determined relative to a reference channel.

Beer-Lambert

At its core, the Beer-Lambert Law postulates that light absorbance (A) is directly proportional to the concentration (C) of the sample:

$$A = -\ln\left(\frac{I_0}{I}\right) = \epsilon l C \quad (1)$$

Where:

I_0 = incident intensity (dimensionless)

I = transmitted intensity (dimensionless)

l = length of the solution the light passes through (distance between IR emitter and detector; cm)

ϵ = molar absorption coefficient ($\text{M}^{-1}\text{cm}^{-1}$)

Practical considerations in the NDIR implementation require modifications to the Beer-Lambert Law, as follows, to obtain accurate readings:

$$A = SPAN(1 - e^{-bc^c}) \quad (2)$$

Where:

$SPAN$ = scaling factor that accounts for non-idealities in the measurement system (dimensionless)

b = slope correction factor non-linear increases in absorption with concentration (dimensionless)

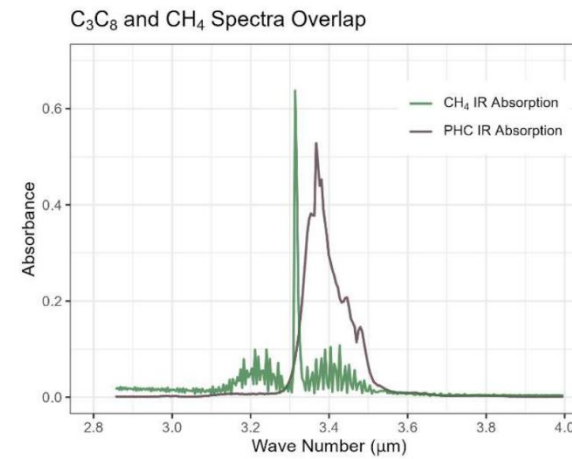
C = concentration (as above)

c = molar absorption coefficient ($\text{M}^{-1}\text{cm}^{-1}$)

The $SPAN$ factor is introduced because not all the IR radiation that impinges upon the active detector is absorbed by the gas, even at high concentrations.

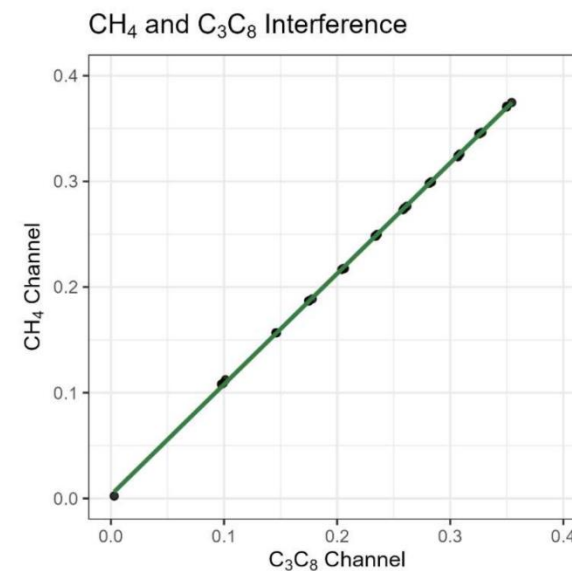
CH₄ and PHC Interference

There is overlap between IR CH₄ and other low molecular weight PHC absorption frequencies (e.g., C₃H₈):



Note the overlap between CH₄ and PHC spectra.

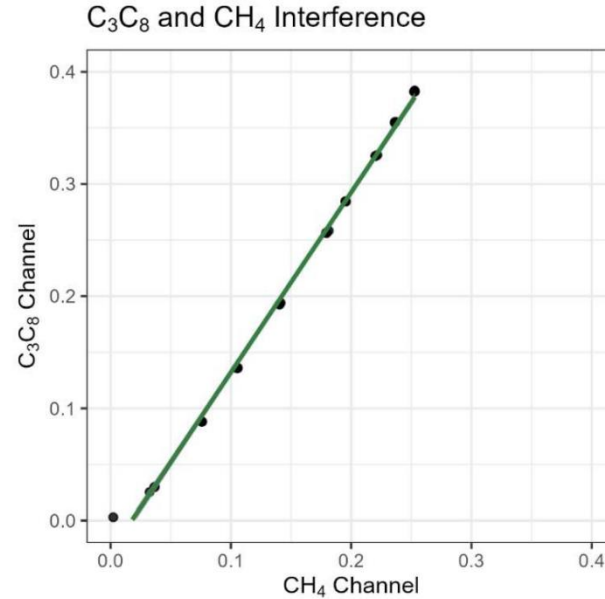
So PHC will be detected on the CH₄ channel:



Above highlights how PHC registers on the CH₄ channel.

CH₄ and PHC Interference

And CH₄ will be detected on the PHC channel:



Above highlights how CH₄ registers on the PHC channel.

To account for this, EMS uses absorption cross-interference algorithms⁷ to determine PHC and methane concentrations in mixtures.

$$NA_{CH_4}^i = \frac{NA_{CH_4} - NA_{C_3H_8}K_{C_3H_8}}{1 - K_{CH_4}K_{C_3H_8}} \quad (8)$$

Where:

$NA_{CH_4}^i$ = interference corrected CH₄ normalized absorbance (dimensionless)

NA = uncorrected normalized absorbance for the corresponding gas (dimensionless)

K = interference correction constant for the corresponding gas (dimensionless)

Water Sense

Henry's Law:

The amount of dissolved gas in a liquid is proportional to its partial pressure above the liquid.

Raoult's Law:

The partial pressure of each component of an ideal mixture of liquids is equal to the vapour pressure of the pure component multiplied by its mole fraction in the mixture.

Water Sense

$$C_a = C_g P K_H^0 \exp \left\{ \delta \left[\frac{1}{T} - \frac{1}{T^0} \right] \right\} M_B \rho_w$$

C_a = Aqueous concentration

C_g = Gaseous concentration

P = Pressure

K_H^0 = Henry's constant for solubility in water

δ = Temperature dependence constant

T = Measured temperature

T^0 = Standard temperature

M_B = Benzene molar mass

ρ_w = Water density

Water Sense

$$C_a = C_g P K_H^0 \exp \left\{ \delta \left[\frac{1}{T} - \frac{1}{T^0} \right] \right\} M_B \rho_w$$

C_a = Aqueous concentration



What we want

C_g = Gaseous concentration

P = Pressure

T = Measured temperature



What we measure

K_H^0 = Henry's constant for solubility in water

δ = Temperature dependence constant

T^0 = Standard temperature

M_B = Benzene molar mass

ρ_w = Water density



Constants