



Technische
Universität
Braunschweig

Master's Thesis

Constraint Optimization for Reservoir Learning of Multivariate Time Series

Yannic Lieder, April 13, 2021



Technische
Universität
Braunschweig

Master's Thesis

Constraint Optimization for Reservoir Learning
of Multivariate Time Series

Yannic Lieder, April 13, 2021

Outline

1. Constraint Definition
2. Embedding Constraints into the Neural Network
3. Example: Forecasting of Satellite Images
4. Future Work & Conclusion

Outline

1. Constraint Definition
2. Embedding Constraints into the Neural Network
3. Example: Forecasting of Satellite Images
4. Future Work & Conclusion

Context: Neumann, Rolf and Steil (2012)

RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS INTO EXTREME LEARNING MACHINES

KLAUS NEUMANN*

*Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
kneumann@cor-lab.de*

MATTHIAS ROLF

*Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
mrolf@cor-lab.de*

JOCHEM JAKOB STEIL

*Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
jsteil@cor-lab.de*

The application of machine learning methods in the engineering of intelligent technical systems often requires the integration of continuous constraints like positivity, monotonicity, or bounded curvature in the learned function to guarantee a reliable performance. We show that the extreme learning machine is particularly well suited for this task. Constraints involving arbitrary derivatives of the learned function are effectively implemented through quadratic optimization because the learned function is linear in its parameters, and derivatives can be derived analytically. We further provide a constructive approach to verify that discretely sampled constraints are generalized to continuous regions and show how local violations of the constraint can be rectified by iterative re-learning. We demonstrate the approach on a practical and challenging control problem from robotics, illustrating also how the proposed method enables learning from few data samples if additional prior knowledge about the problem is available.

Keywords: extreme learning machine, neural network, prior knowledge, continuous constraints, regression.

Context: Neumann, Rolf and Steil (2012)

RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS
INTO EXTREME LEARNING MACHINES

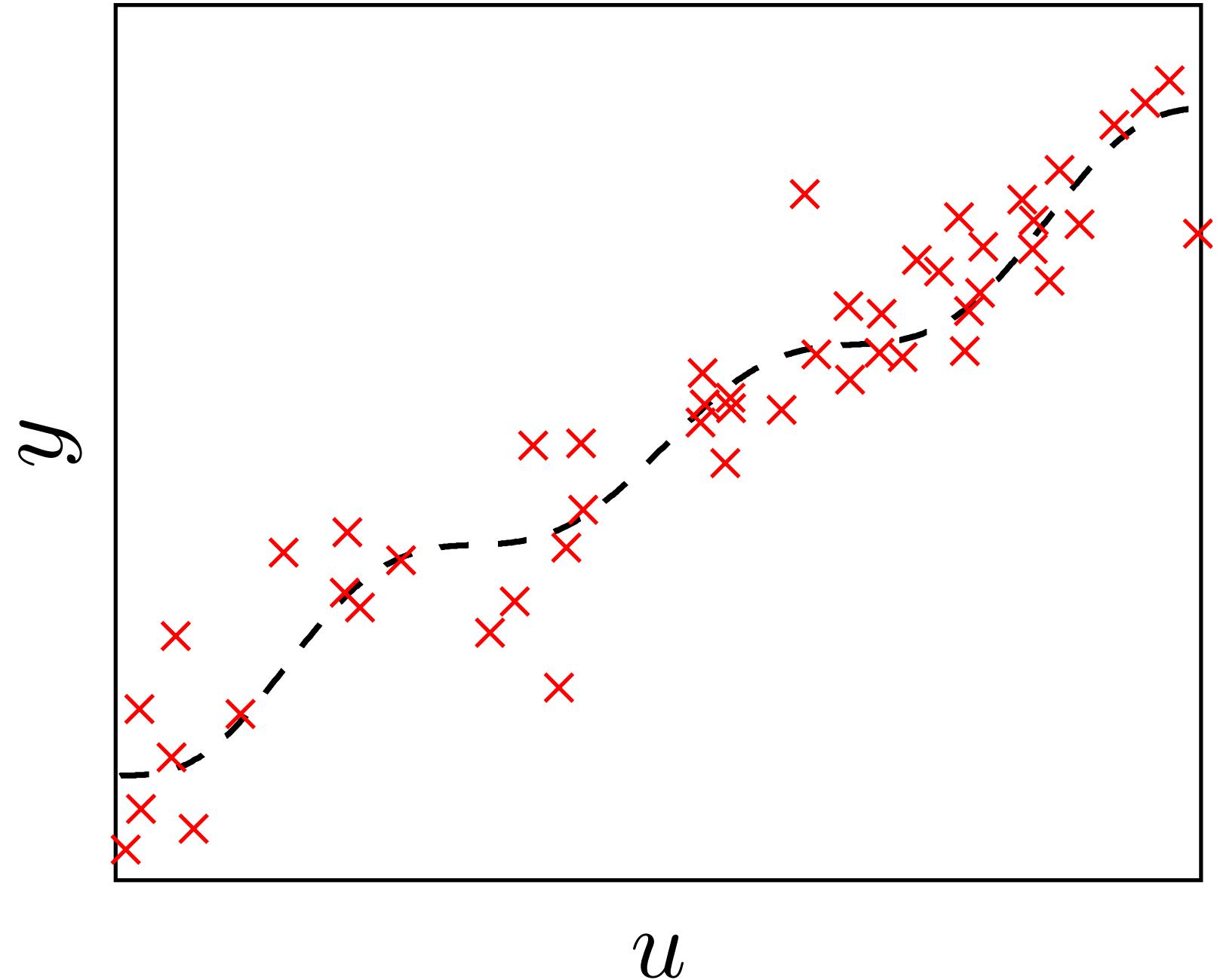
KLAUS NEUMANN^{*}
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
neumann@co-lab.de

MATTHIAS ROLF
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jrolf@co-lab.de

JÖCHEN JACOB STEIL
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jsteil@co-lab.de

The application of machine learning methods in the engineering of intelligent technical systems often requires the integration of continuous constraints like physical measures or bounds on variables in the learning task to guarantee reliable performance. We show that the extreme learning machine is particularly well suited for this task. The ELM is a linear model that can be trained very quickly. It is also easily implemented through quadratic optimization because the learned function is linear in its parameters. In this paper we propose a novel approach to handle continuous constraints in ELMs. This approach is based on a convex relaxation of the non-convex quadratic programming problem that arises from the ELM's quadratic loss function. We verify that discretely sampled constraints are generalized to continuous regions by means of a novel quadratic programming approach to verify the constraints. We demonstrate the approach on a practical and challenging control problem from robotics. We show that the proposed approach is able to learn from few data samples if additional prior knowledge about the problem is available.

Keywords: extreme learning machine, neural network, prior knowledge, continuous constraints, regression.



Adapted from Neumann (2013)

Context: Neumann, Rolf and Steil (2012)

RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS
INTO EXTREME LEARNING MACHINES

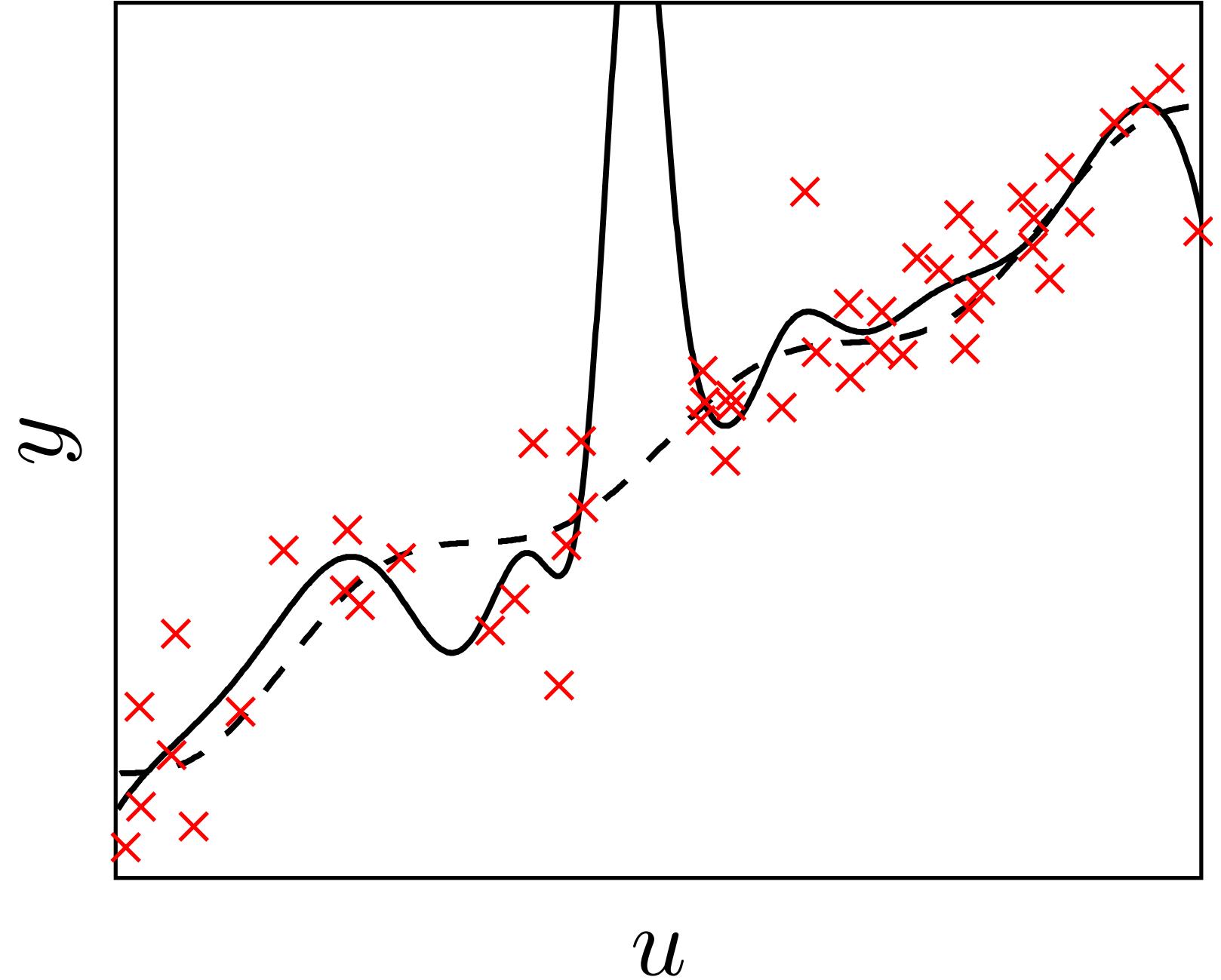
KLAUS NEUMANN^{*}
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
kneumann@co-lab.de

MATTHIAS ROLF
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jrolf@co-lab.de

JÖCHEN JACOB STEIL
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jsteil@co-lab.de

The application of machine learning methods in the engineering of intelligent technical systems often requires the integration of continuous constraints like position, momentum or velocity limits. In this paper we propose a reliable machine learning approach to implement such constraints in the learning process. We show that the extreme learning machine is particularly well suited for this task. We demonstrate that the proposed approach is more reliable than a quadratic programming approach implemented through quadratic optimization because the learned function is linear in its parameters. We also propose a novel approach to verify that discrete sampled constraints are generalized to continuous regions. This is done by applying a local quadratic approximation to the learned function. We demonstrate the approach on a practical and challenging control problem from robotics. We show that the proposed approach is able to learn a smooth function from few data samples if additional prior knowledge about the problem is available.

Keywords: extreme learning machine, neural network, prior knowledge, continuous constraints, regression.



Adapted from Neumann (2013)

Context: Neumann, Rolf and Steil (2012)

RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS
INTO EXTREME LEARNING MACHINES

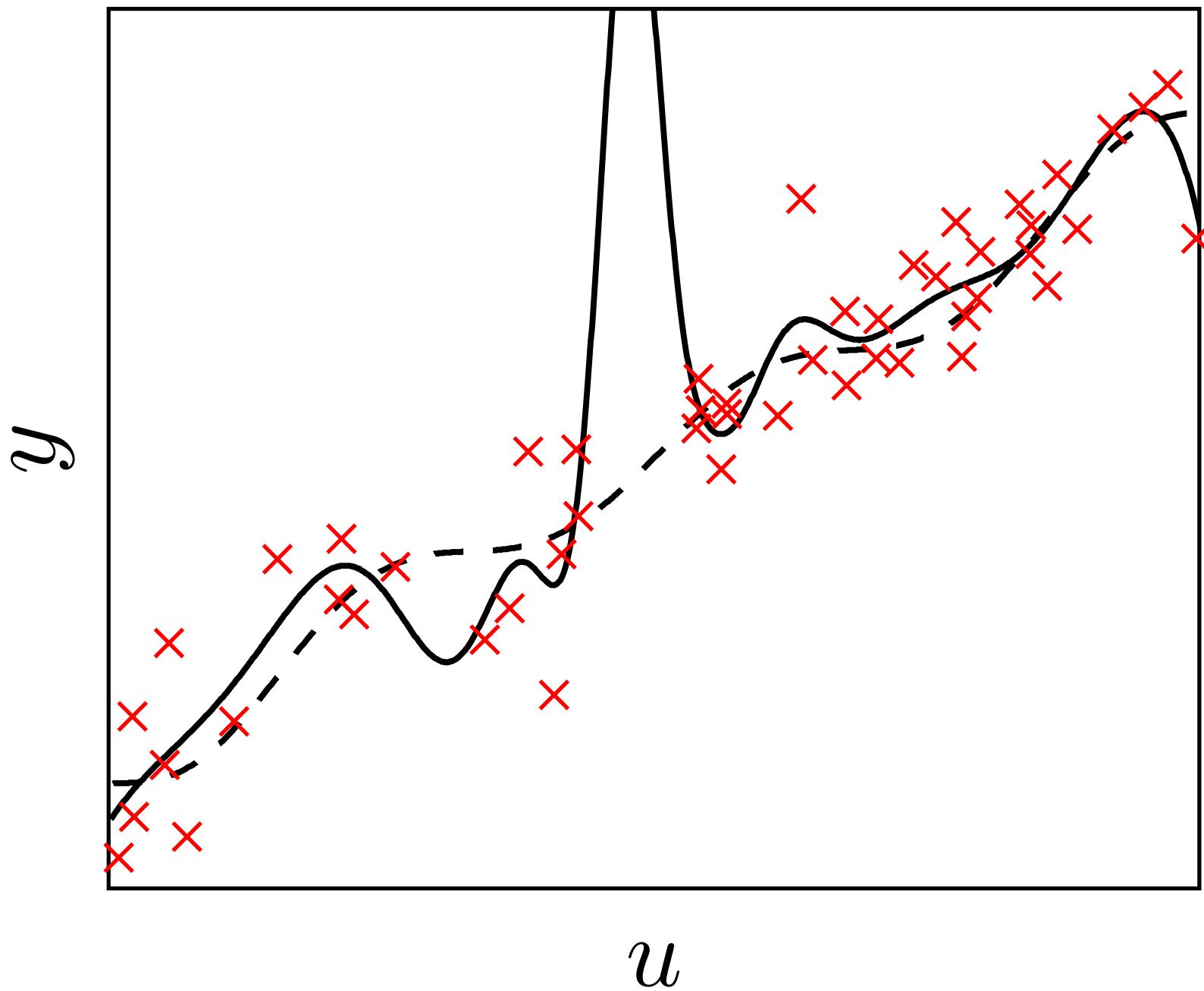
KLAUS NEUMANN^{*}
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
kneumann@co-lab.de

MATTHIAS ROLF
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jrolf@co-lab.de

JÖCHEN JACOB STEIL
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jsteil@co-lab.de

The application of machine learning methods in the engineering of intelligent technical systems often requires the integration of continuous constraints like positive/negative orthants or bounded variables in the learning process to reliably measure performance. We show that the extreme learning machine is particularly well suited for this task. We demonstrate that the ELM can be easily extended to incorporate continuous constraints by implementing quadratic optimization because the learned function is linear in its parameters. We propose a novel approach to verify that discretely sampled constraints are generalized to continuous regions. This is done by applying a quadratic programming approach to verify the constraints. We demonstrate the approach on a practical and challenging control problem from robotics. We show that the proposed approach is more reliable than learning from few data samples if additional prior knowledge about the problem is available.

Keywords: extreme learning machine, neural network, prior knowledge, continuous constraints, regression.

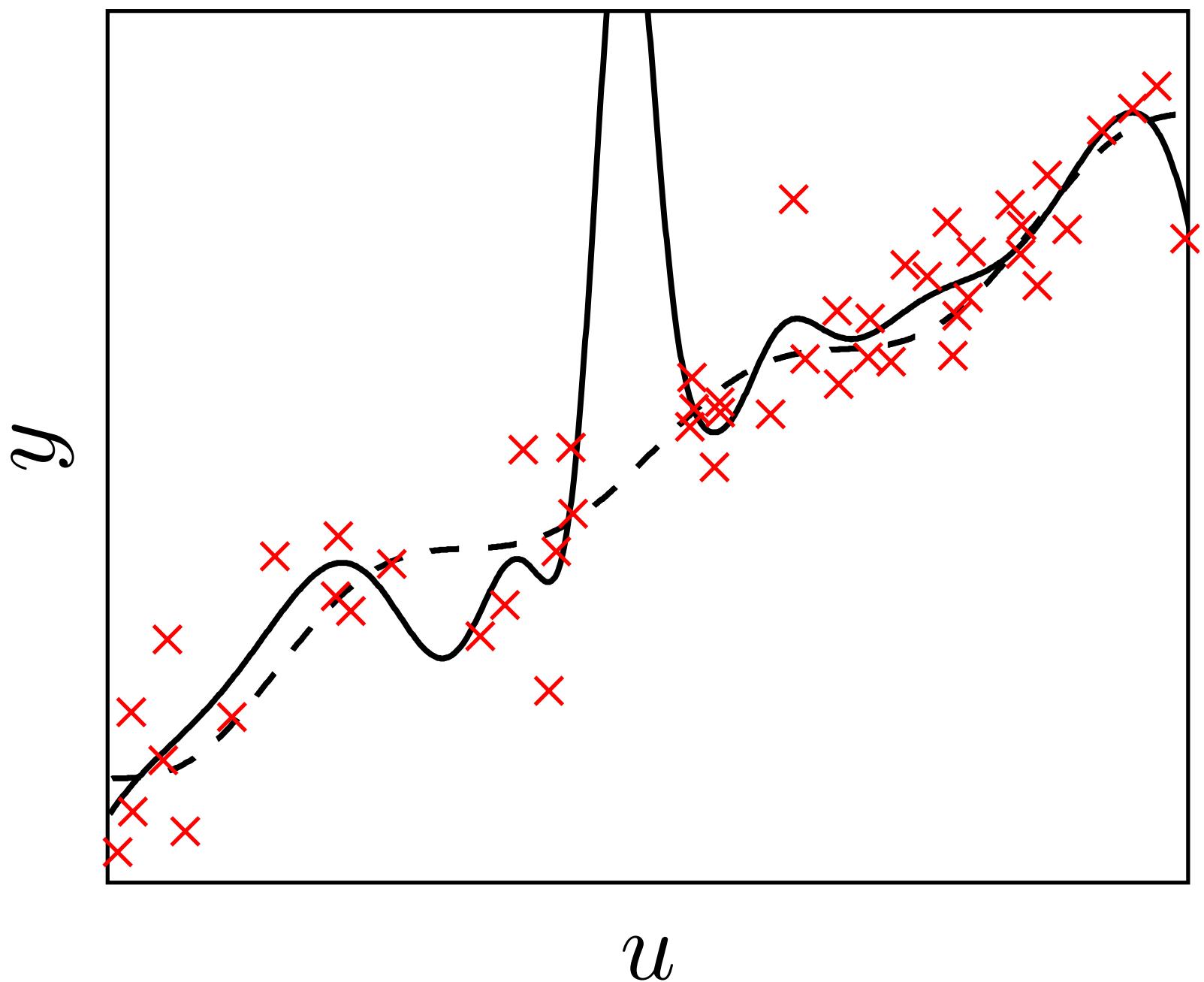


Target function steadily increasing:

$$u_1 \leq u_2 \Rightarrow y(u_1) \leq y(u_2)$$

Adapted from Neumann (2013)

Context: Neumann, Rolf and Steil (2012)



Target function steadily increasing:

$$u_1 \leq u_2 \Rightarrow y(u_1) \leq y(u_2)$$

or

$$\frac{\partial}{\partial u} y(u) \geq 0$$

Adapted from Neumann (2013)

Context: Neumann, Rolf and Steil (2012)

RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS
INTO EXTREME LEARNING MACHINES

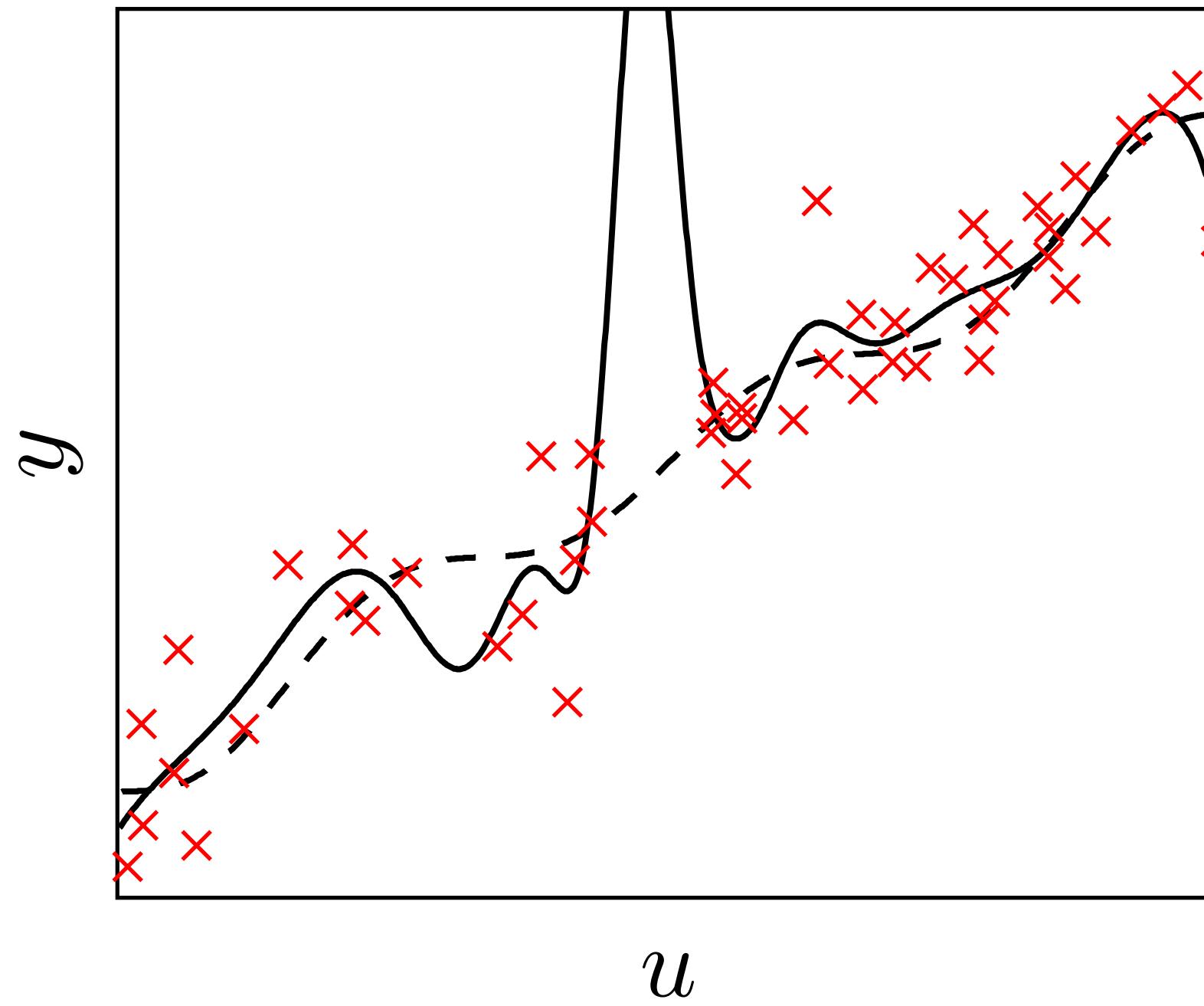
KLAUS NEUMANN^{*}
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
kneumann@co-lab.de

MATTHIAS ROLF
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jrolf@co-lab.de

JÖCHEN JACOB STEIL
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jsteil@co-lab.de

The application of machine learning methods in the engineering of intelligent technical systems often requires the integration of continuous constraints like piecewise monotonicity or convexity into the learning task in order to guarantee reliable performance. We show that the extreme learning machine is particularly well suited for this task. It is able to learn piecewise linear functions from discrete training samples implemented through quadratic optimization because the learned function is linear in its neurons. We propose a novel approach to verify that discretely sampled constraints are generalized to continuous regions. This is done by applying a quadratic programming approach to a convex active set problem. We demonstrate the approach on a practical and challenging control problem from robotics. We show that the proposed approach is more robust than learning from few data samples if additional prior knowledge about the problem is available.

Keywords: extreme learning machine, neural network, prior knowledge, continuous constraints, regression.



Target function steadily increasing:

$$u_1 \leq u_2 \Rightarrow y(u_1) \leq y(u_2)$$

or

$$\frac{\partial}{\partial u} y(u) \geq 0$$

Constraint

Adapted from Neumann (2013)

Context: Neumann, Rolf and Steil (2012)

RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS
INTO EXTREME LEARNING MACHINES

KLAUS NEUMANN^{*}
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
kneumann@co-lab.de

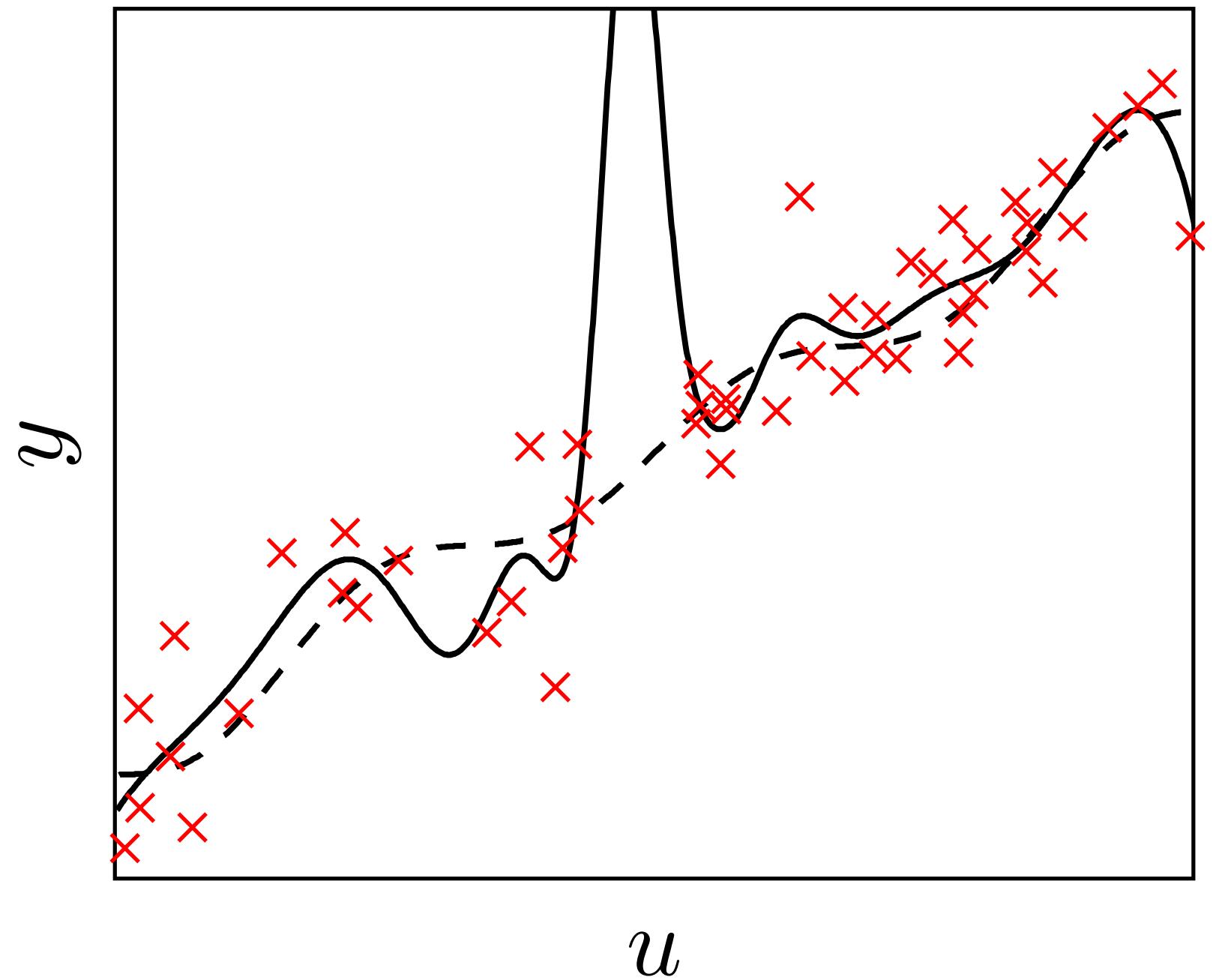
MATTHIAS ROLF
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jrolf@co-lab.de

JÖCHEN JACOB STEIL
Research Institute for Cognition and Robotics (Co-Lab), Bielefeld University
Universitätsstraße 25, 33513 Bielefeld, Germany
jsteil@co-lab.de

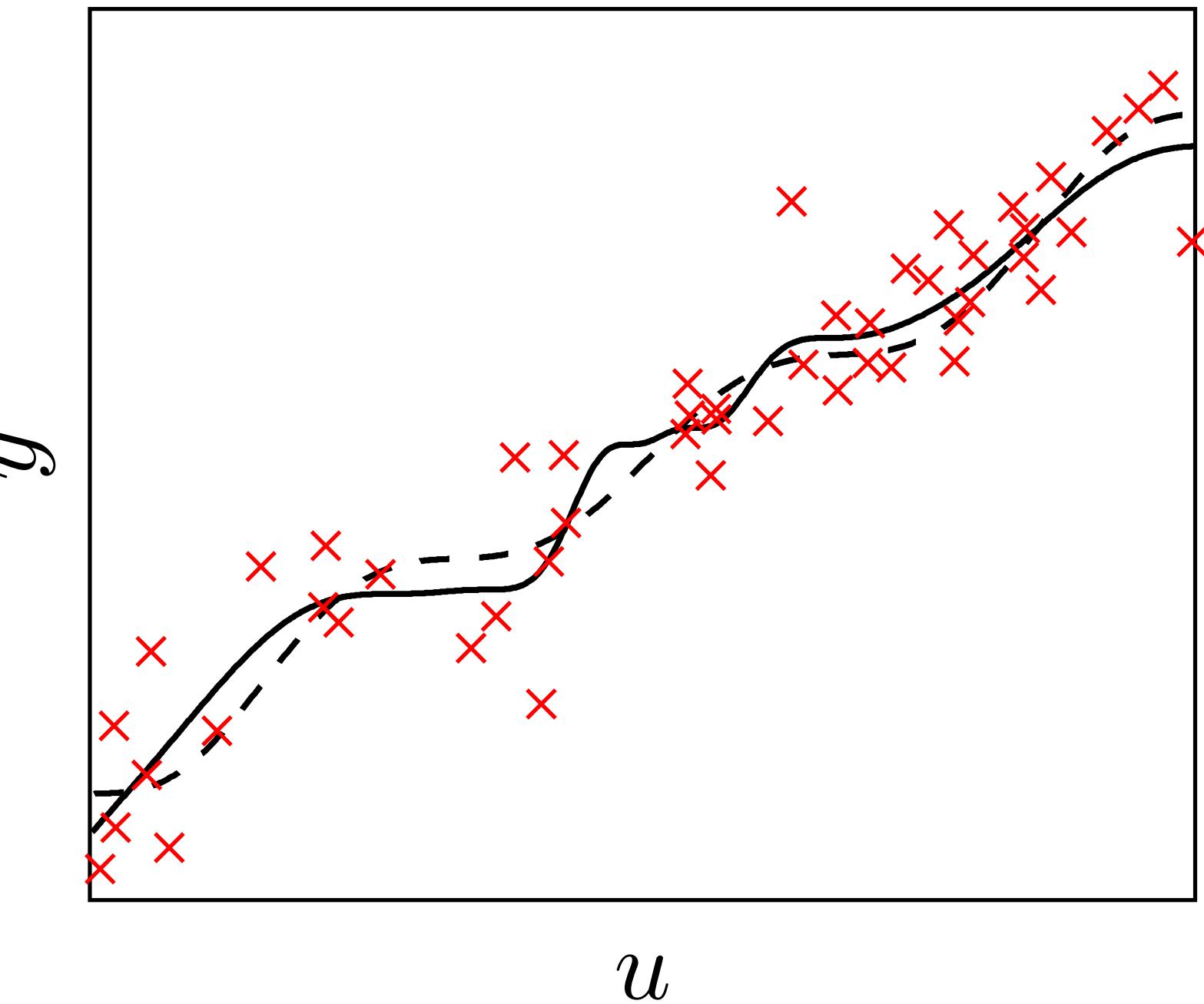
The application of machine learning methods in the engineering of intelligent technical systems often requires the integration of continuous constraints like monotonicity or boundedness. In this paper we propose a reliable approach to implement such constraints in the learning process of an extreme learning machine. We show that the extreme learning machine is particularly well suited for this task. The proposed approach is based on quadratic programming and can be easily implemented through quadratic optimization because the learned function is linear in its parameters. We also propose a method to verify that discretely sampled constraints are satisfied by the learned function. This allows us to verify the constraints in an iterative approach to verify that discretely sampled constraints are generalized to continuous regions. We demonstrate the approach on a practical and challenging control problem from the field of robotics. We demonstrate the approach on a practical and challenging control problem from the field of robotics. We demonstrate the approach on a practical and challenging control problem from the field of robotics. We demonstrate the approach on a practical and challenging control problem from the field of robotics.

Keywords: extreme learning machine, neural network, prior knowledge, continuous constraints, regression.

Without constraints



With monotonicity constraints



Adapted from Neumann (2013)

Adapting constraint to time series case

Static Case

(Feedforward Neural Network)

Input u

Output y

VS.

Dynamic Case

(Recurrent Neural Network)

Input $u(1), \dots, u(t-1), u(t)$

Output $y(1), \dots, y(t-1), y(t)$

Adapting constraint to time series case

Static Case

(Feedforward Neural Network)

Input u

Output y

VS.

Dynamic Case

(Recurrent Neural Network)

Input $u(1), \dots, u(t-1), u(t)$

Output $y(1), \dots, y(t-1), y(t)$



Constraints describe sensitivity
of y w.r.t. u

?

Adapting constraint to time series case

Static Case

(Feedforward Neural Network)

Input u

Output y

VS.

Dynamic Case

(Recurrent Neural Network)

Input $u(1), \dots, u(t-1), u(t)$

Output $y(1), \dots, y(t-1), y(t)$



Constraints describe sensitivity
of y w.r.t. u

Constraints describe sensitivity
of y w.r.t. time t

Time-dependent Constraint*

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$

*simplified to one-dimensional output

Time-dependent Constraint*

Number of time steps in history

H

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$

coefficients

bound

network output

A mathematical expression representing a time-dependent constraint. The expression is $\sum_{h=0}^H \gamma_h y(t-h) \leq c$. The term H is labeled 'Number of time steps in history'. The term γ_h is labeled 'coefficients'. The term $y(t-h)$ is labeled 'network output'. The term c is labeled 'bound'. Red arrows indicate the correspondence between the labels and the respective terms in the equation.

*simplified to one-dimensional output

Examples

Constraint Definition:

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$

Examples

- Upper (or lower) bound to the output:

$$y(t) \leq 0.4$$

Constraint Definition:

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$

Examples

- Upper (or lower) bound to the output:

$$y(t) \leq 0.4$$

- Steadily decreasing (or increasing) output:

$$y(t) - y(t-1) \leq 0$$

Constraint Definition:

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$

Examples

- Upper (or lower) bound to the output:

$$y(t) \leq 0.4$$

- Steadily decreasing (or increasing) output:

$$y(t) - y(t - 1) \leq 0$$

- Periodically repeating output (with period P):

$$y(t) - y(t - P) \leq 0$$

$$-y(t) + y(t - P) \leq 0$$

Constraint Definition:

$$\sum_{h=0}^H \gamma_h y(t - h) \leq c$$

Examples

- Upper (or lower) bound to the output:

$$y(t) \leq 0.4$$

- Steadily decreasing (or increasing) output:

$$y(t) - y(t - 1) \leq 0$$

- Periodically repeating output (with period P):

$$y(t) - y(t - P) \leq 0$$

$$-y(t) + y(t - P) \leq 0$$

- Difference Quotient of arbitrary order

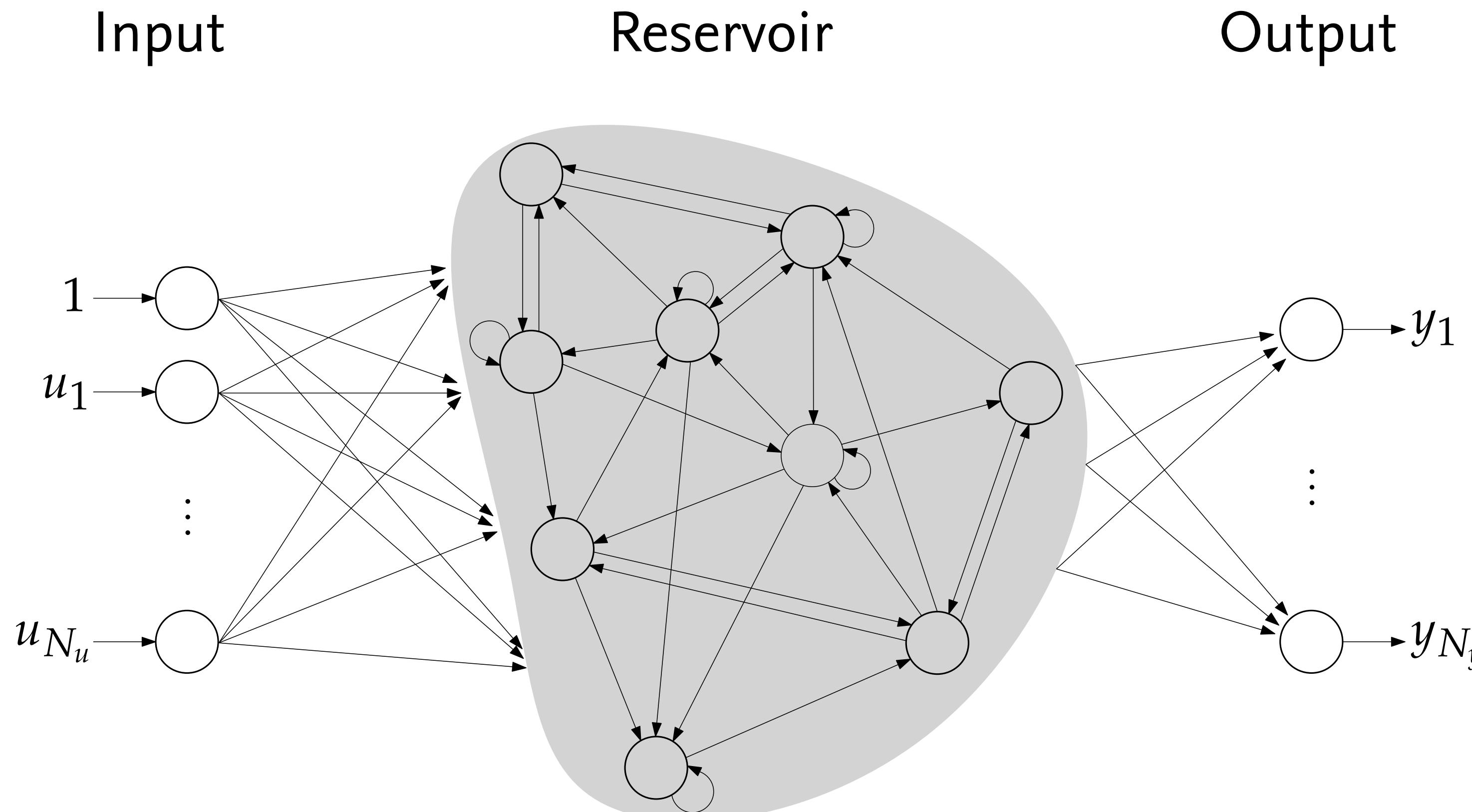
Constraint Definition:

$$\sum_{h=0}^H \gamma_h y(t - h) \leq c$$

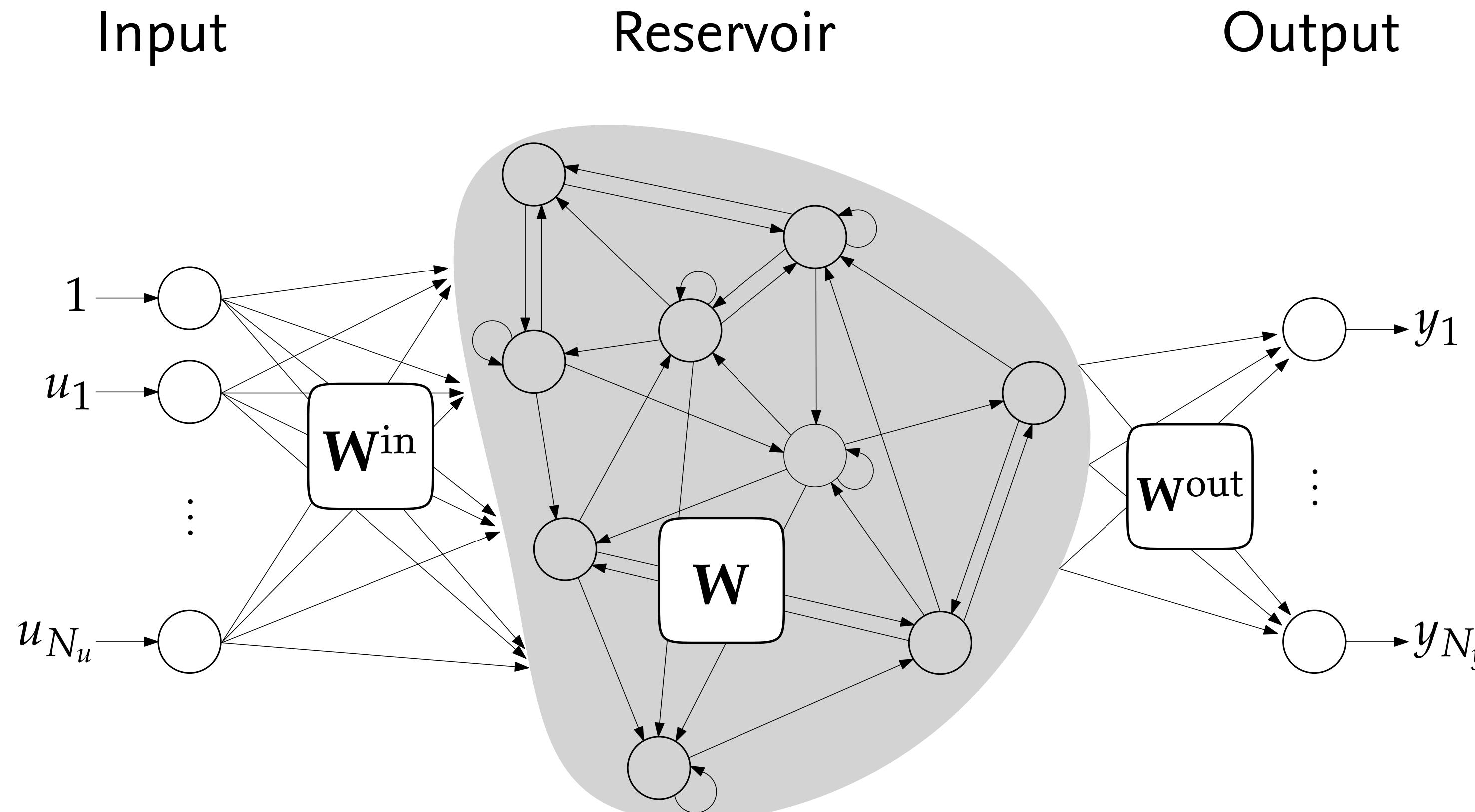
Outline

1. Constraint Definition
2. Embedding Constraints into the Neural Network
3. Example: Forecasting of Satellite Images
4. Future Work & Conclusion

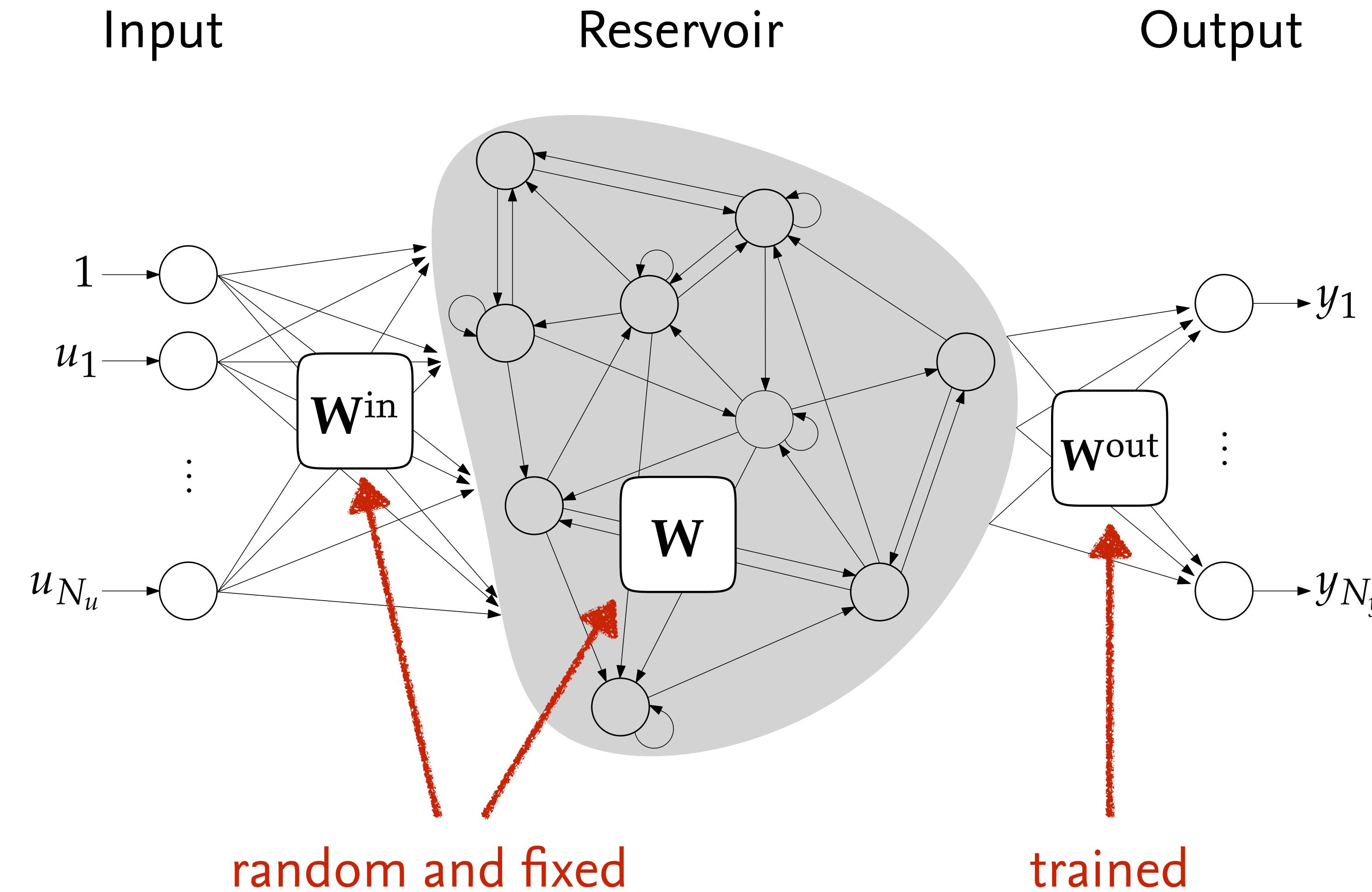
The Echo State Network (1/3)

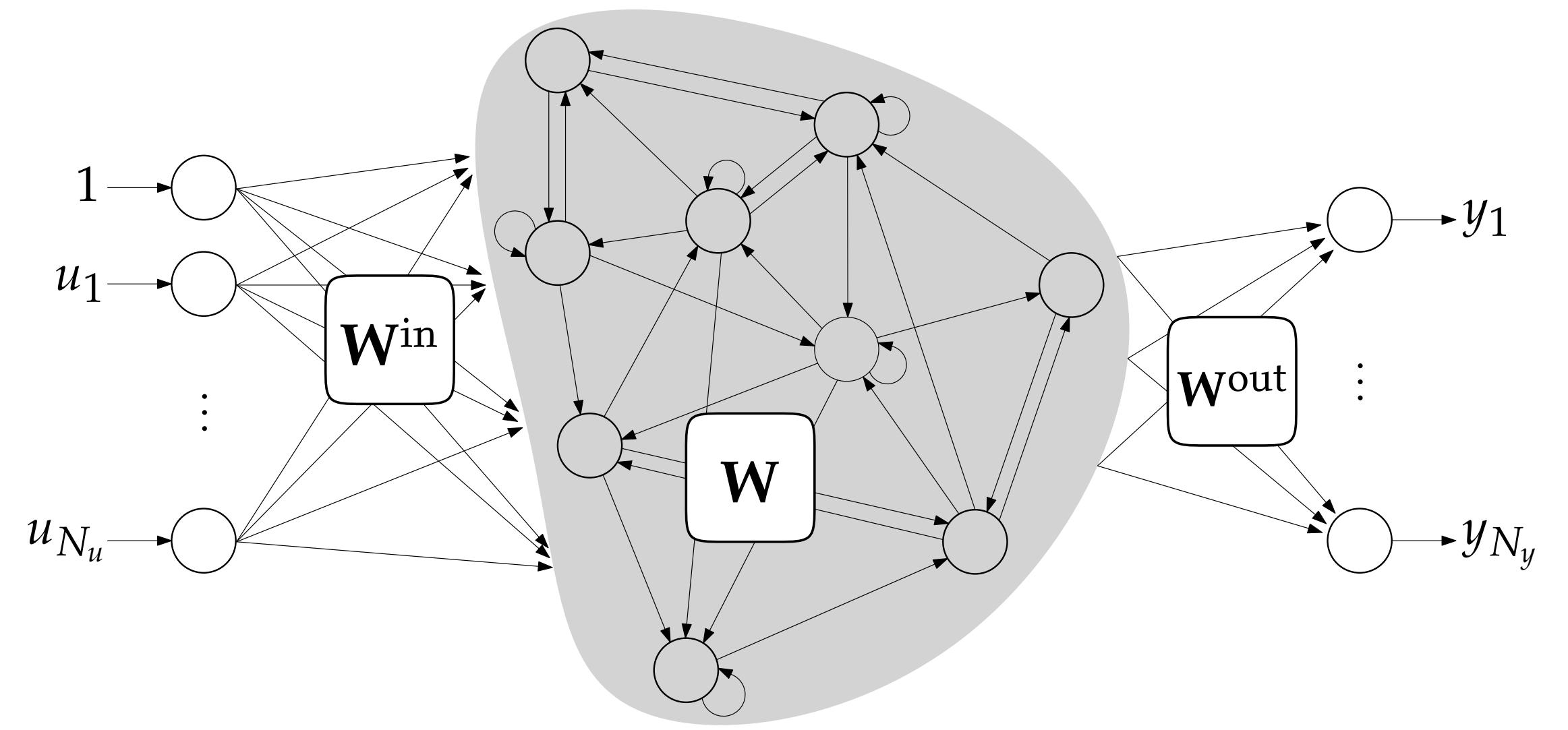


The Echo State Network (1/3)

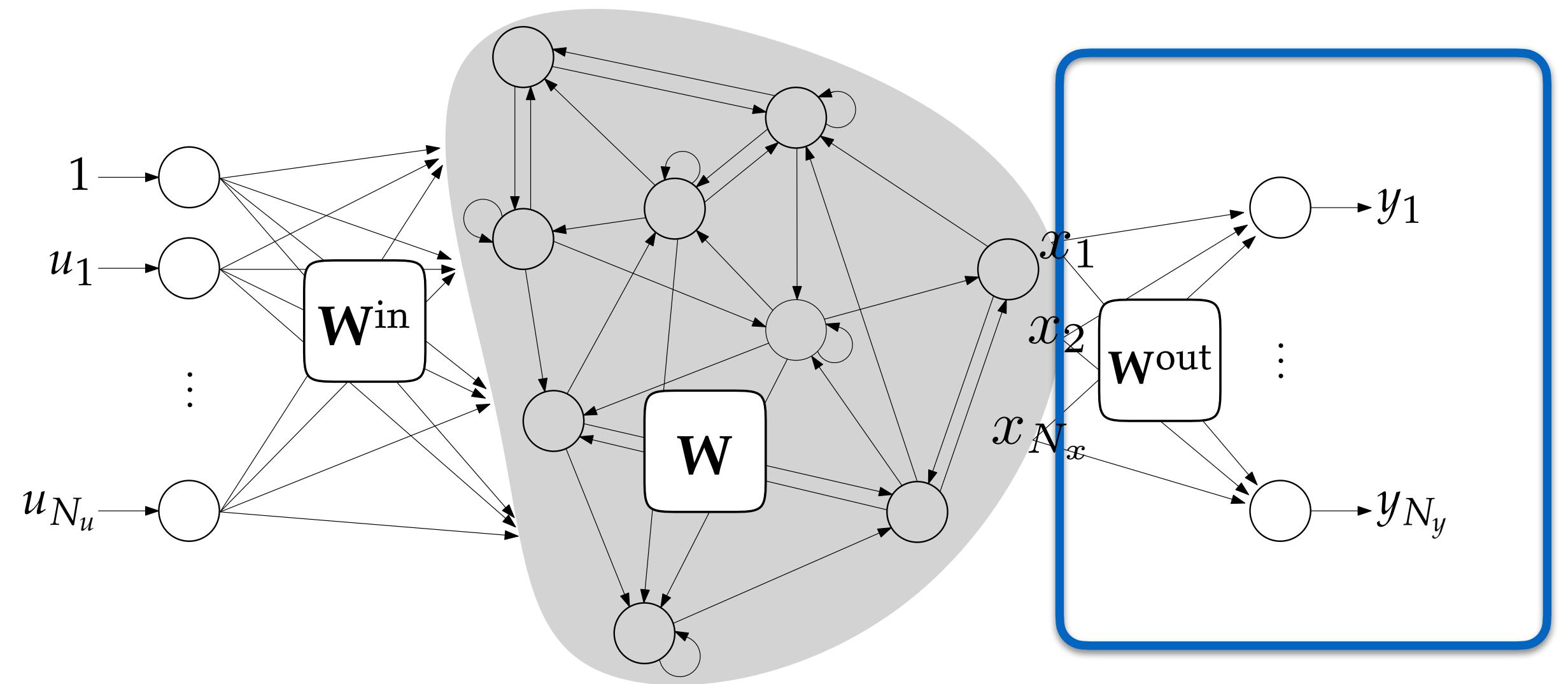


The Echo State Network (1/3)





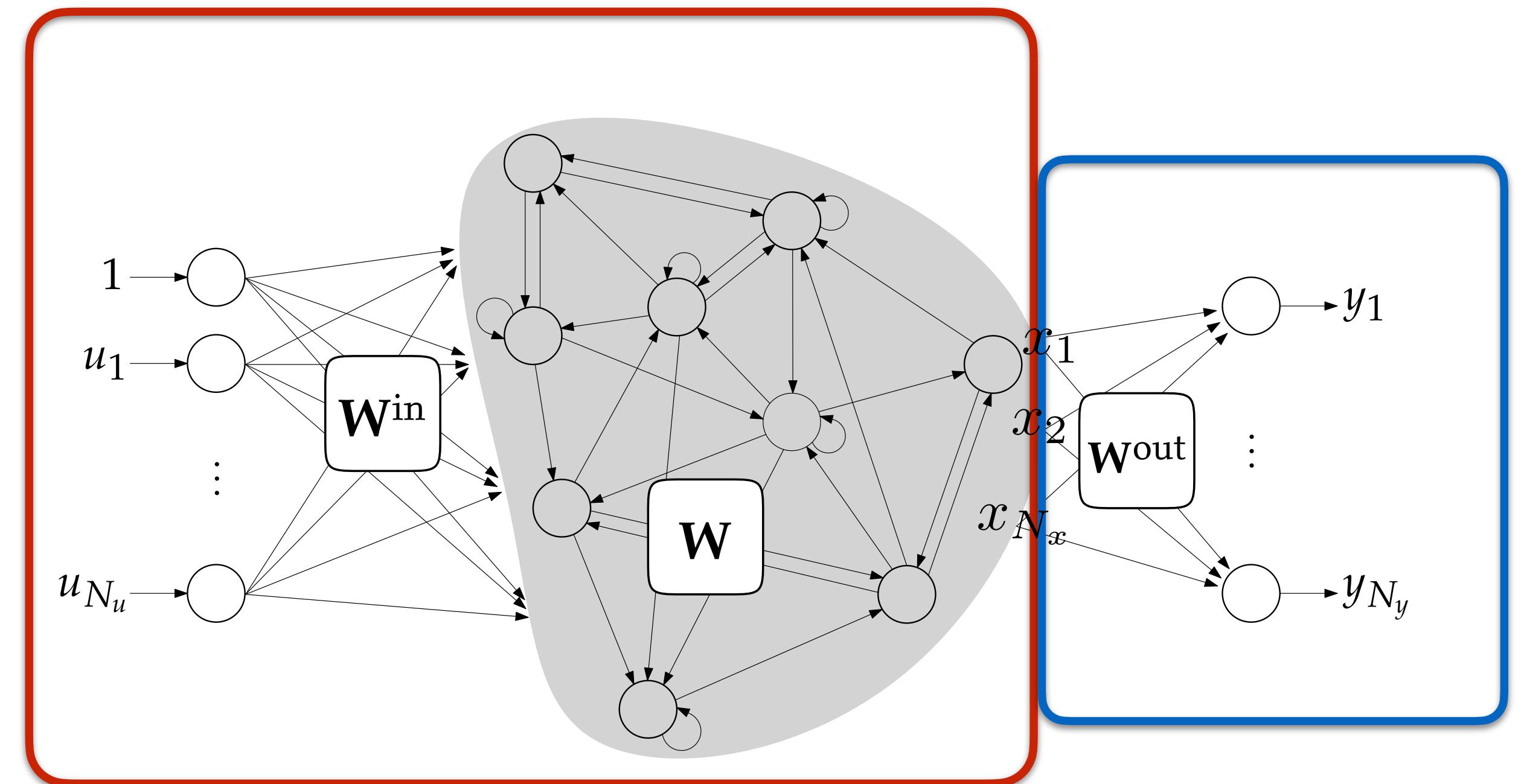
$$\mathbf{y}(t) = \mathbf{W}^{\text{out}} \mathbf{x}(t)$$



- feedforward
- linear read-out

$$\mathbf{x}(t) = \sigma (\mathbf{W}^{\text{in}} \mathbf{u}(t) + \mathbf{W} \mathbf{x}(t - 1))$$

$$\mathbf{y}(t) = \mathbf{W}^{\text{out}} \mathbf{x}(t)$$



- high-dimensional
- non-linear
- recurrent
- feedforward
- linear read-out

The Echo State Network (3/3)

The Echo State Network (3/3)

- Minimize Training Error:

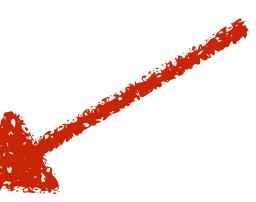
$$\epsilon_{\text{RMSE}} = \|\mathbf{Y} - \mathbf{Y}^{\text{target}}\|^2$$

The Echo State Network (3/3)

- Minimize Training Error:

$$\begin{aligned}\epsilon_{\text{RMSE}} &= \|\mathbf{Y} - \mathbf{Y}^{\text{target}}\|^2 \\ &= \|\mathbf{W}^{\text{out}} \mathbf{x} - \mathbf{Y}^{\text{target}}\|^2\end{aligned}$$

$\mathbf{y}(t) = \mathbf{W}^{\text{out}} \mathbf{x}(t)$

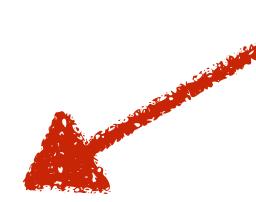


The Echo State Network (3/3)

- Minimize Training Error:

$$\begin{aligned}\epsilon_{\text{RMSE}} &= \|\mathbf{Y} - \mathbf{Y}^{\text{target}}\|^2 \\ &= \|\mathbf{W}^{\text{out}} \mathbf{x} - \mathbf{Y}^{\text{target}}\|^2 \\ &= \|\mathbf{x}^T (\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\|^2\end{aligned}$$

$\mathbf{y}(t) = \mathbf{W}^{\text{out}} \mathbf{x}(t)$



The Echo State Network (3/3)

- Minimize Training Error:

$$\begin{aligned}\epsilon_{\text{RMSE}} &= \|\mathbf{Y} - \mathbf{Y}^{\text{target}}\|^2 \\ &= \|\mathbf{W}^{\text{out}} \mathbf{x} - \mathbf{Y}^{\text{target}}\|^2 \quad \text{y}(t) = \mathbf{W}^{\text{out}} \mathbf{x}(t) \\ &= \|\mathbf{x}^T (\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\|^2\end{aligned}$$

- Equivalent to a Linear Least Squares Problem:

$$\mathbf{W}^{\text{out}} = \arg \min_{\mathbf{W}^{\text{out}}} \|\mathbf{x}^T (\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\|^2$$

The Echo State Network (3/3)

- Minimize Training Error:

$$\begin{aligned}\epsilon_{\text{RMSE}} &= \|\mathbf{Y} - \mathbf{Y}^{\text{target}}\|^2 \\ &= \|\mathbf{W}^{\text{out}} \mathbf{X} - \mathbf{Y}^{\text{target}}\|^2 \quad \text{y}(t) = \mathbf{W}^{\text{out}} \mathbf{x}(t) \\ &= \|\mathbf{X}^T (\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\|^2\end{aligned}$$

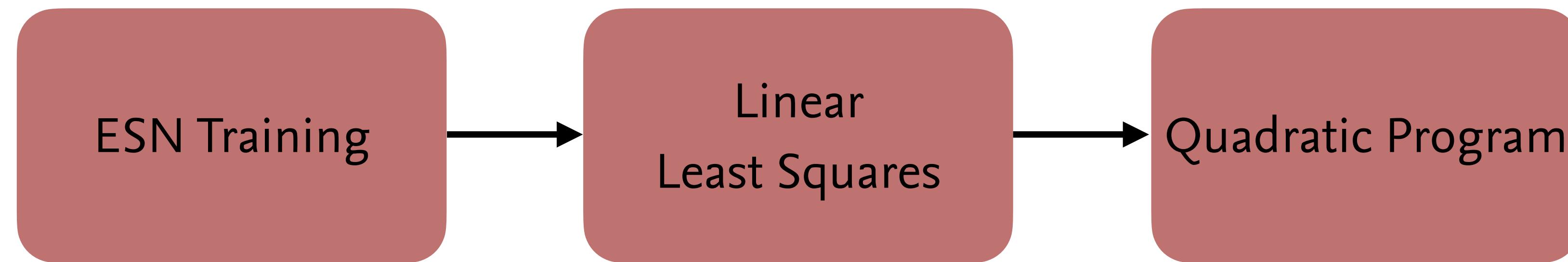
- Equivalent to a Linear Least Squares Problem:

$$\mathbf{W}^{\text{out}} = \arg \min_{\mathbf{W}^{\text{out}}} \|\mathbf{X}^T (\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\|^2$$

Linear Least Squares:

$$\min \|\mathbf{AX} - \mathbf{b}\|^2$$

Reduction to Quadratic Program



$$\min \epsilon_{\text{RMSE}}$$

$$\min \| \mathbf{AX} - \mathbf{b} \|^2$$

$$\begin{aligned} \min & \mathbf{x}^T \mathbf{Q} \mathbf{x} \\ \text{s.t. } & \mathbf{G} \mathbf{x} \leq \mathbf{h} \end{aligned}$$

Constraints* to QP constraints

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$

*simplified to one-dimensional output

Constraints* to QP constraints

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$
$$\Leftrightarrow \sum_{h=0}^H \gamma_h \mathbf{W}^{\text{out}} \mathbf{x}(t-h) \leq c$$

*simplified to one-dimensional output

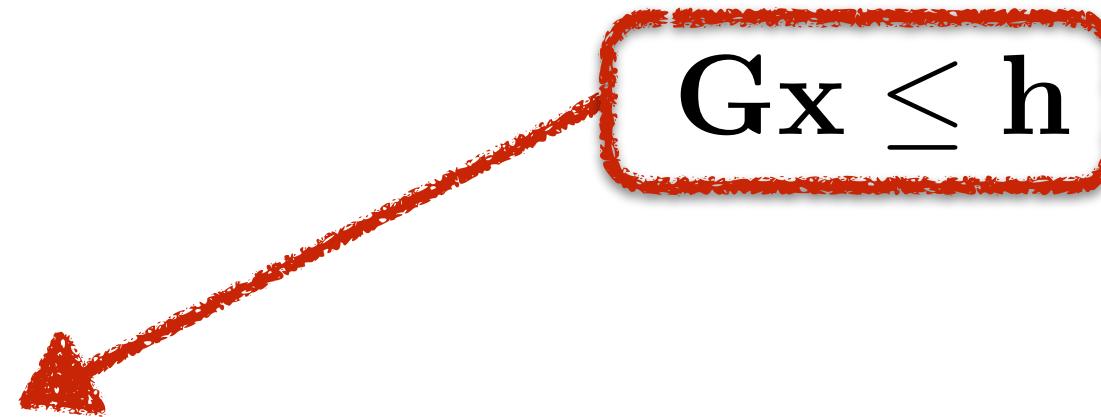
Constraints* to QP constraints

$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$

$$\Leftrightarrow \sum_{h=0}^H \gamma_h \mathbf{W}^{\text{out}} \mathbf{x}(t-h) \leq c$$

$$\Leftrightarrow \left(\sum_{h=0}^H \gamma_h \mathbf{x}(t-h)^T \right) \cdot (\mathbf{W}^{\text{out}})^T \leq c$$

$$\mathbf{Gx} \leq \mathbf{h}$$

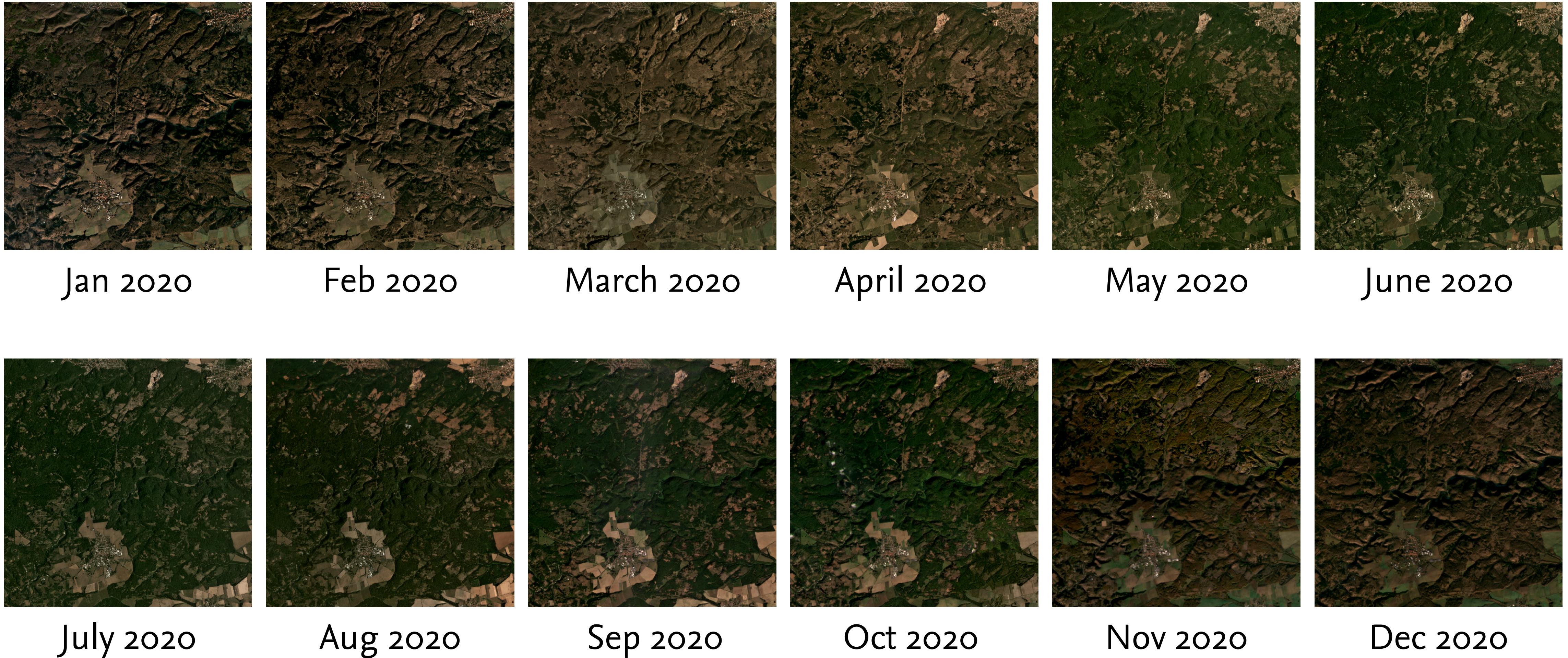


*simplified to one-dimensional output

Outline

1. Constraint Definition
2. Embedding Constraints into the Neural Network
3. Example: Forecasting of Satellite Images
4. Future Work & Conclusion

Satellite Image Time Series of the Harz (monthly)



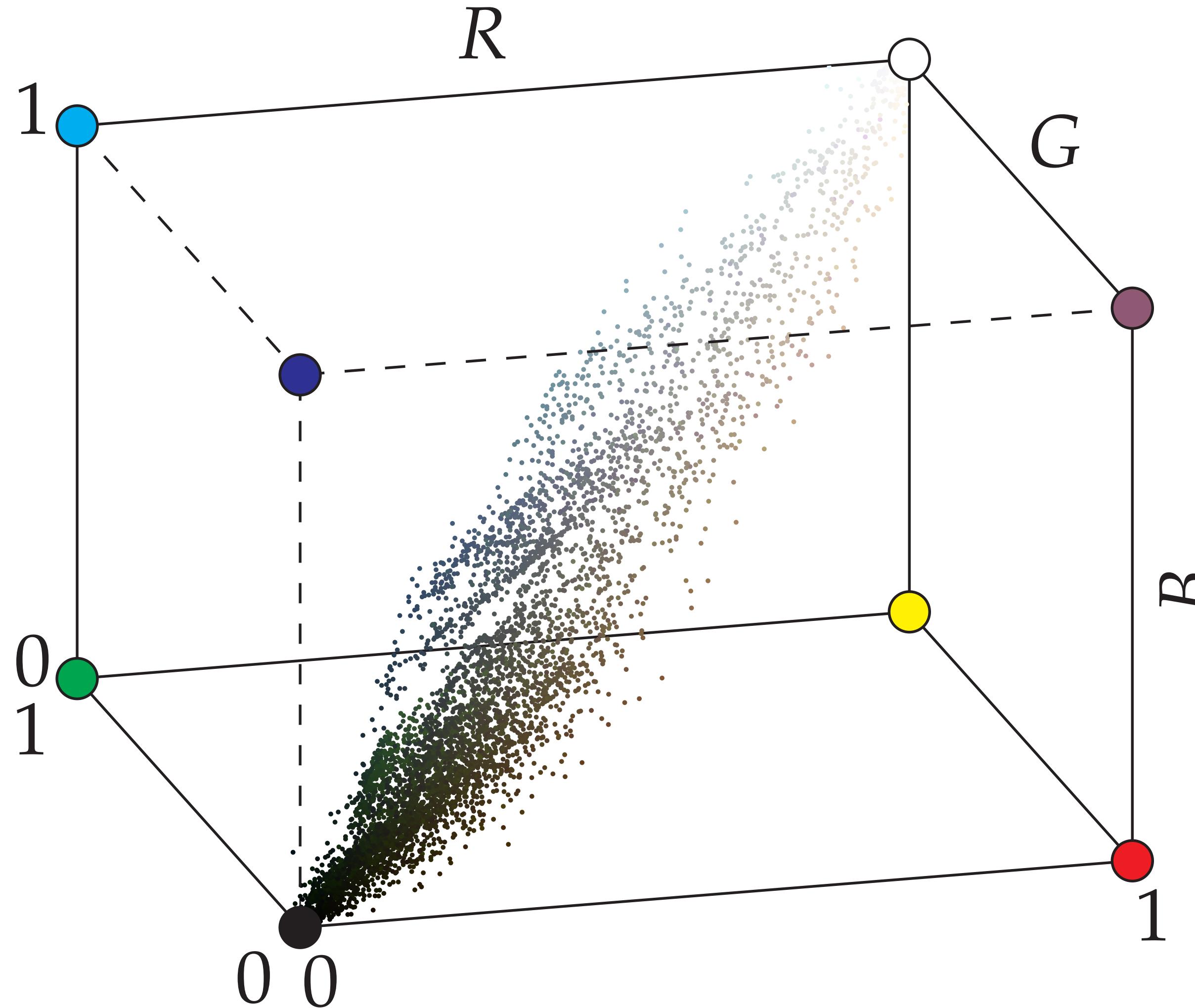
Noisy Images



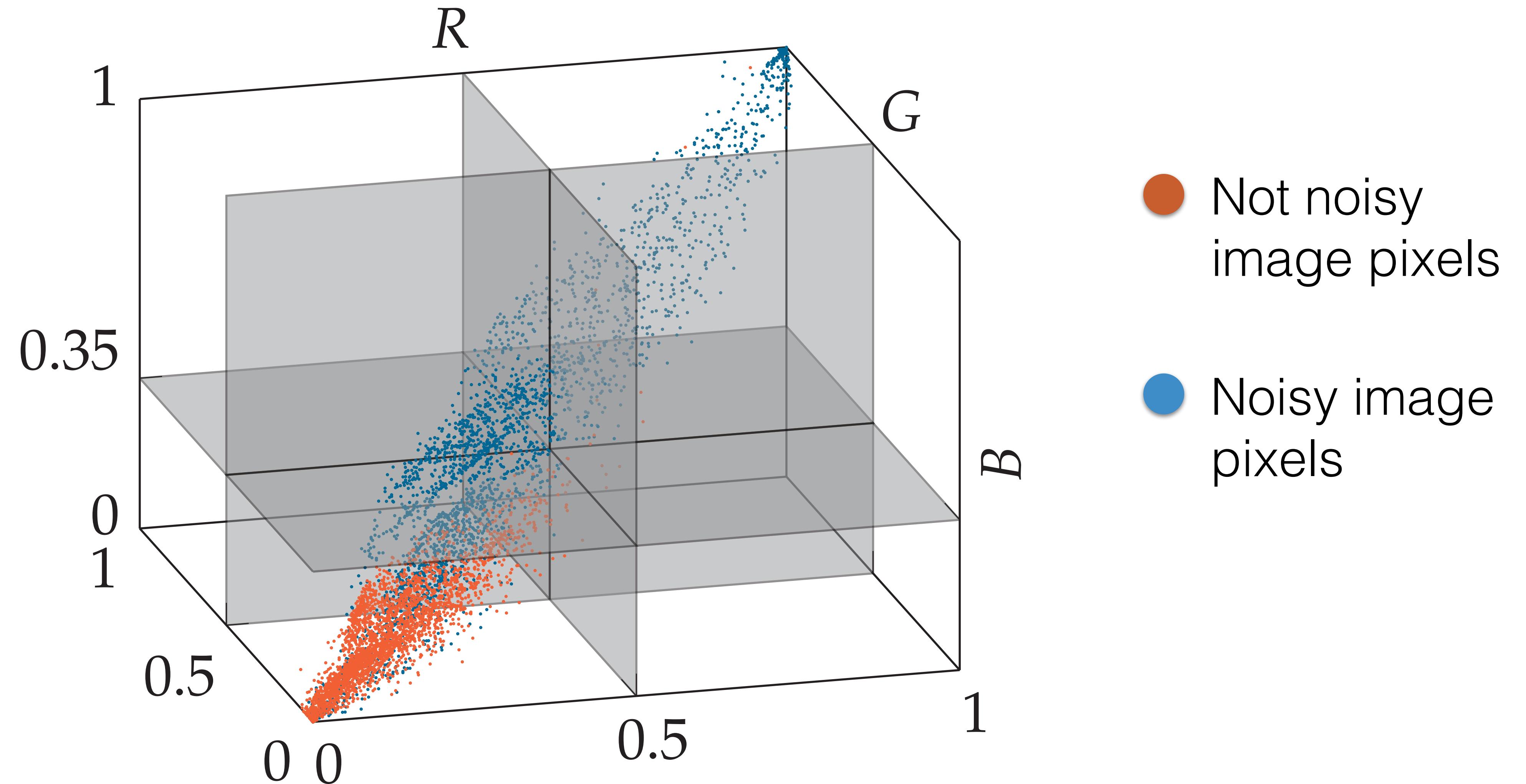
Training Task

- Given a time series of satellite images, predict upcoming month
- Train on single pixels
- Three inputs and three outputs per pixel (RGB)

Boundary Constraints



Boundary Constraints



Constraints

Boundary Constraints

$$y_R(t) \leq 0.5$$

$$y_G(t) \leq 0.5$$

$$y_B(t) \leq 0.35$$

Constraints

Boundary Constraints

$$y_R(t) \leq 0.5$$

$$y_G(t) \leq 0.5$$

$$y_B(t) \leq 0.35$$

Difference Constraints

$$|y_R(t) - y_R(t-1)| \leq 0.05$$

$$|y_G(t) - y_G(t-1)| \leq 0.05$$

$$|y_B(t) - y_B(t-1)| \leq 0.05$$

Constraints

Boundary Constraints

$$y_R(t) \leq 0.5$$

$$y_G(t) \leq 0.5$$

$$y_B(t) \leq 0.35$$

Difference Constraints

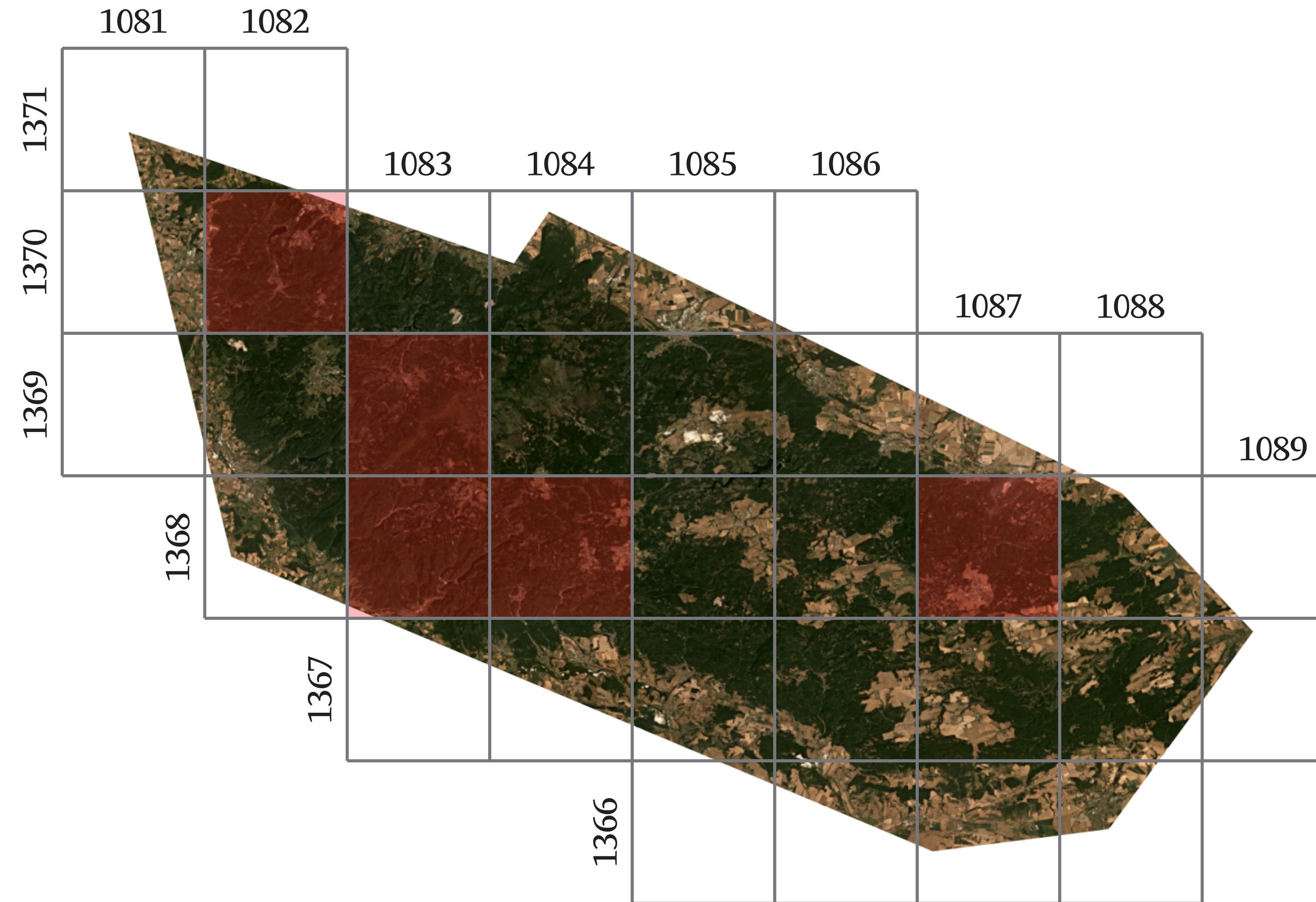
$$|y_R(t) - y_R(t-1)| \leq 0.05$$

$$|y_G(t) - y_G(t-1)| \leq 0.05$$

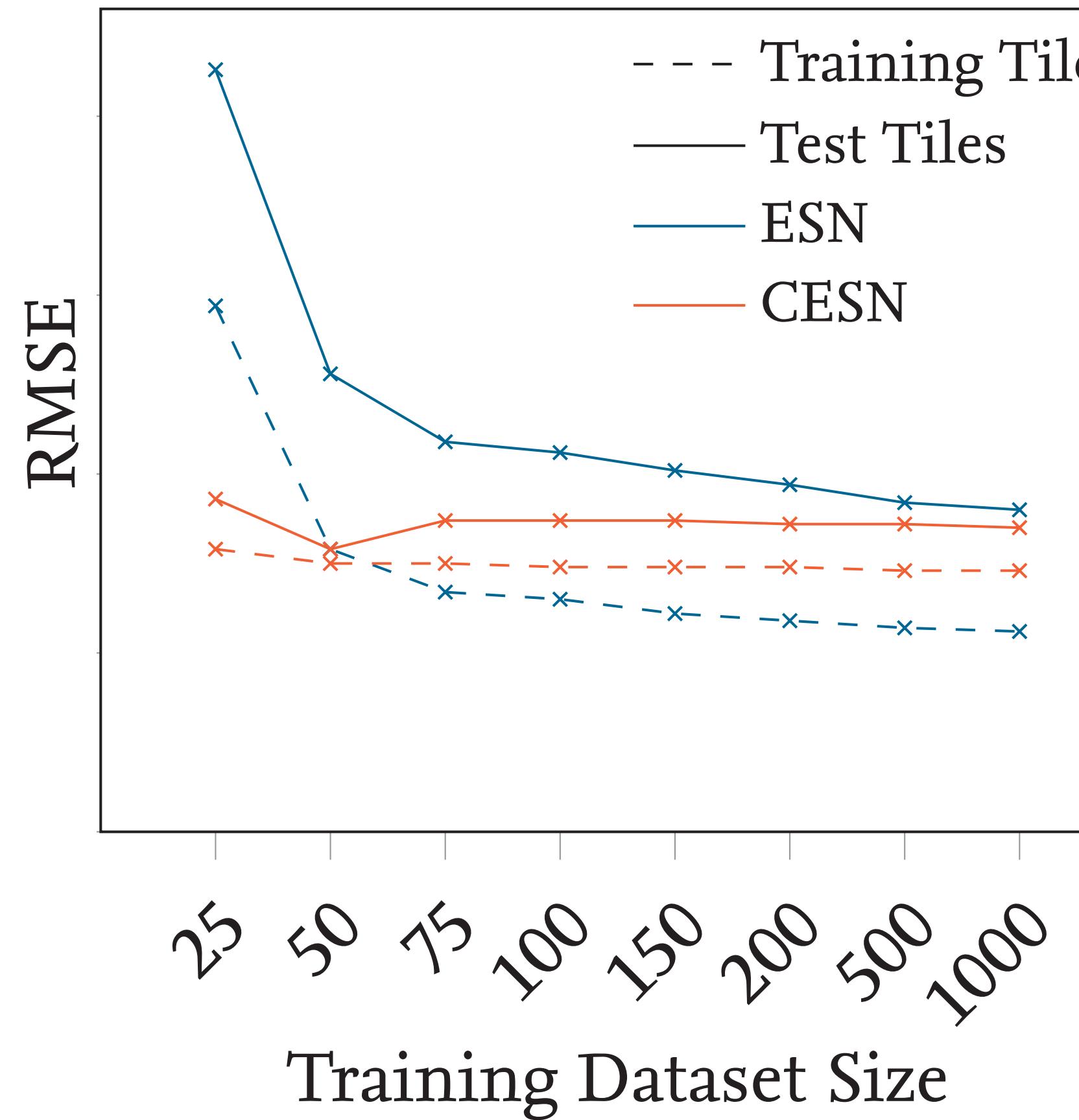
$$|y_B(t) - y_B(t-1)| \leq 0.05$$

→ 9 constraints at each time step in total

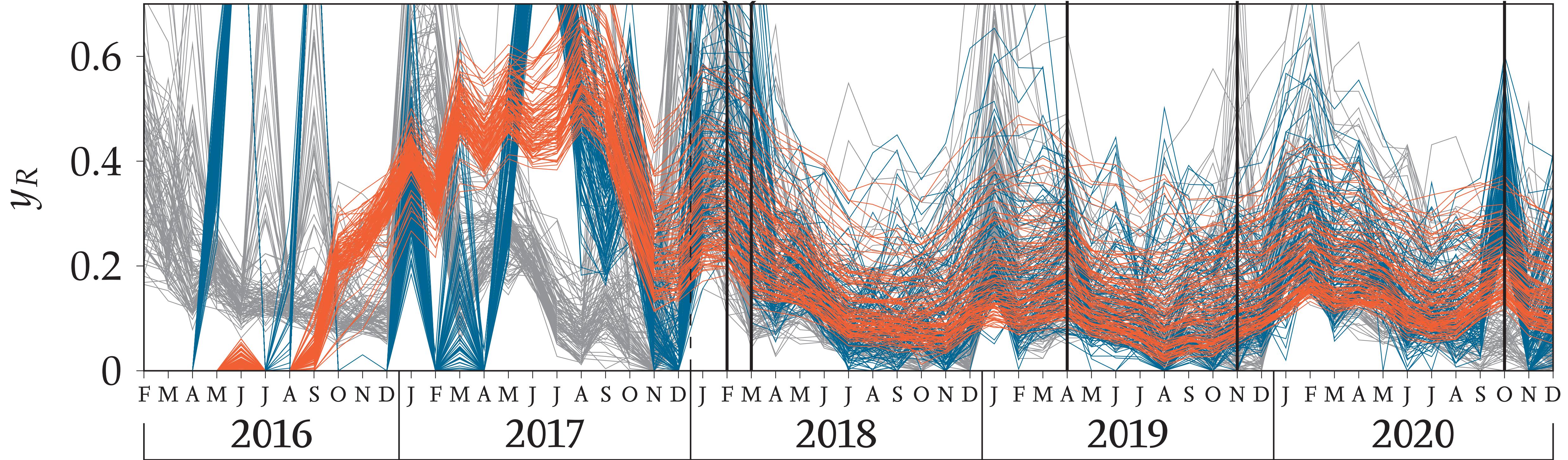
Training Method



Results (1/2)



Results (2/2)



Prediction Examples (1/3)

Prediction



ESN

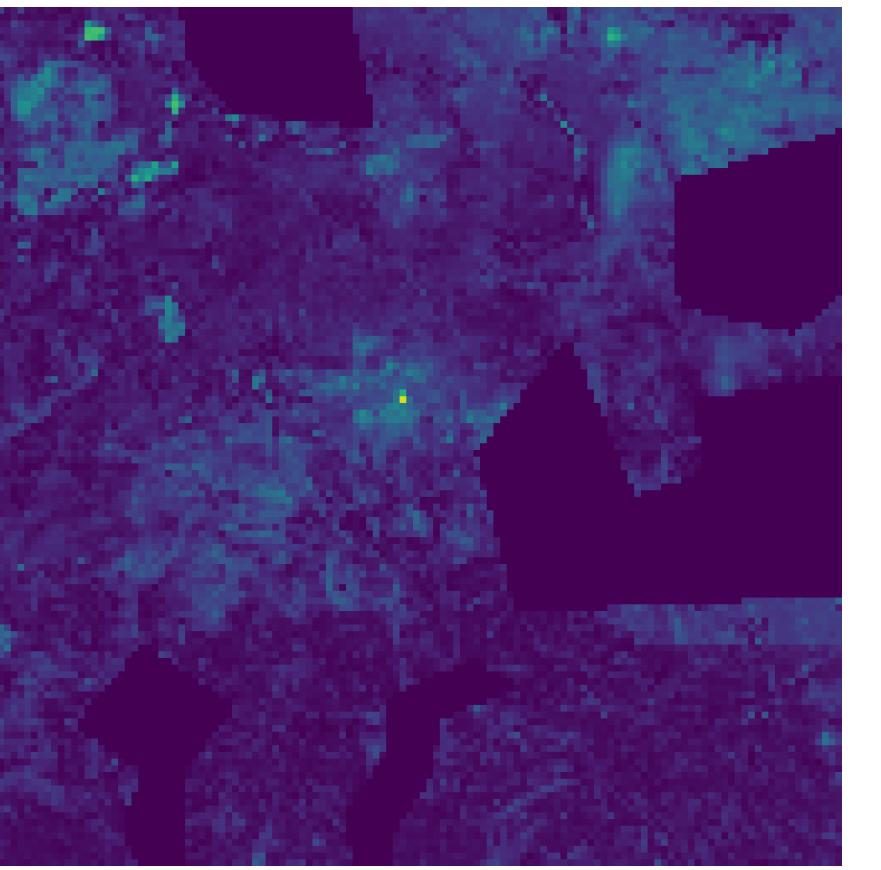
Target



CESN



Error

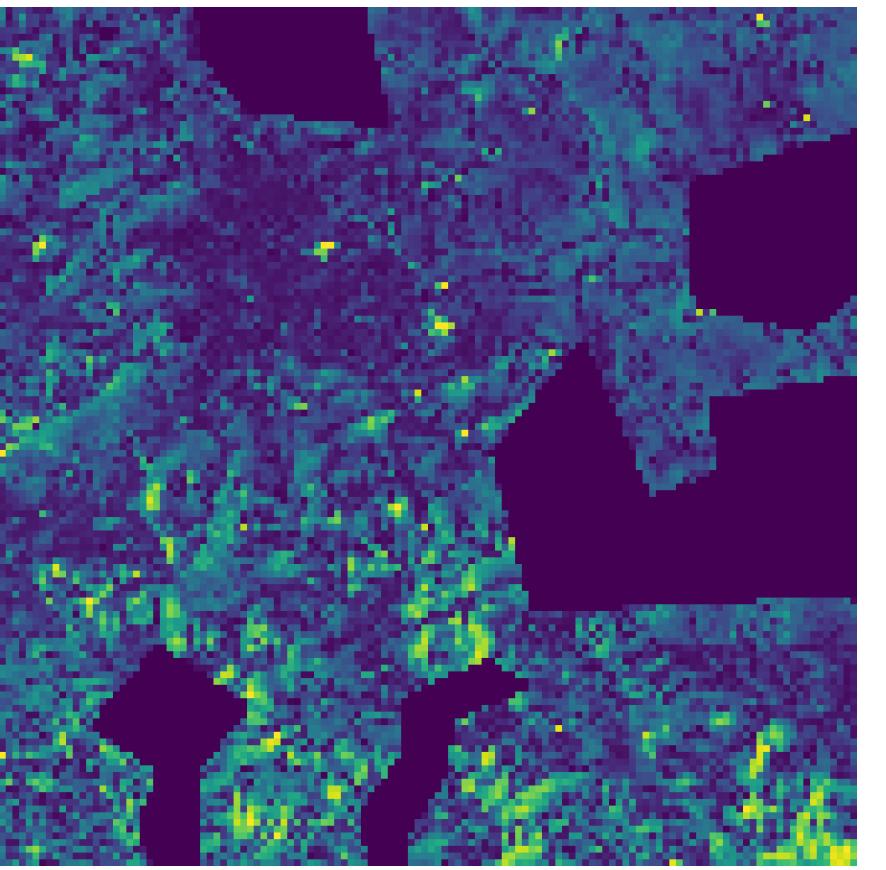


Prediction Examples (2/3)

ESN
Prediction



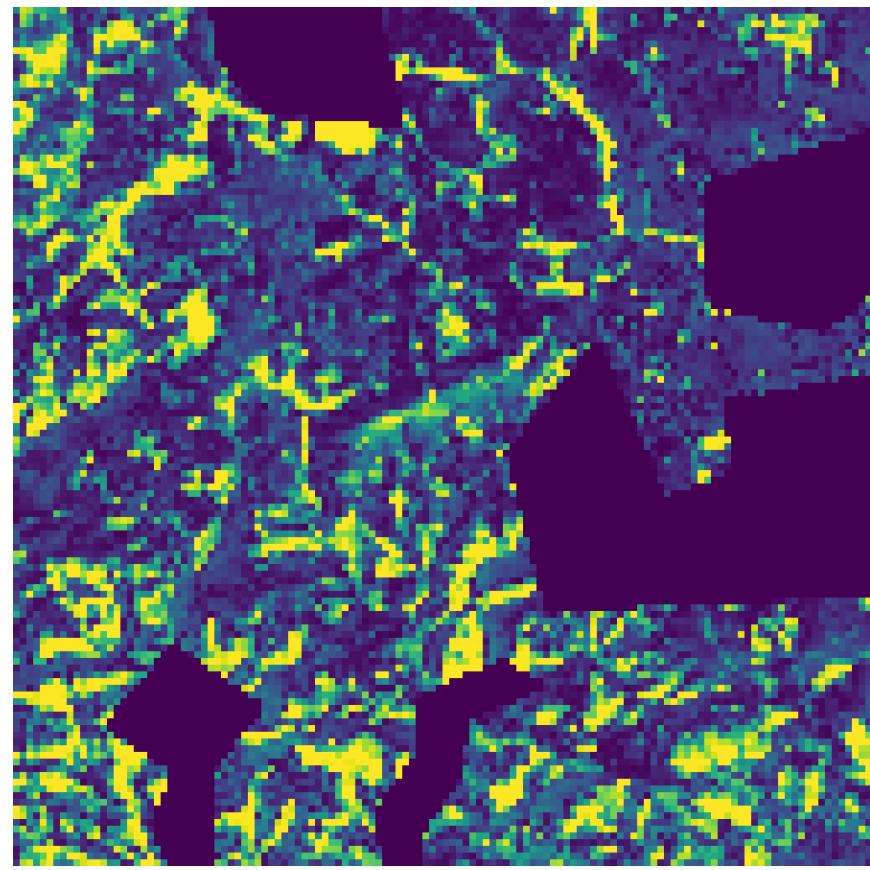
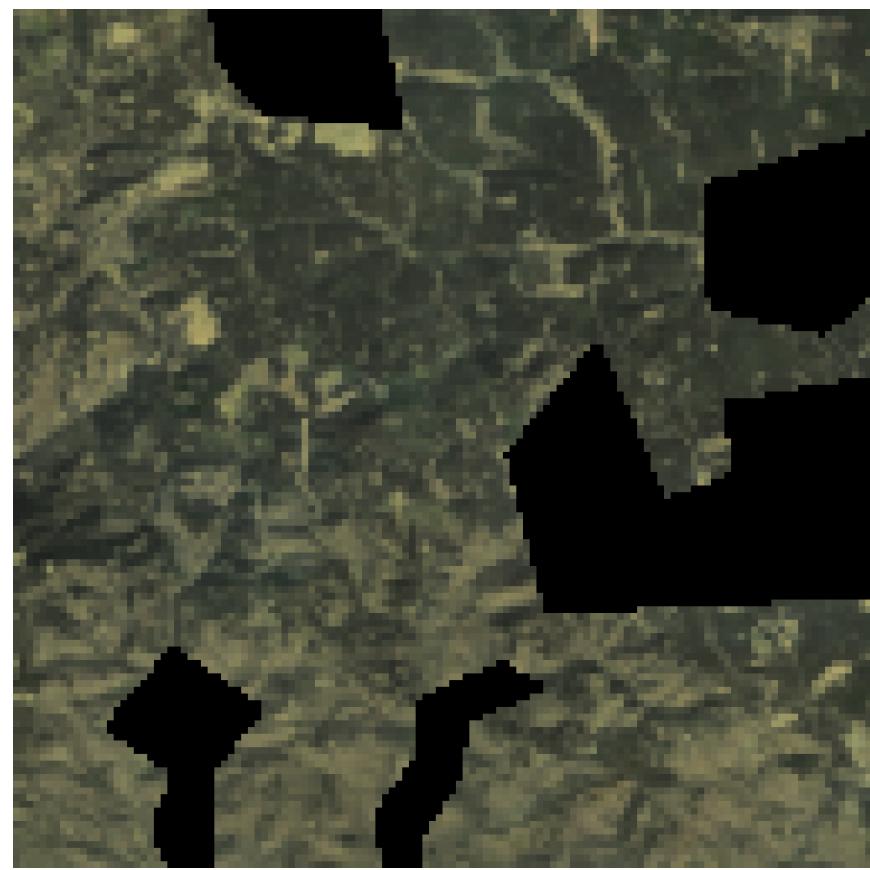
Error



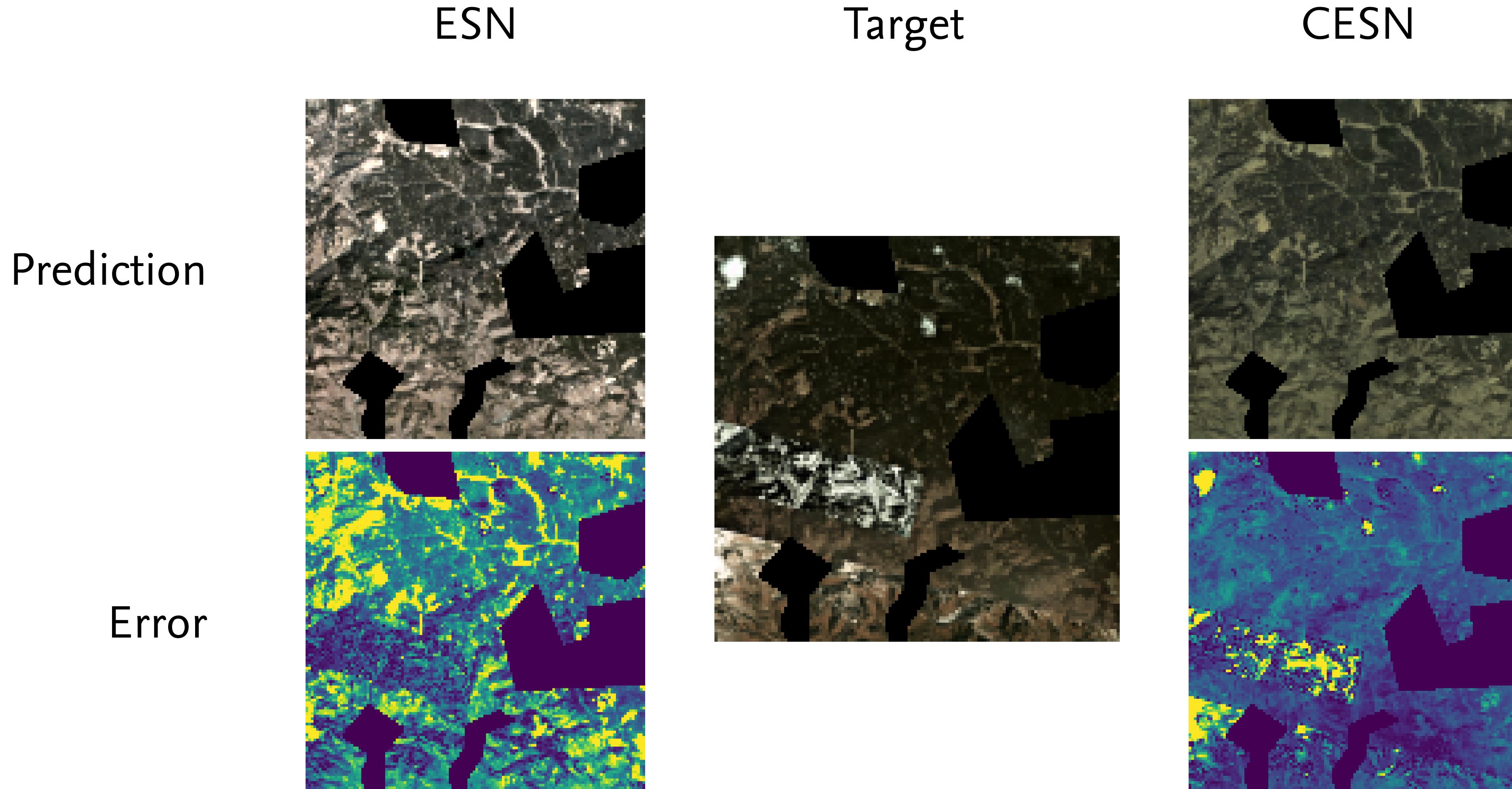
Target



CESN



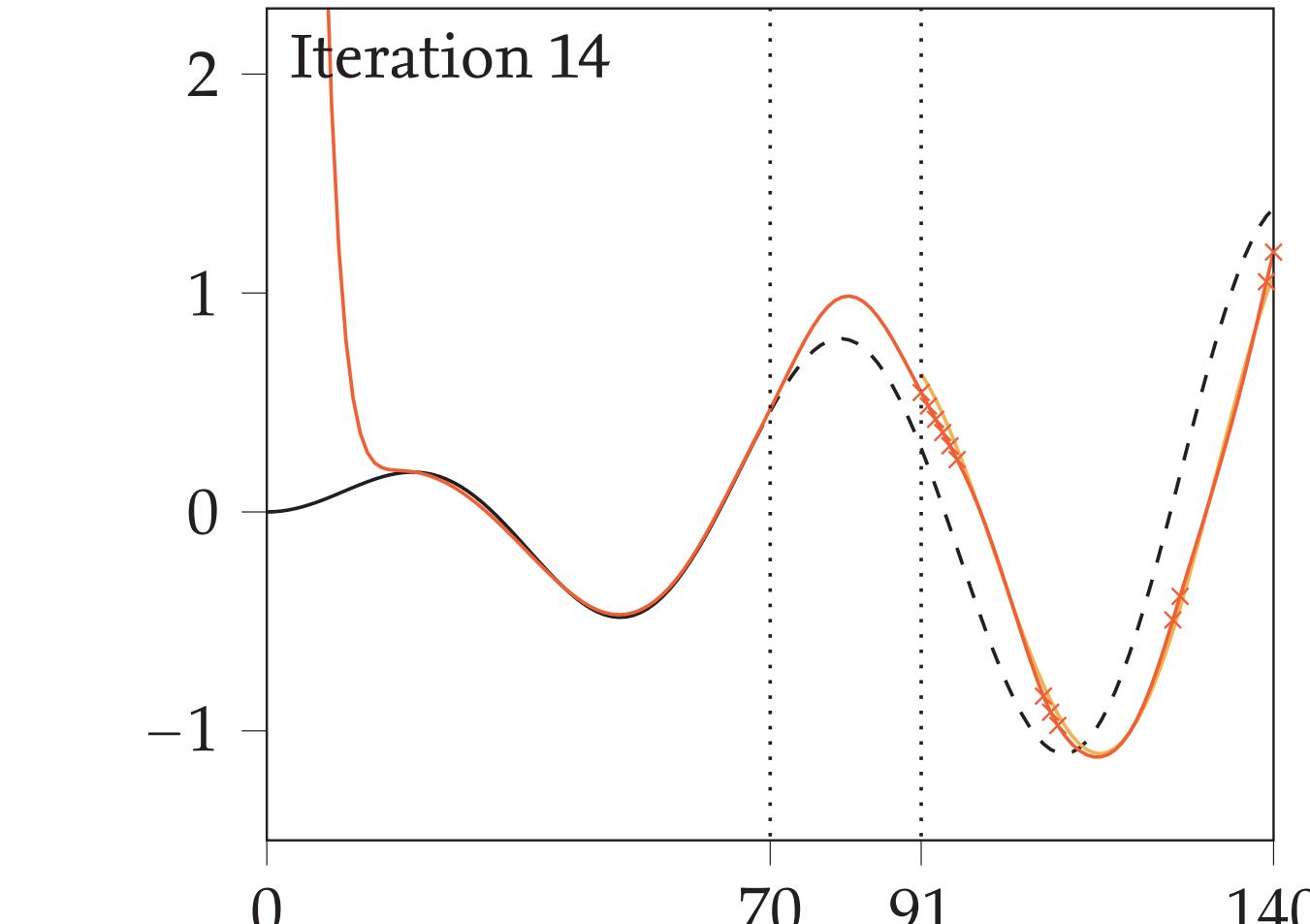
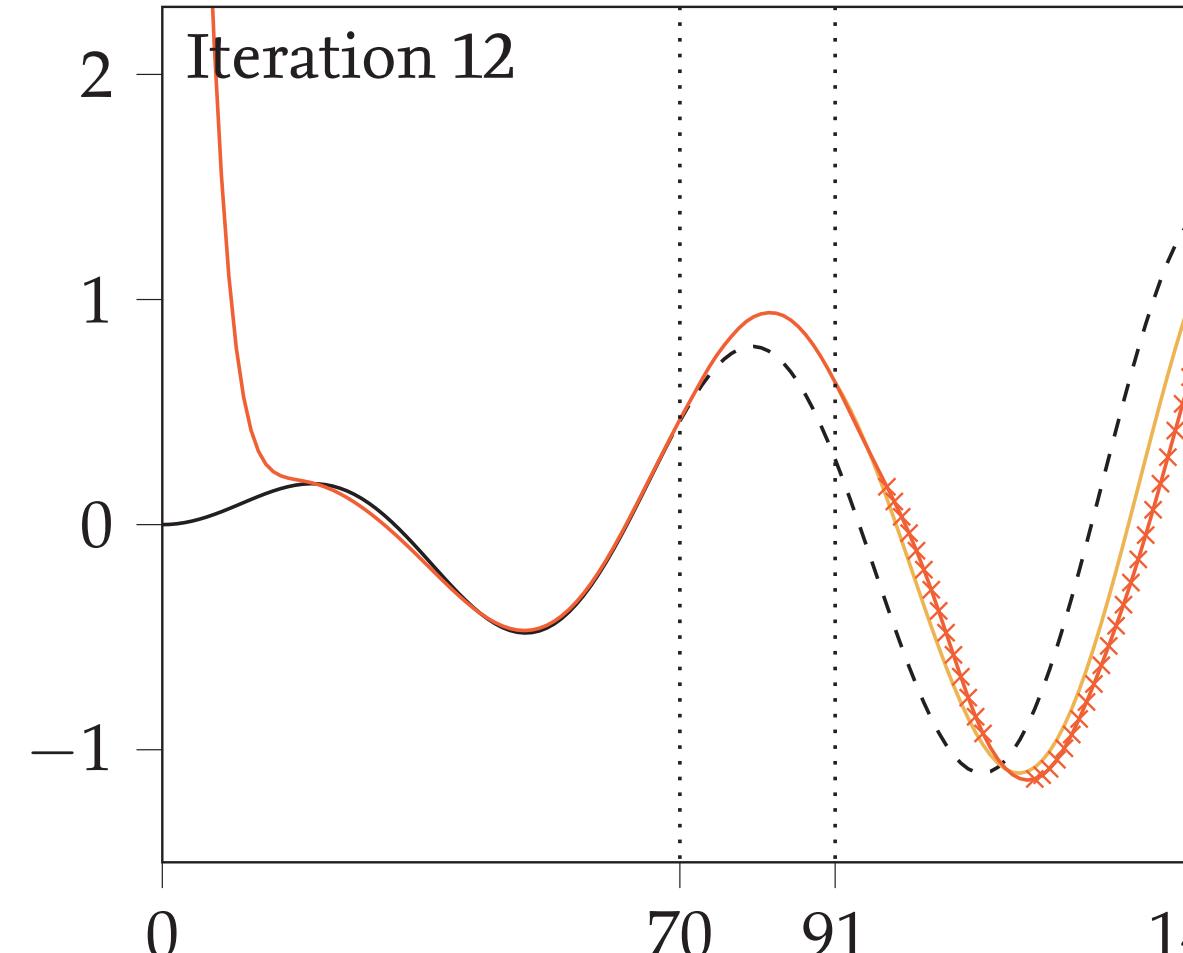
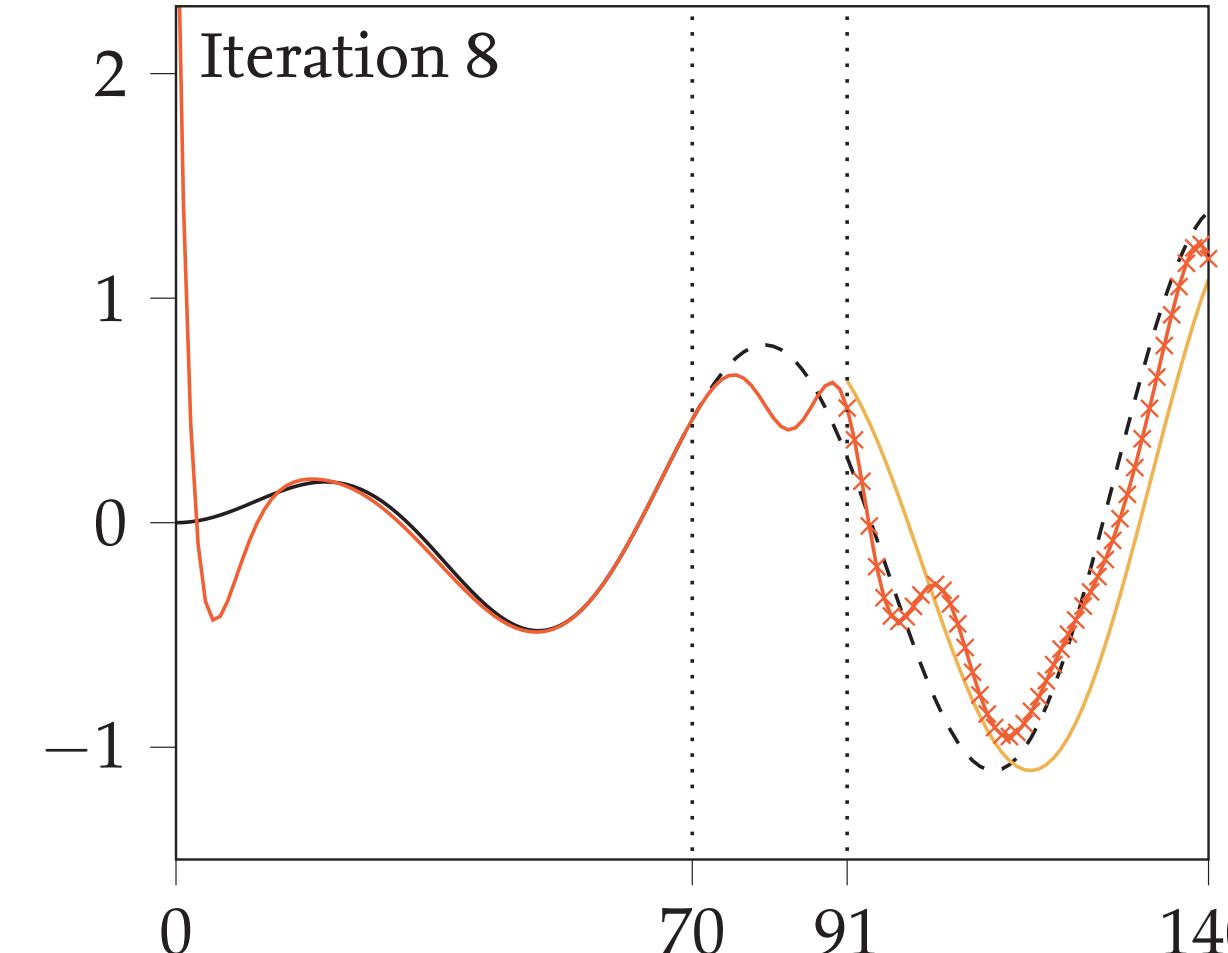
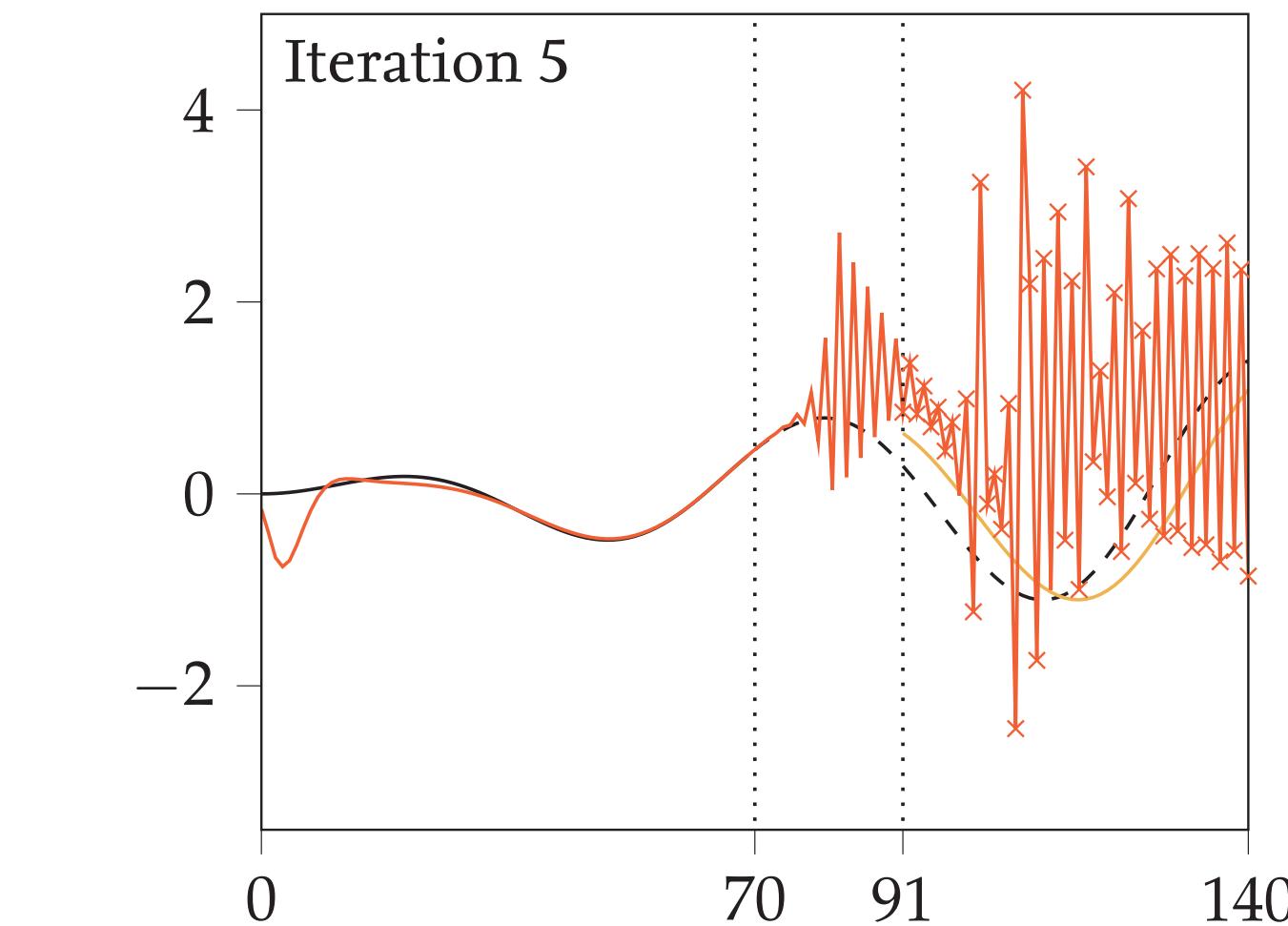
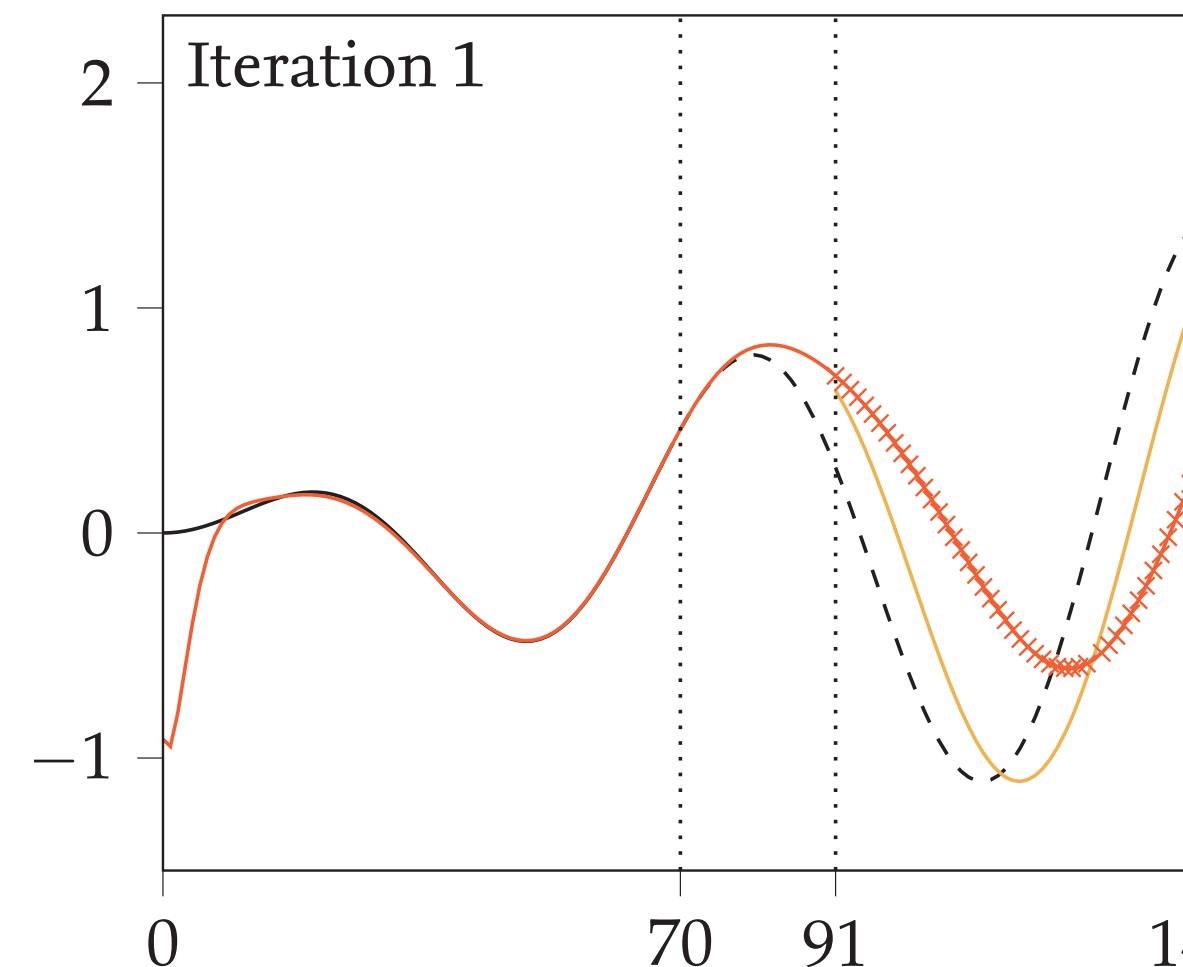
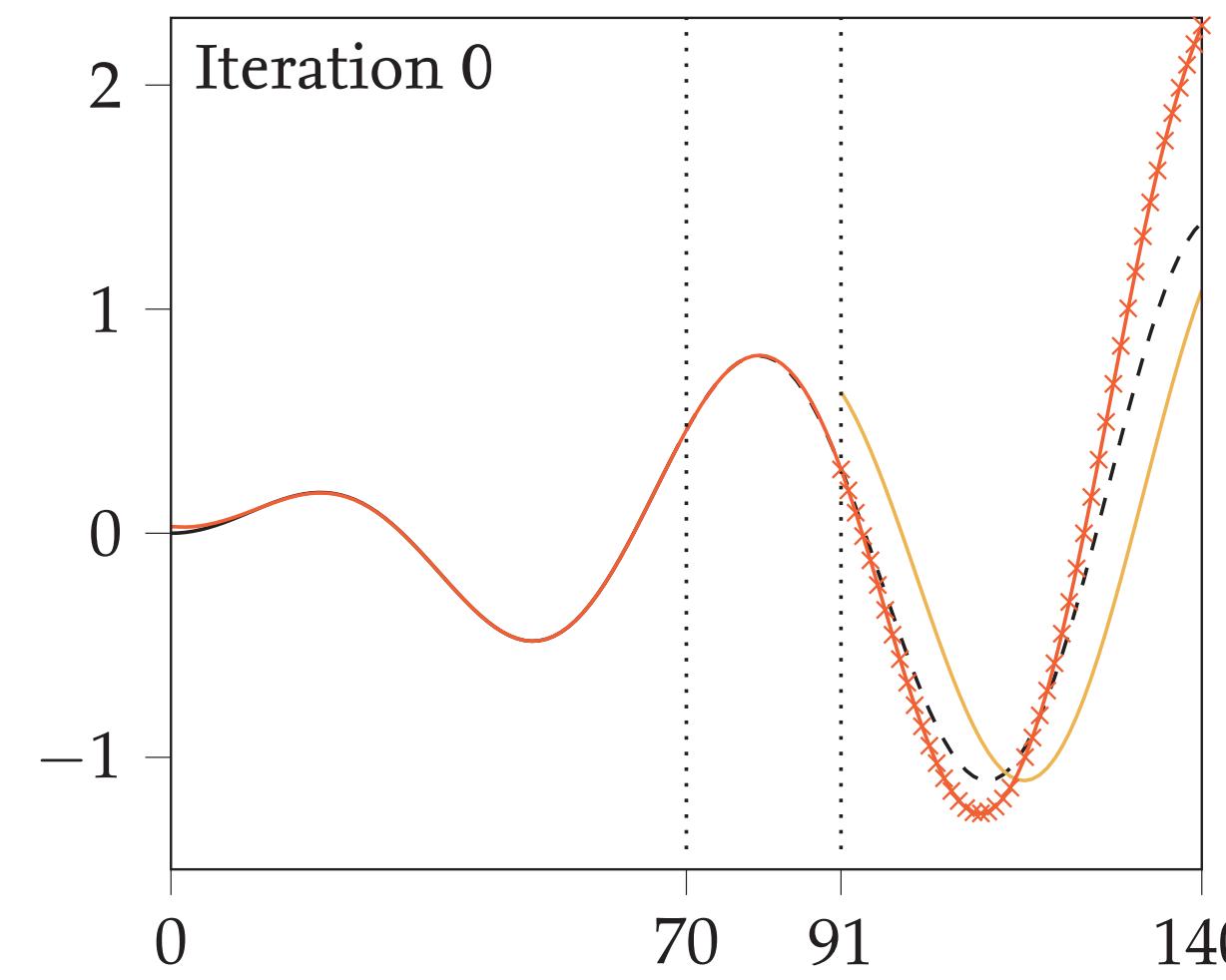
Prediction Examples (3/3)



Outline

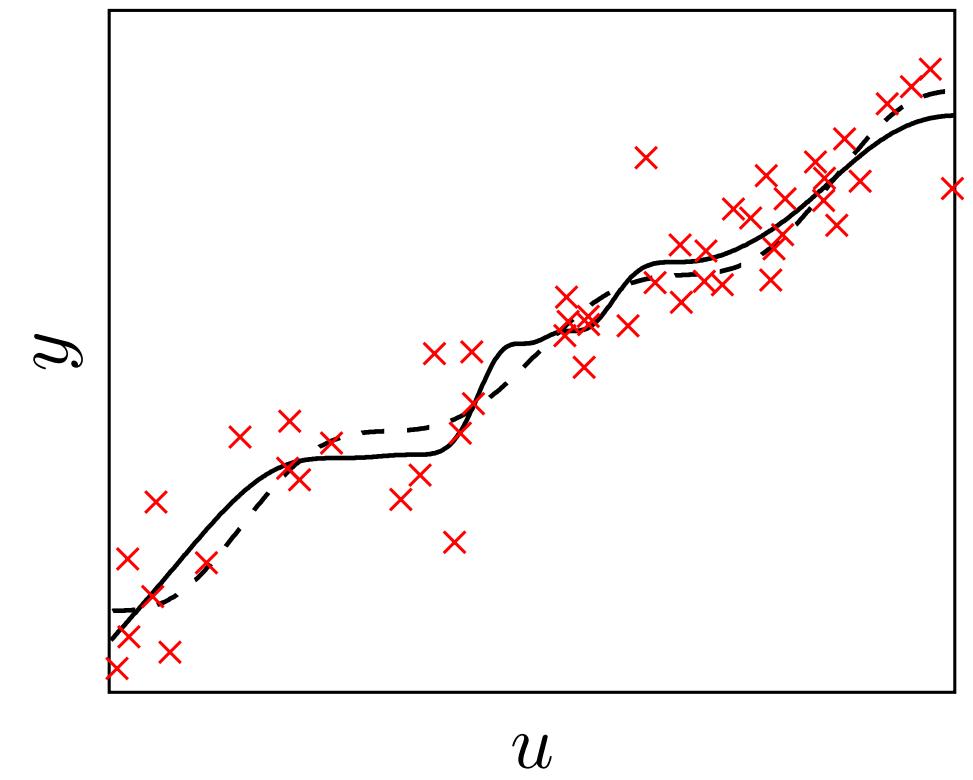
1. Constraint Definition
2. Embedding Constraints into the Neural Network
3. Example: Forecasting of Satellite Images
4. Future Work & Conclusion

Future Work: Recursive Predictions

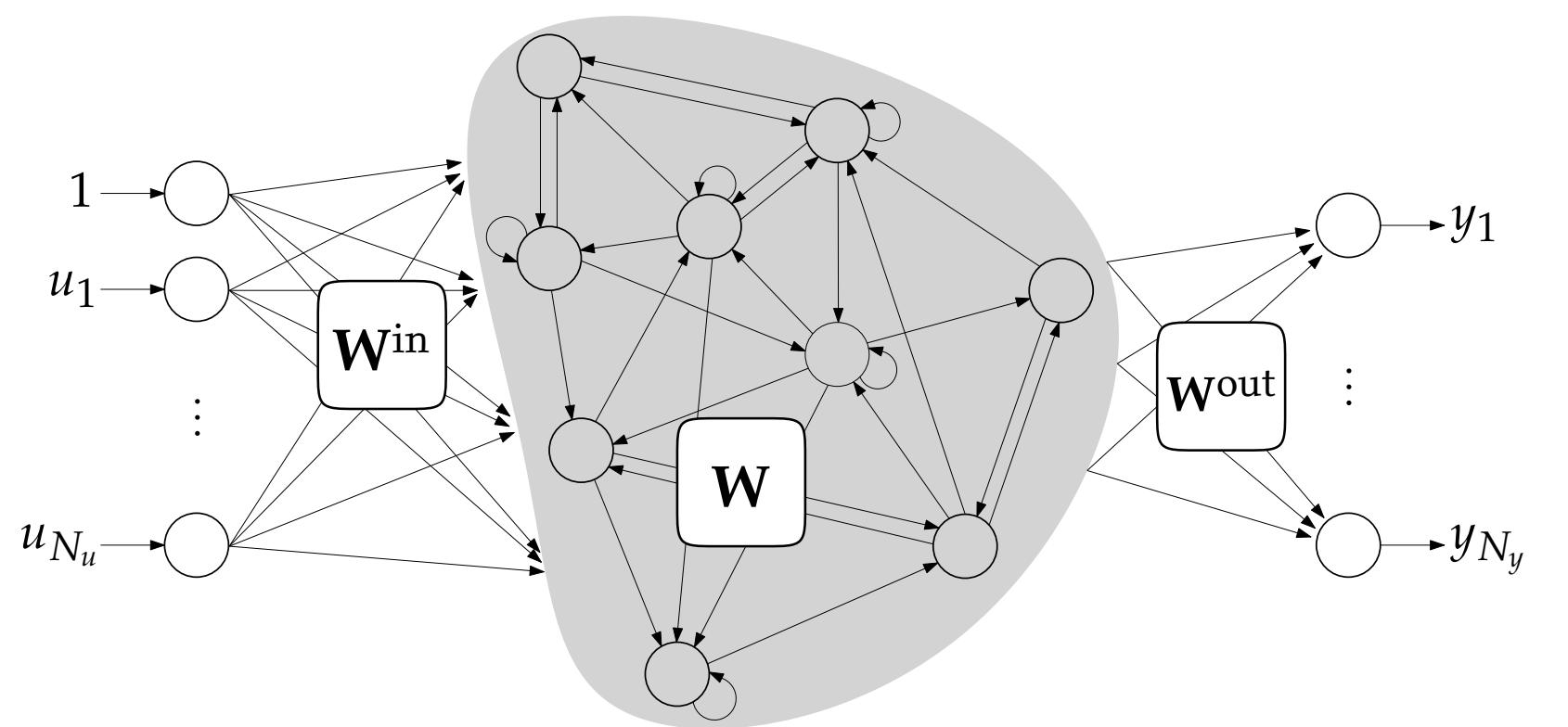
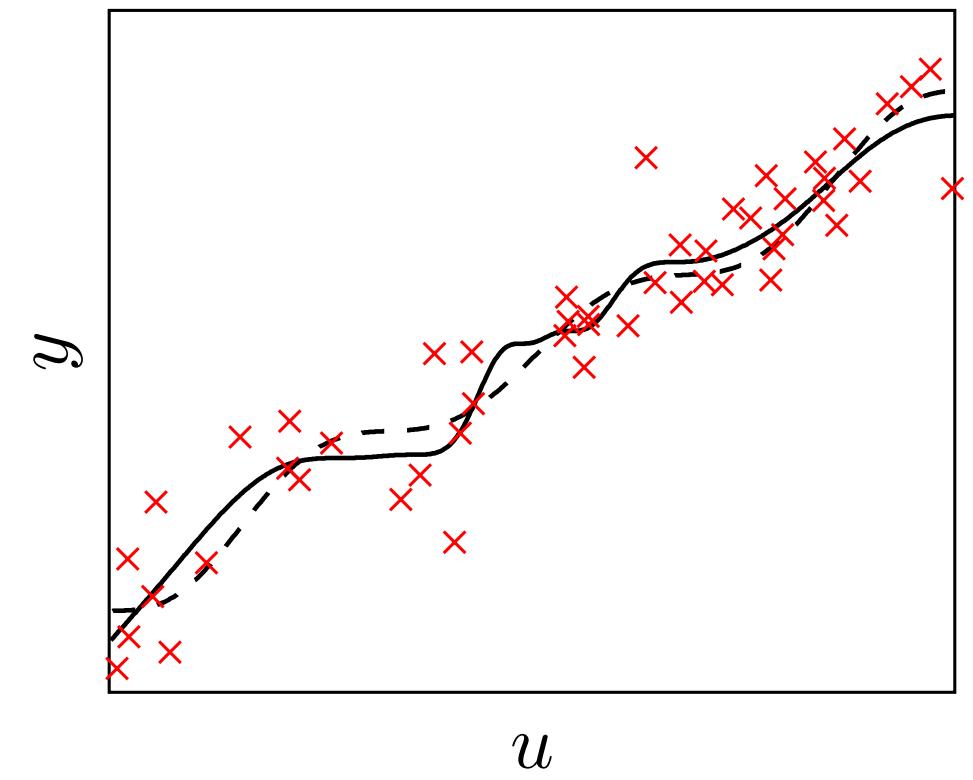


Conclusion

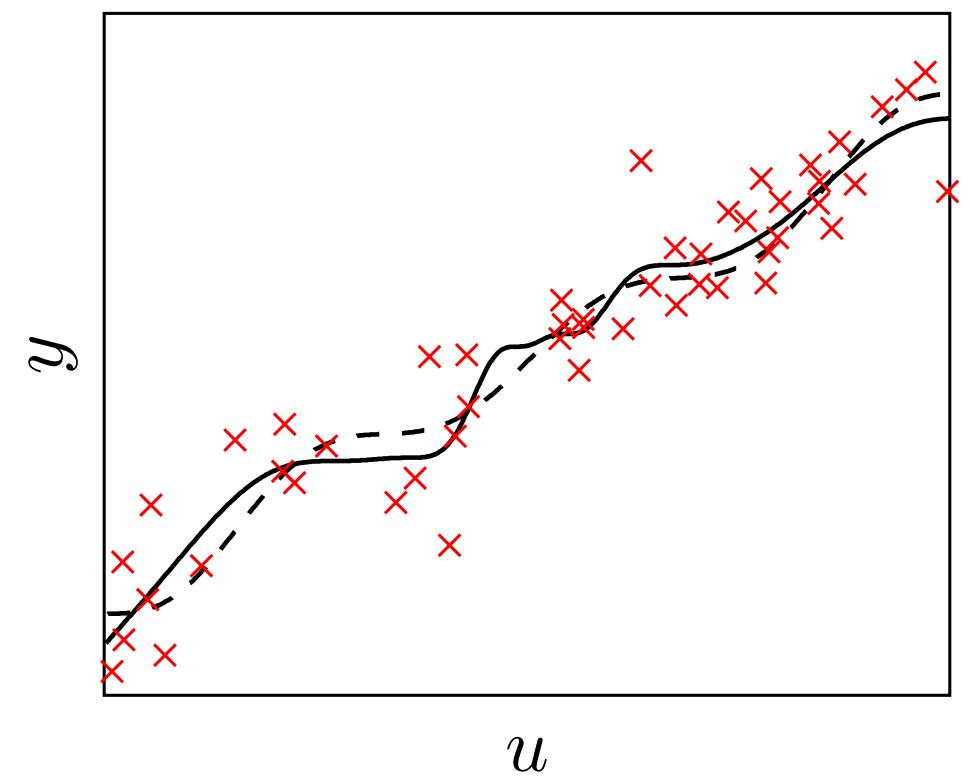
Conclusion



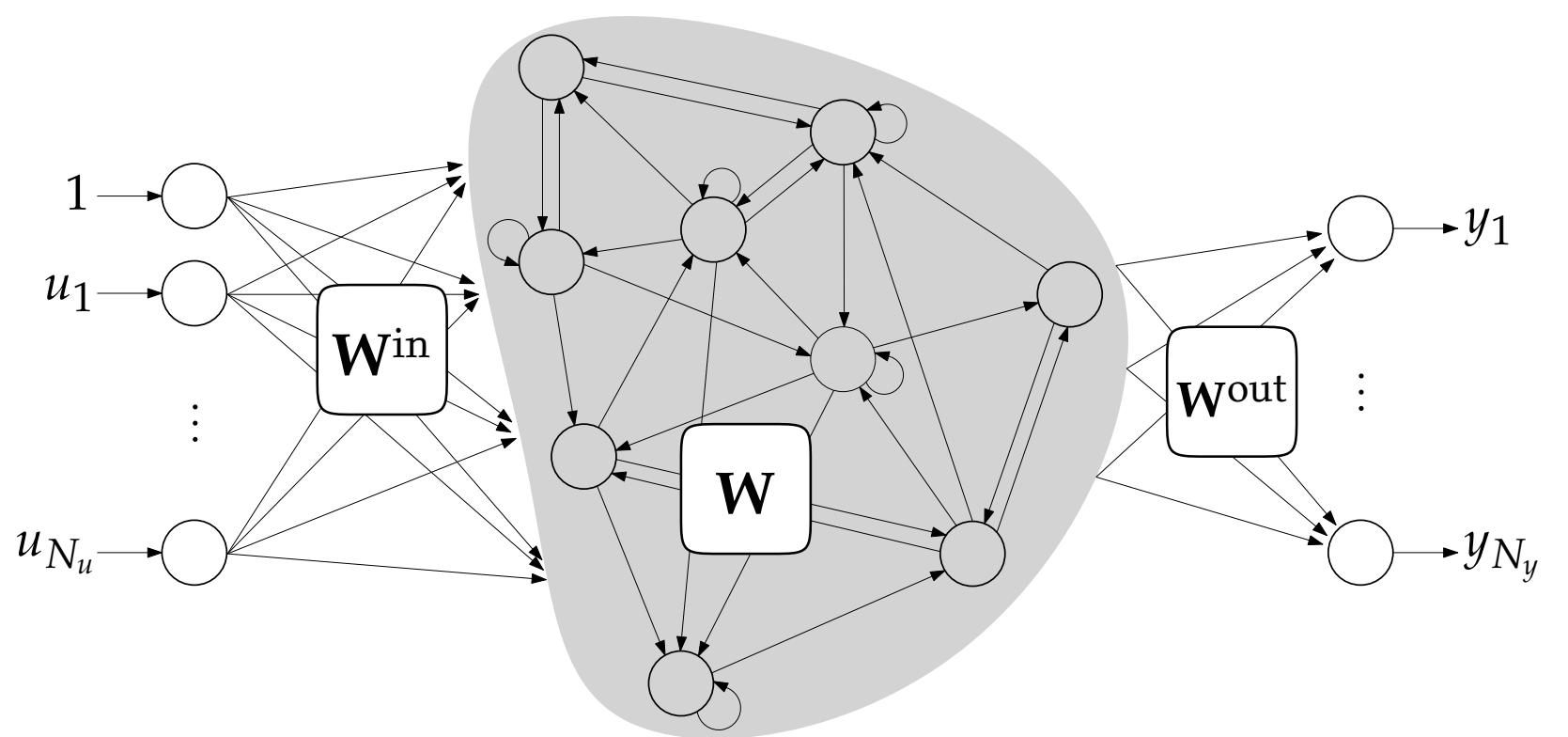
Conclusion



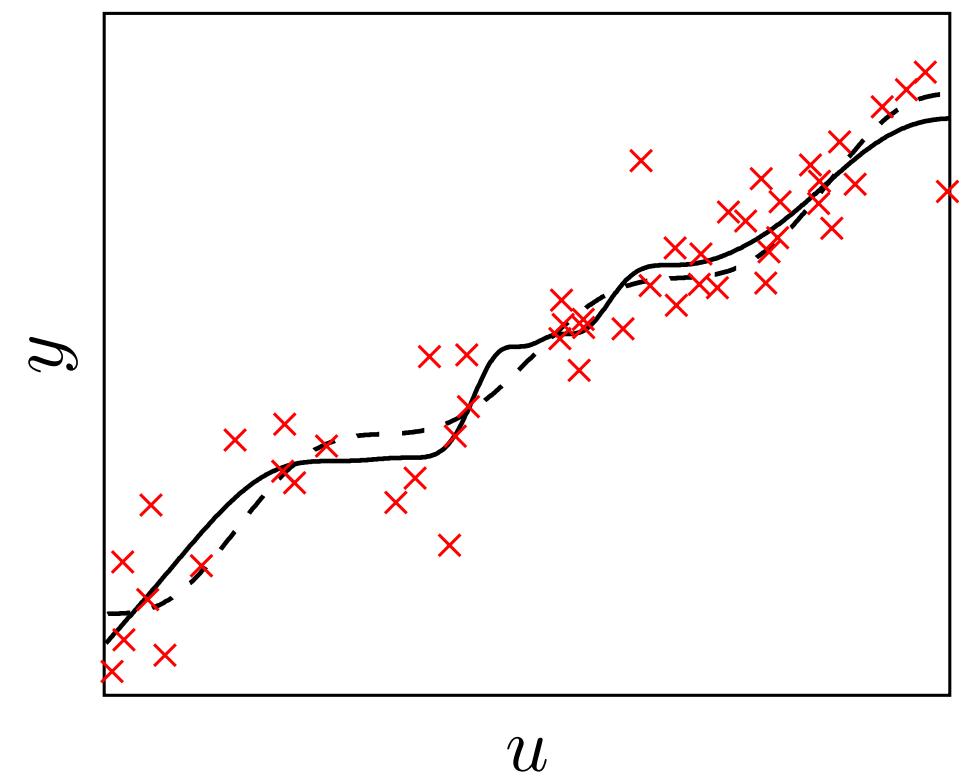
Conclusion



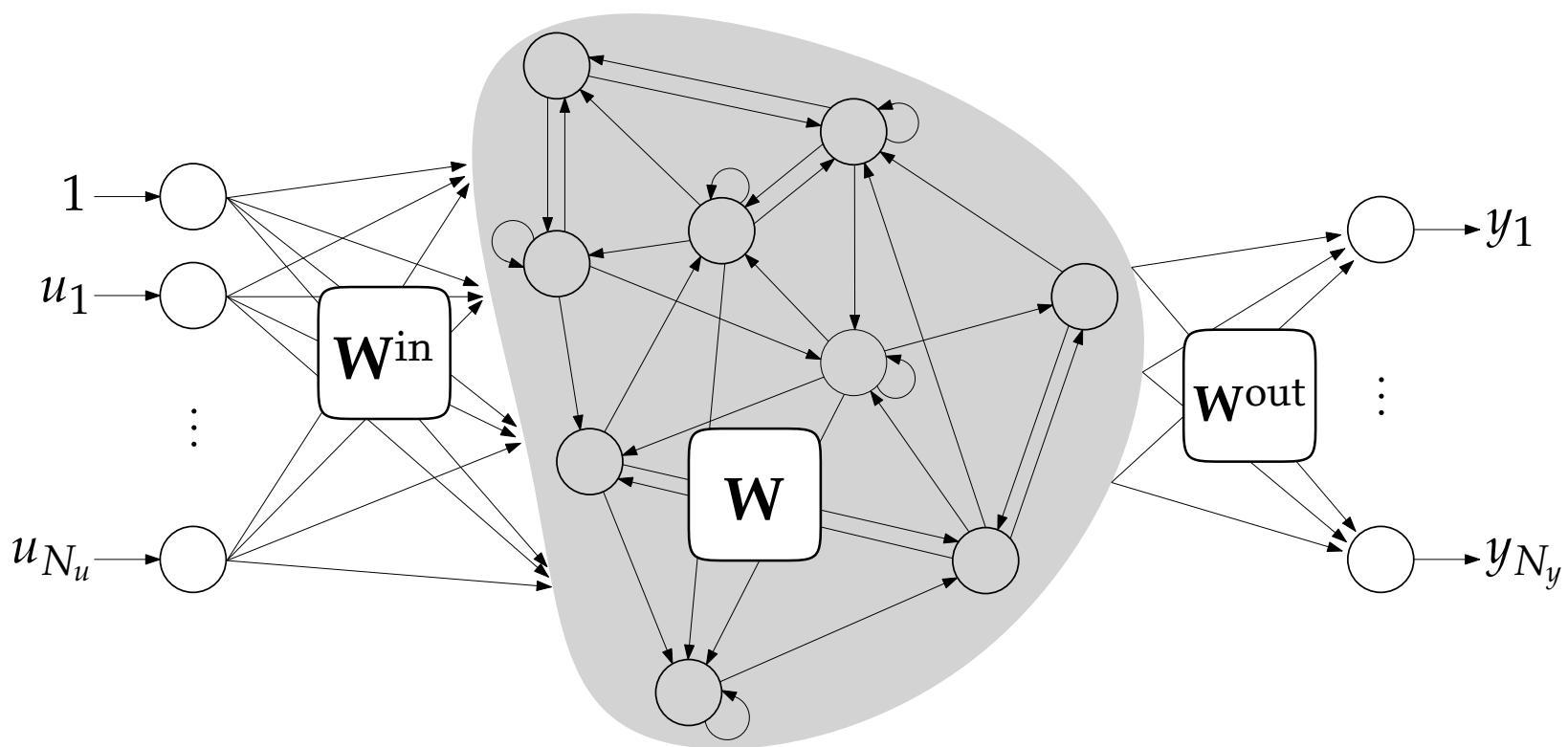
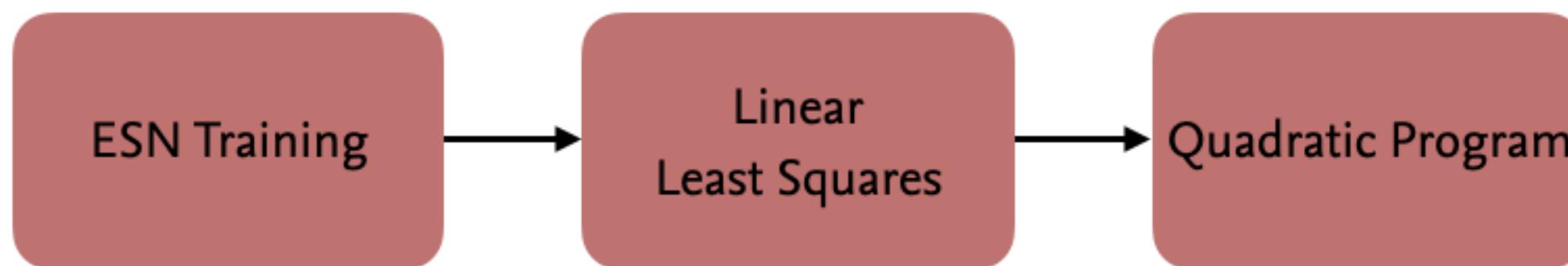
$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$



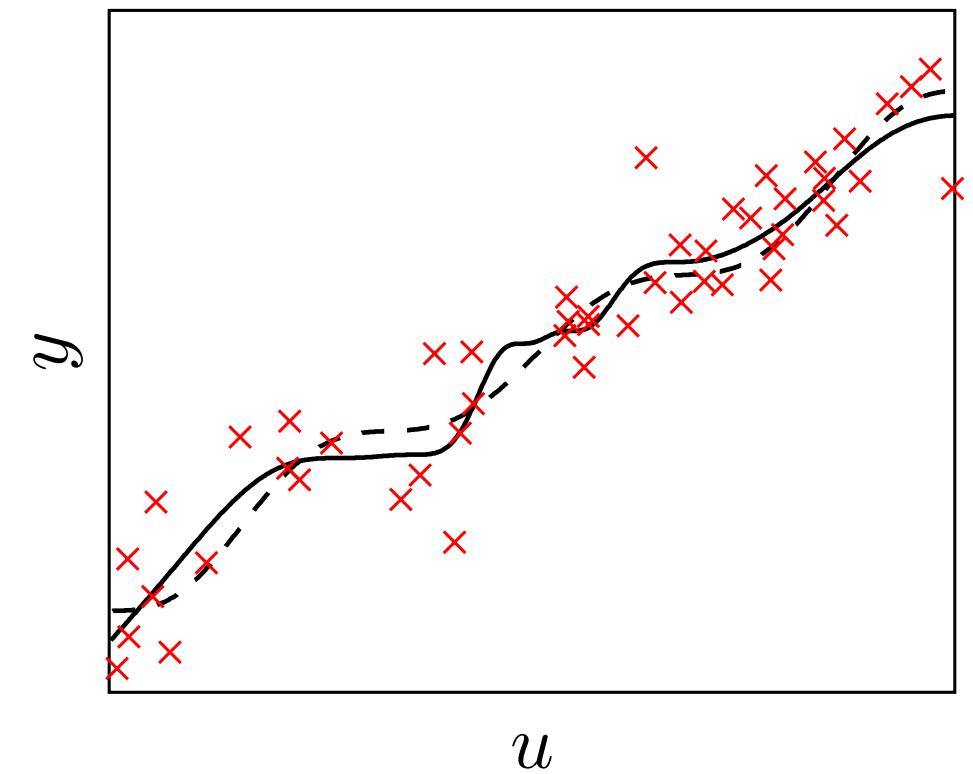
Conclusion



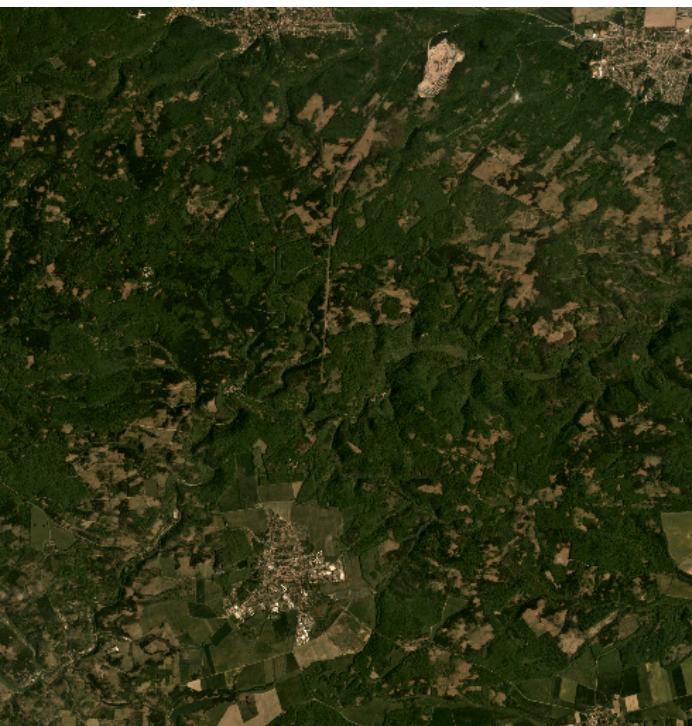
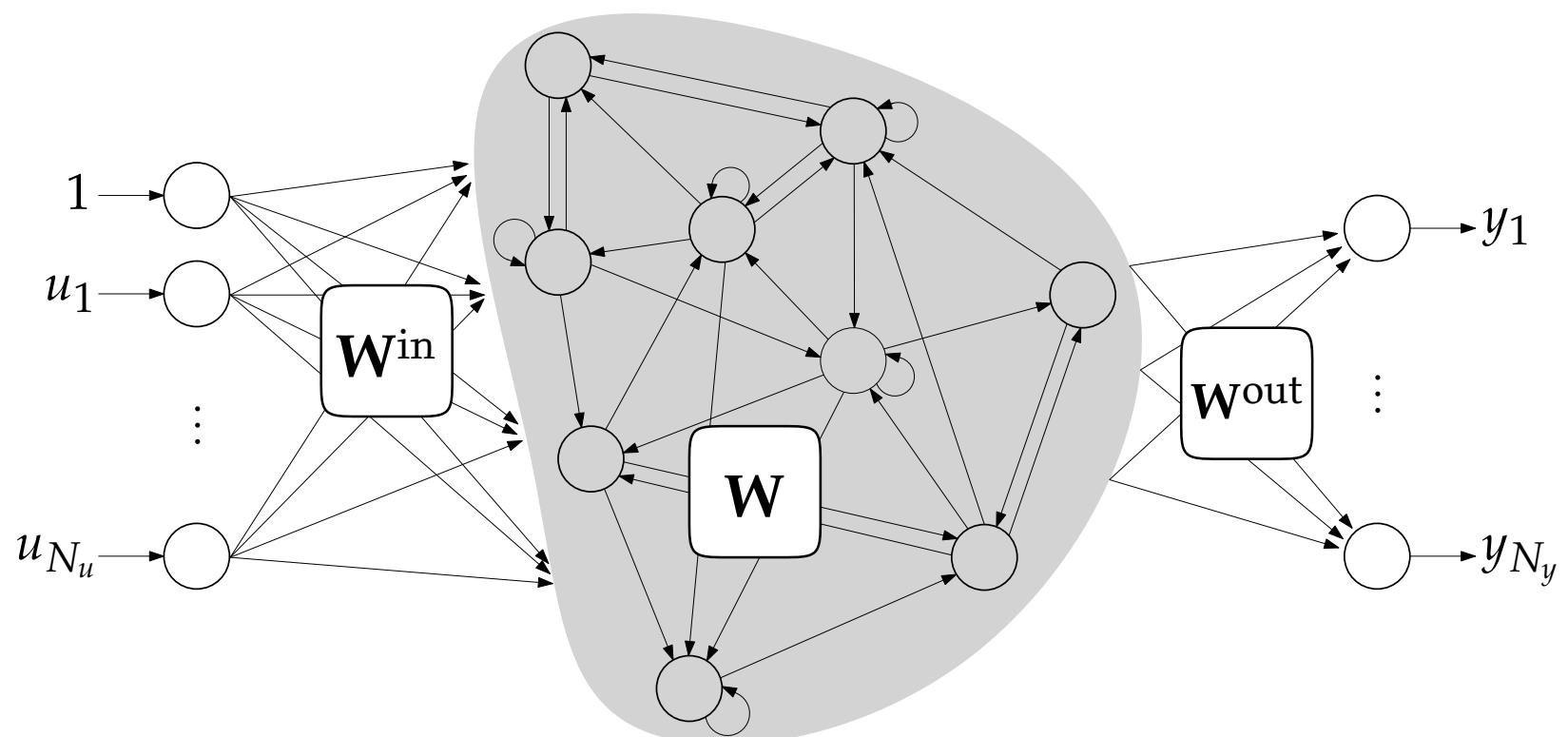
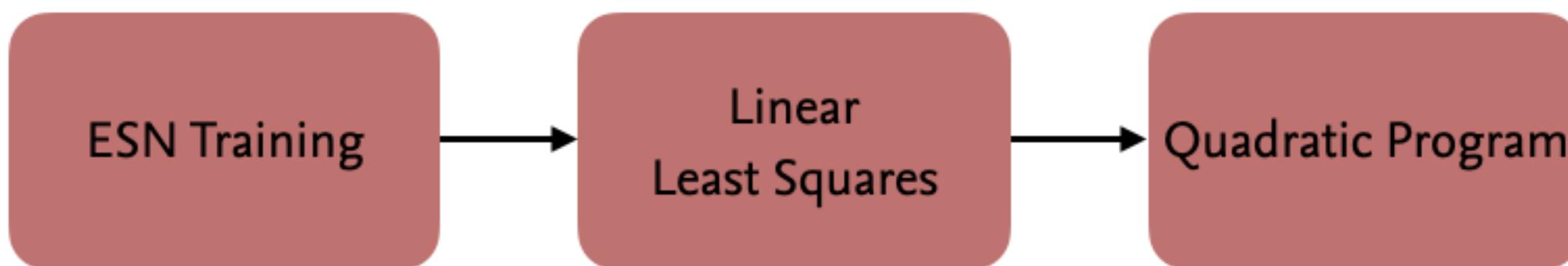
$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$



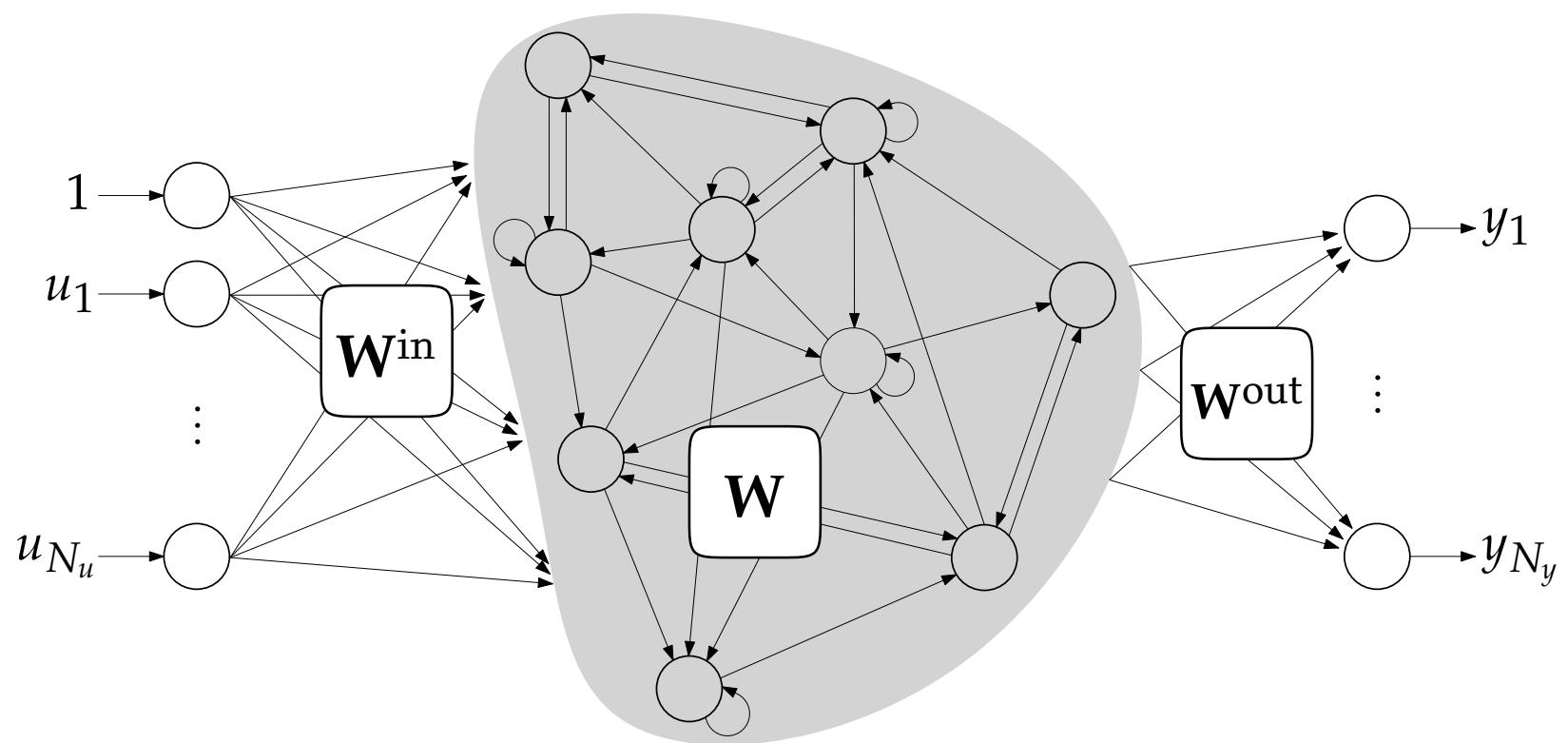
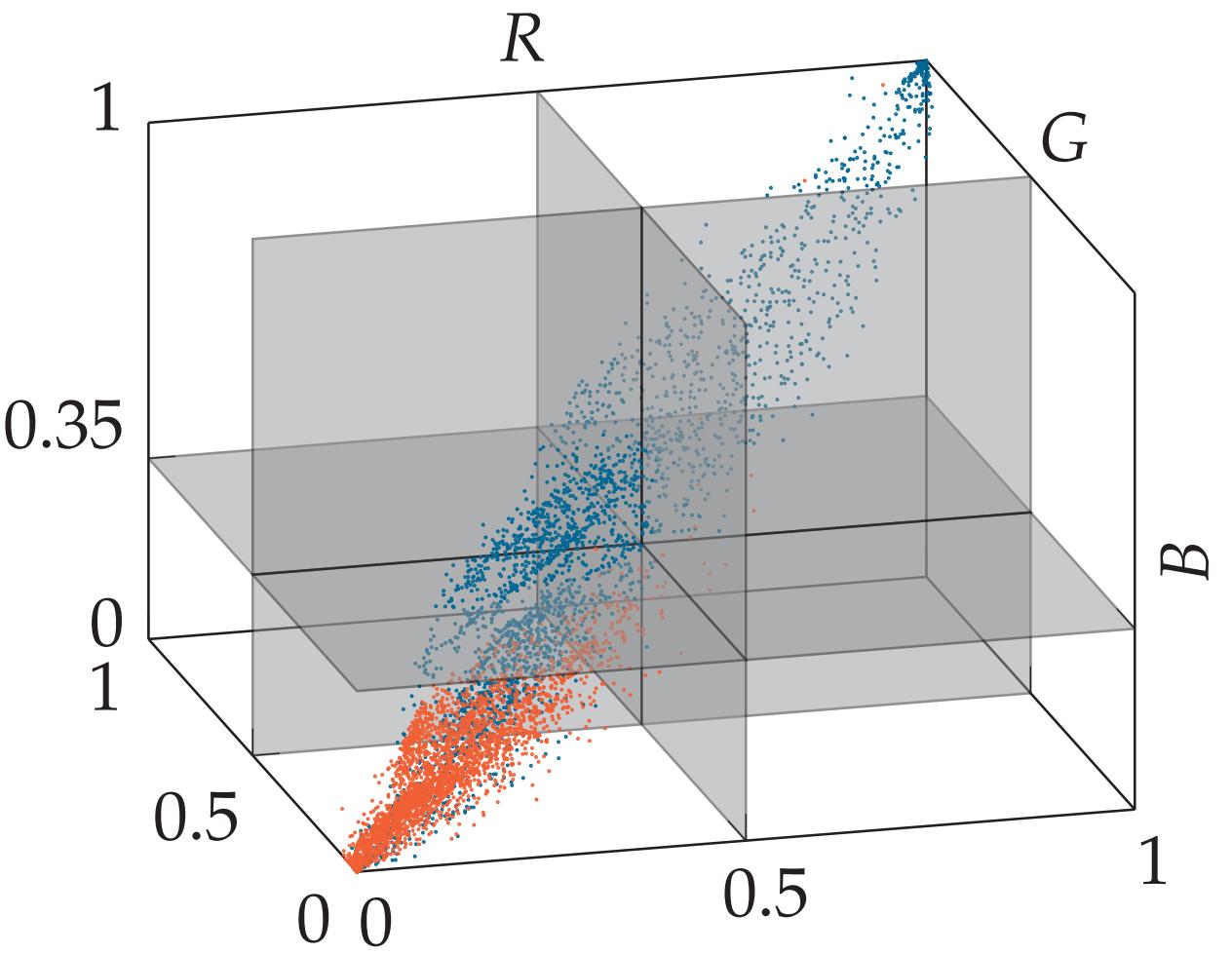
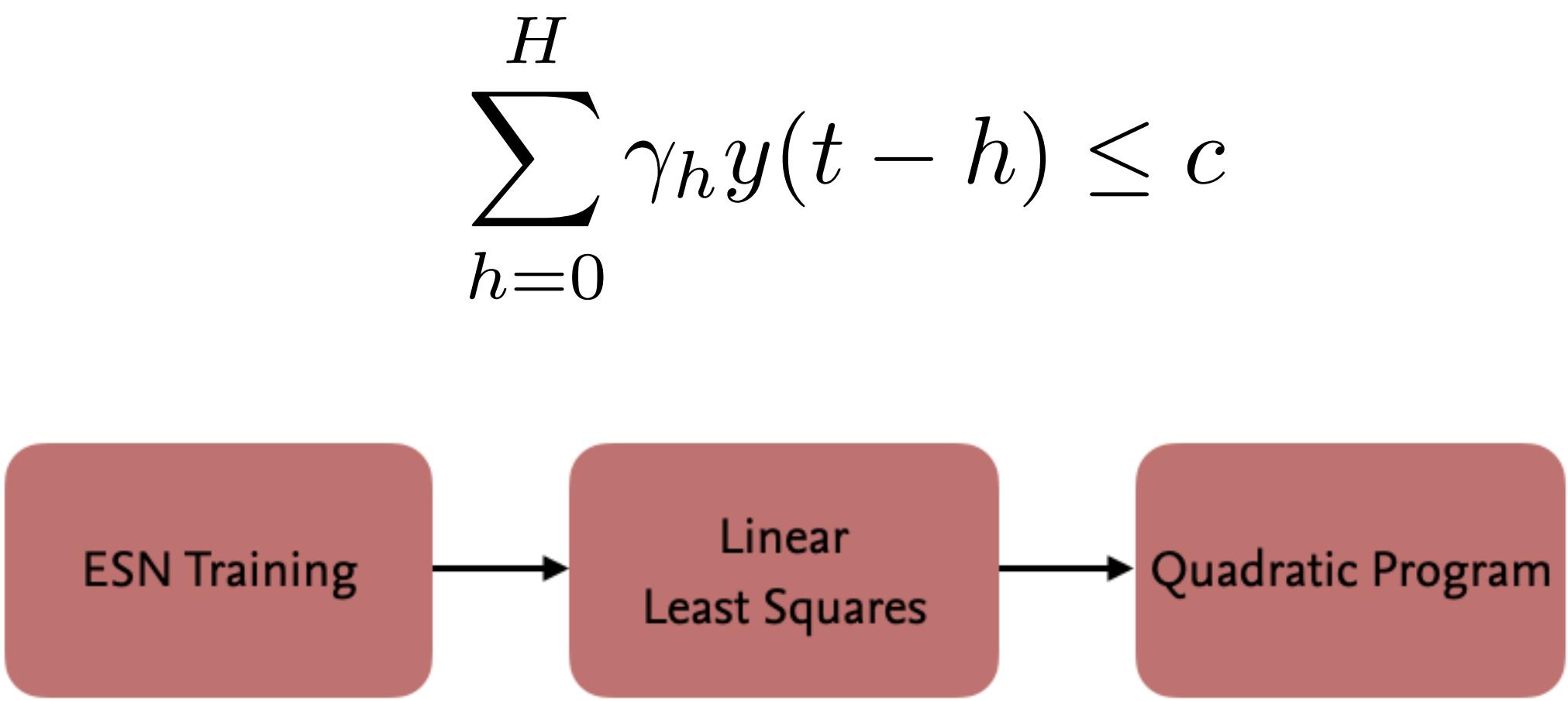
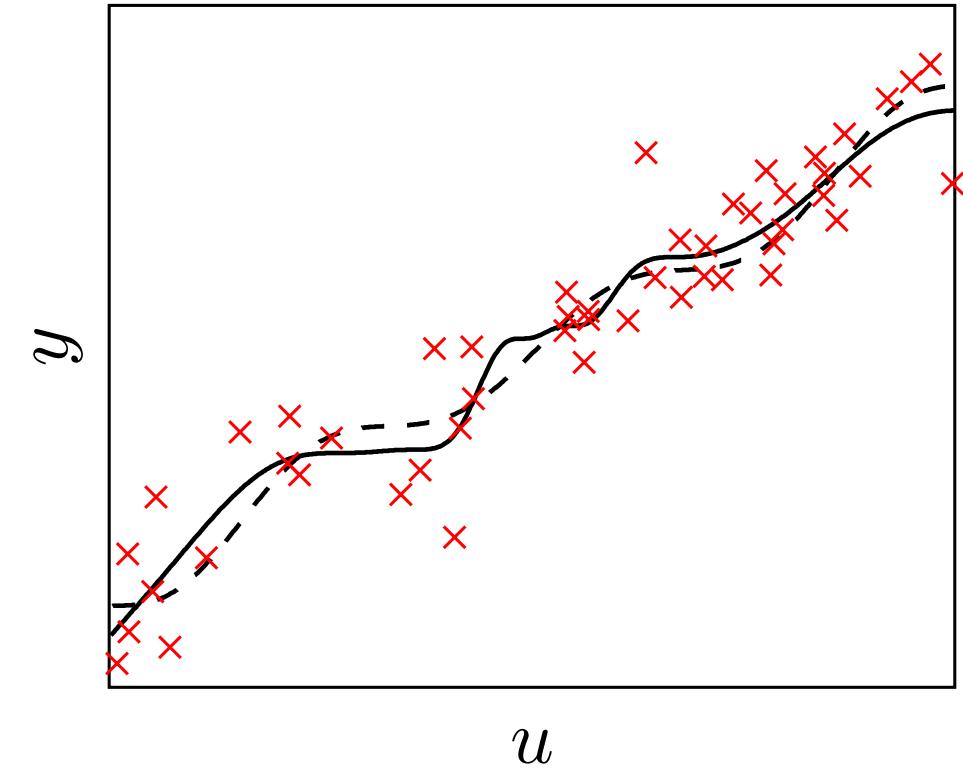
Conclusion



$$\sum_{h=0}^H \gamma_h y(t-h) \leq c$$



Conclusion



Conclusion

