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

JOCHEN JAKOB STEIL

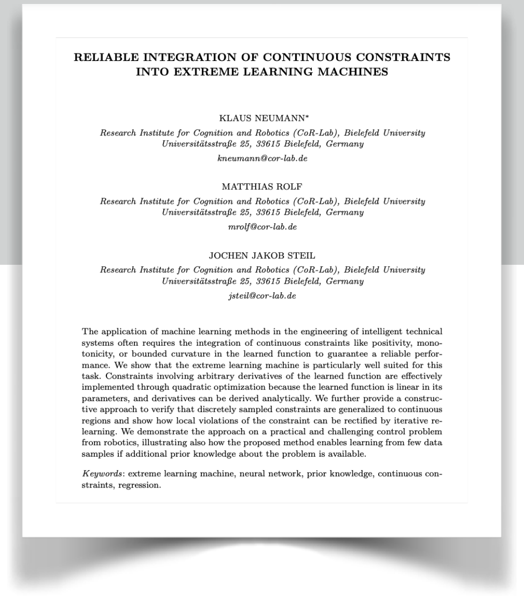
Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany

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

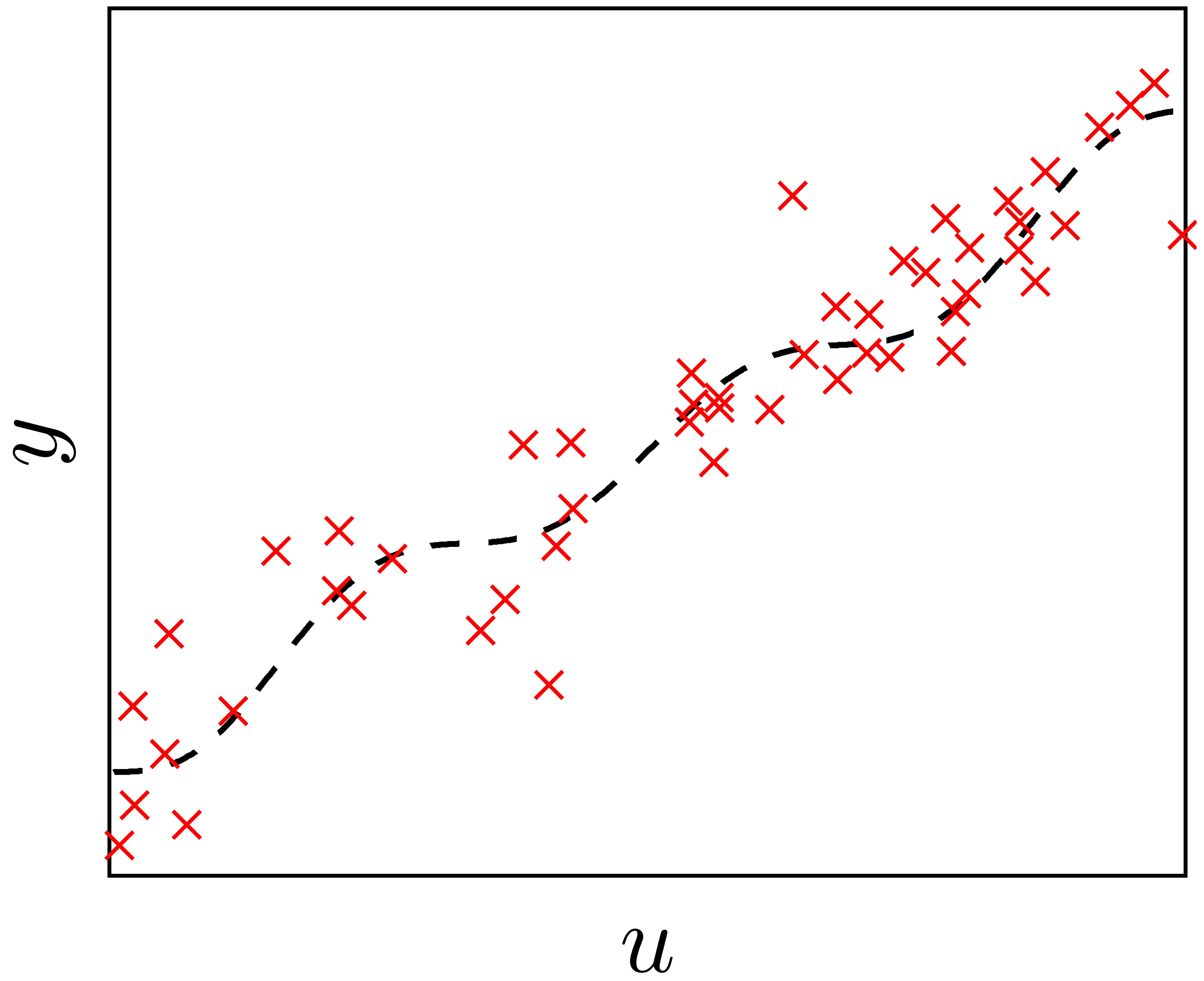
KLAUS NEUMANN*
Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
kneumann@cor-lab.de

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

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

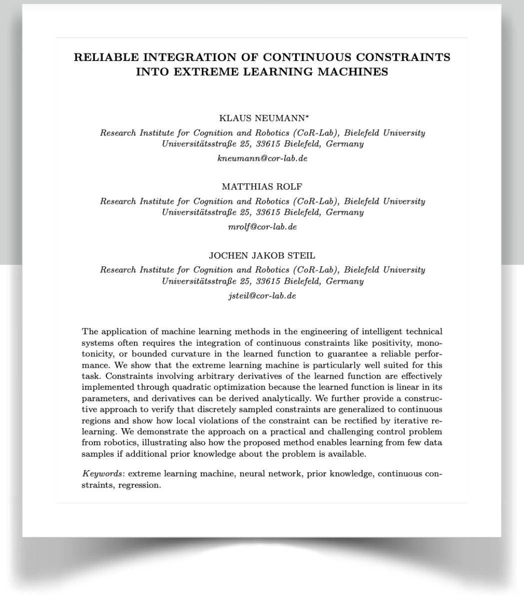
The application of machine learning methods in the engineering of intelligent technical systems often requires the integration of continuous constraints like positivity, monotonicity, or bounded curvature in the learned function to guarantee a reliable performance. We show that the extreme learning machine 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 relearning. 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.



Adapted from Neumann (2013)

Context: Neumann, Rolf and Steil (2012)



RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS INTO EXTREME LEARNING MACHINES

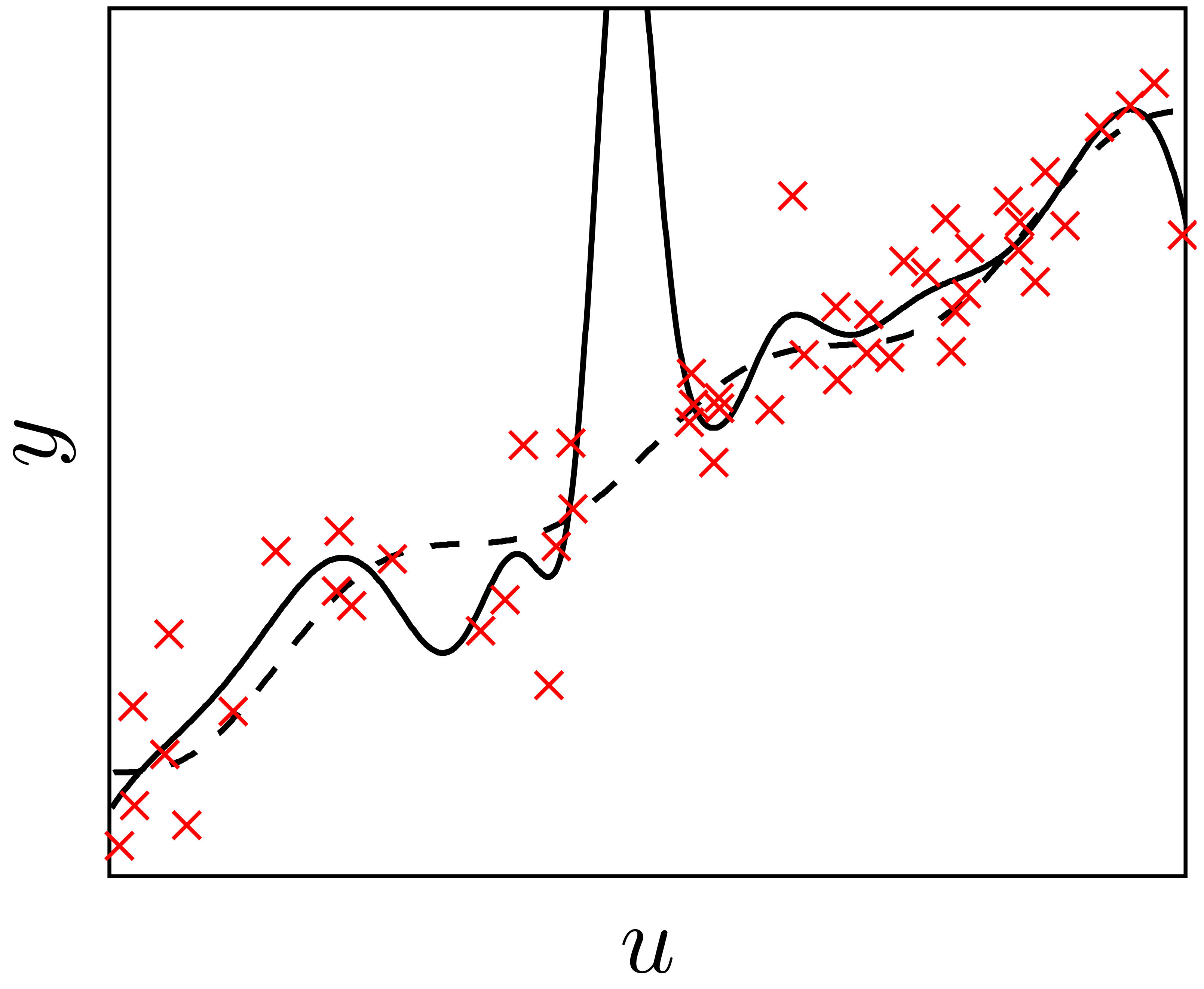
KLAUS NEUMANN*
Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
kneumann@cor-lab.de

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

JOCHEN 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.



Adapted from Neumann (2013)

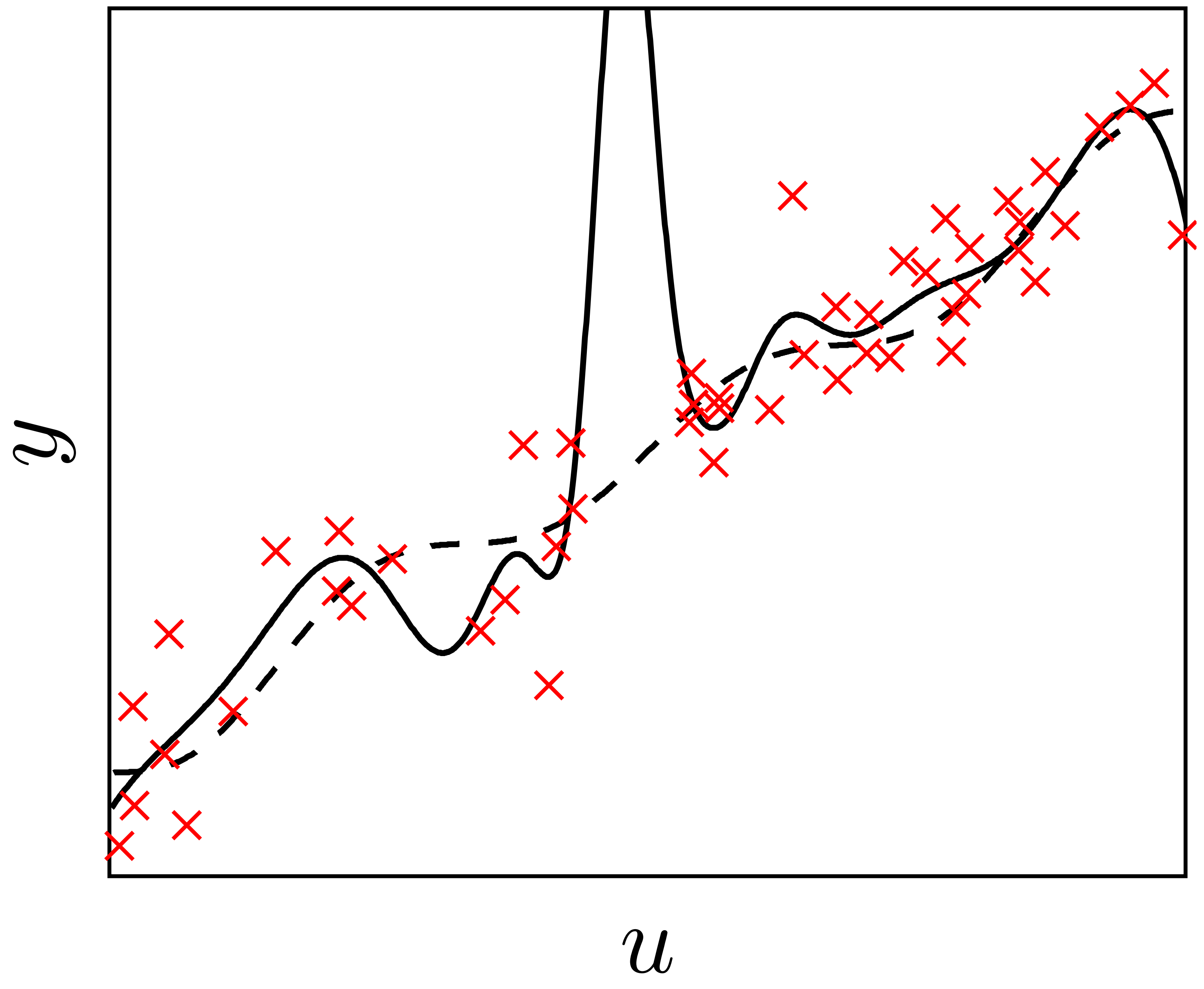
Context: Neumann, Rolf and Steil (2012)


RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS INTO EXTREME LEARNING MACHINES

KLAUS NEUMANN

MATTHIAS ROLF

JOCHEN JAKOB STEIL

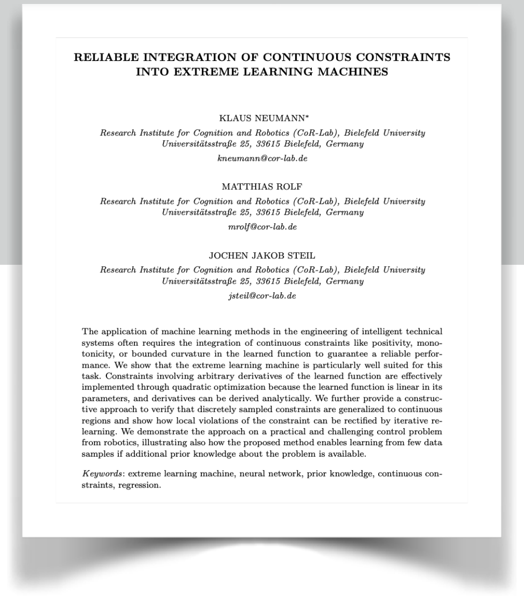


Target function steadily increasing:

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

Adapted from Neumann (2013)

Context: Neumann, Rolf and Steil (2012)


RELIABLE INTEGRATION OF CONTINUOUS CONSTRAINTS INTO EXTREME LEARNING MACHINES

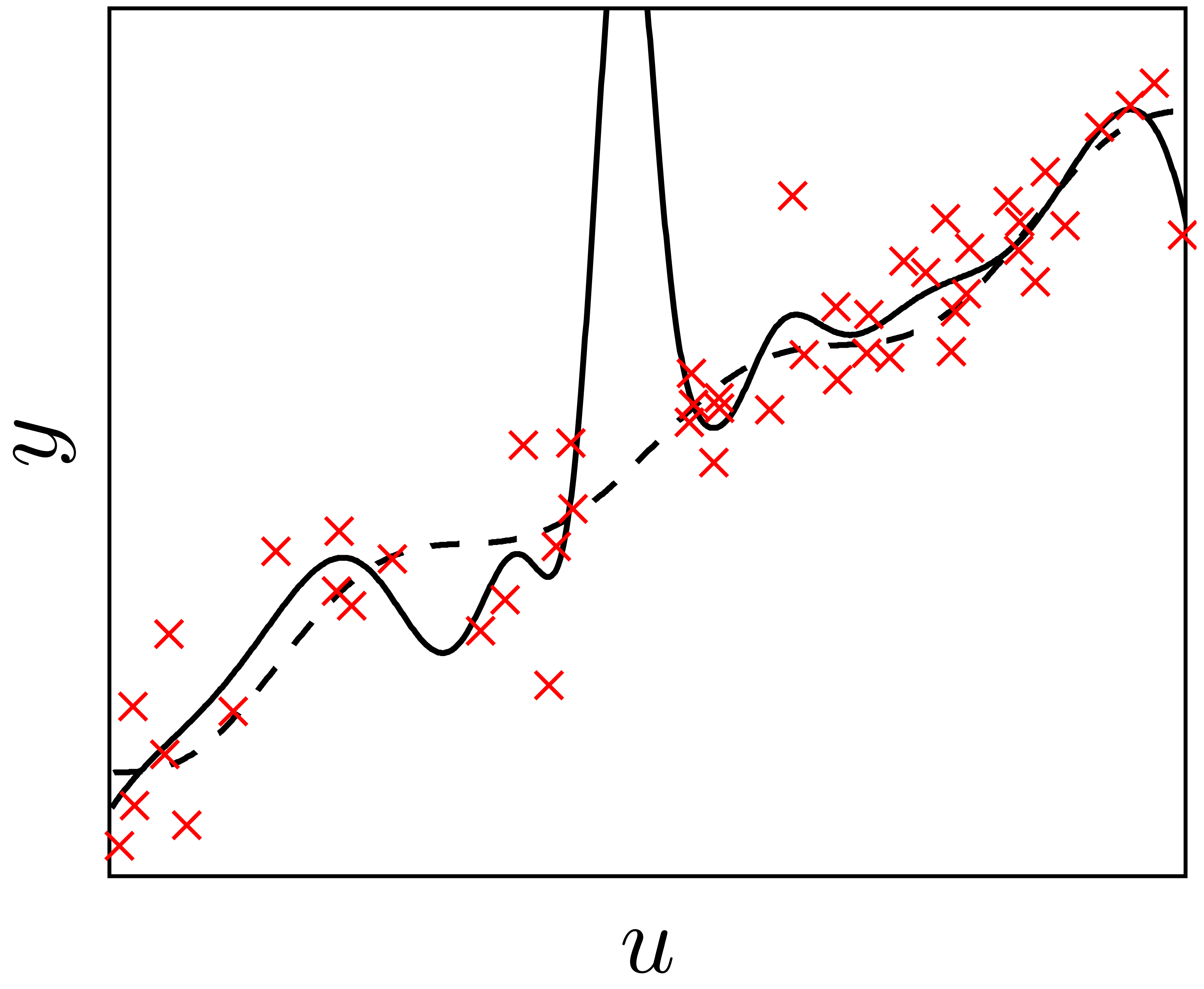
KLAUS NEUMANN*
Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
kneumann@cor-lab.de

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

JOCHEN 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 relearning. 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.



Target function steadily increasing:

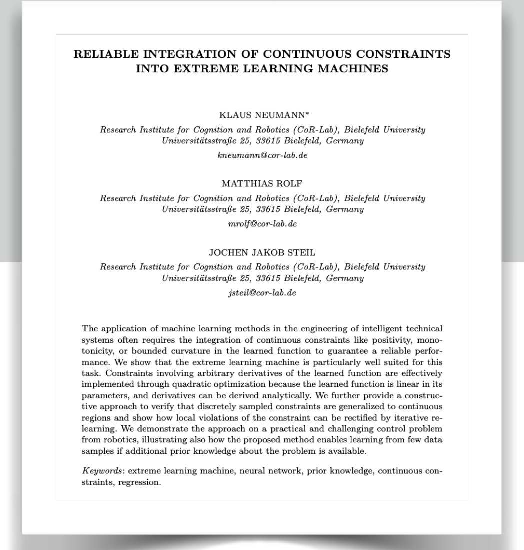
$$u_1 \leq u_2 \Rightarrow y(u_1) \leq y(u_2)$$

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

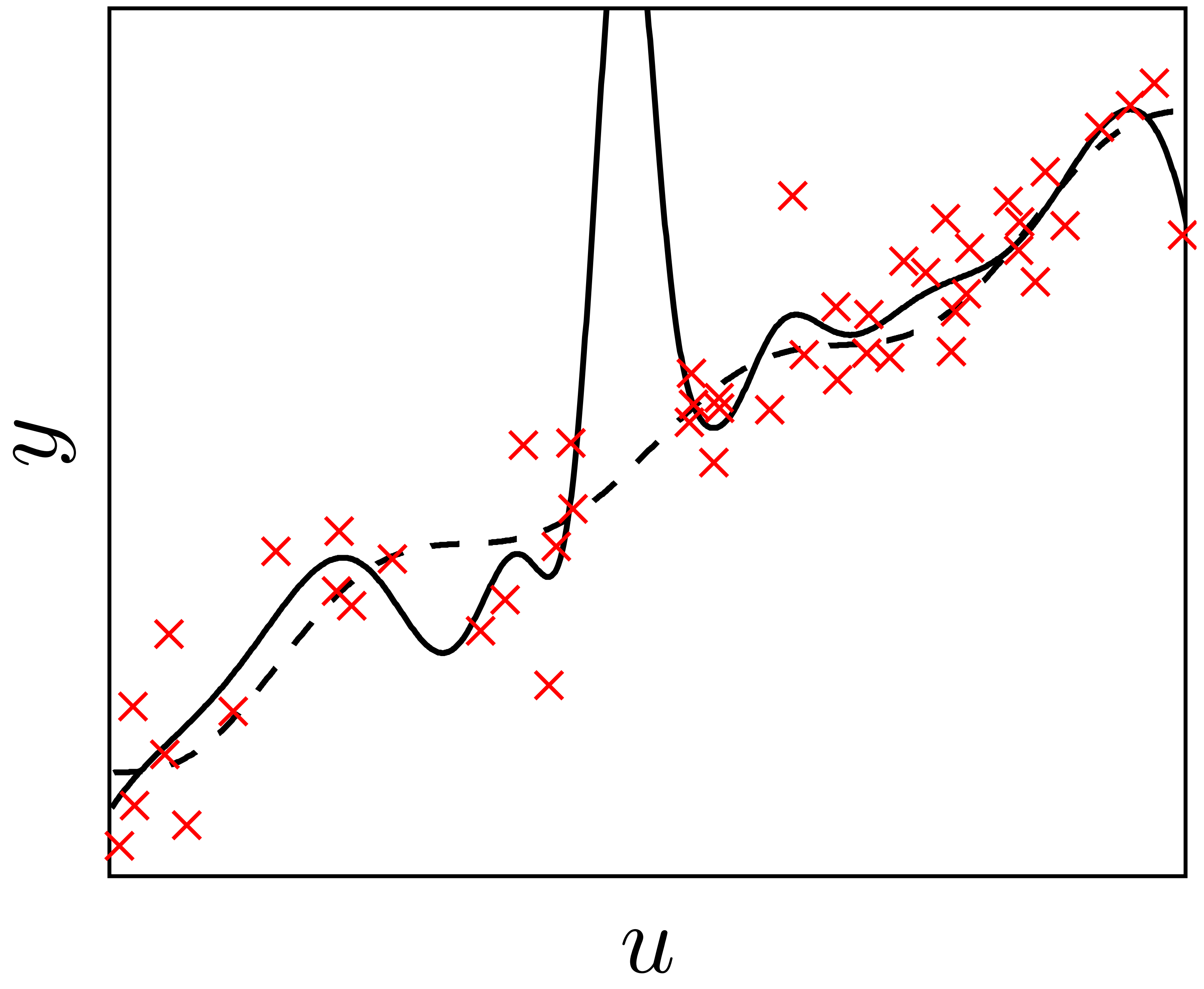
KLAUS NEUMANN*
Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
kneumann@cor-lab.de

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

JOCHEN 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 relearning. 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.



Target function steadily increasing:

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

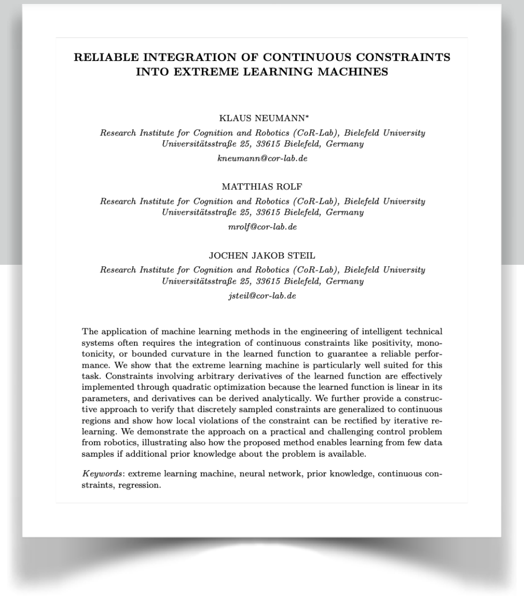
or

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

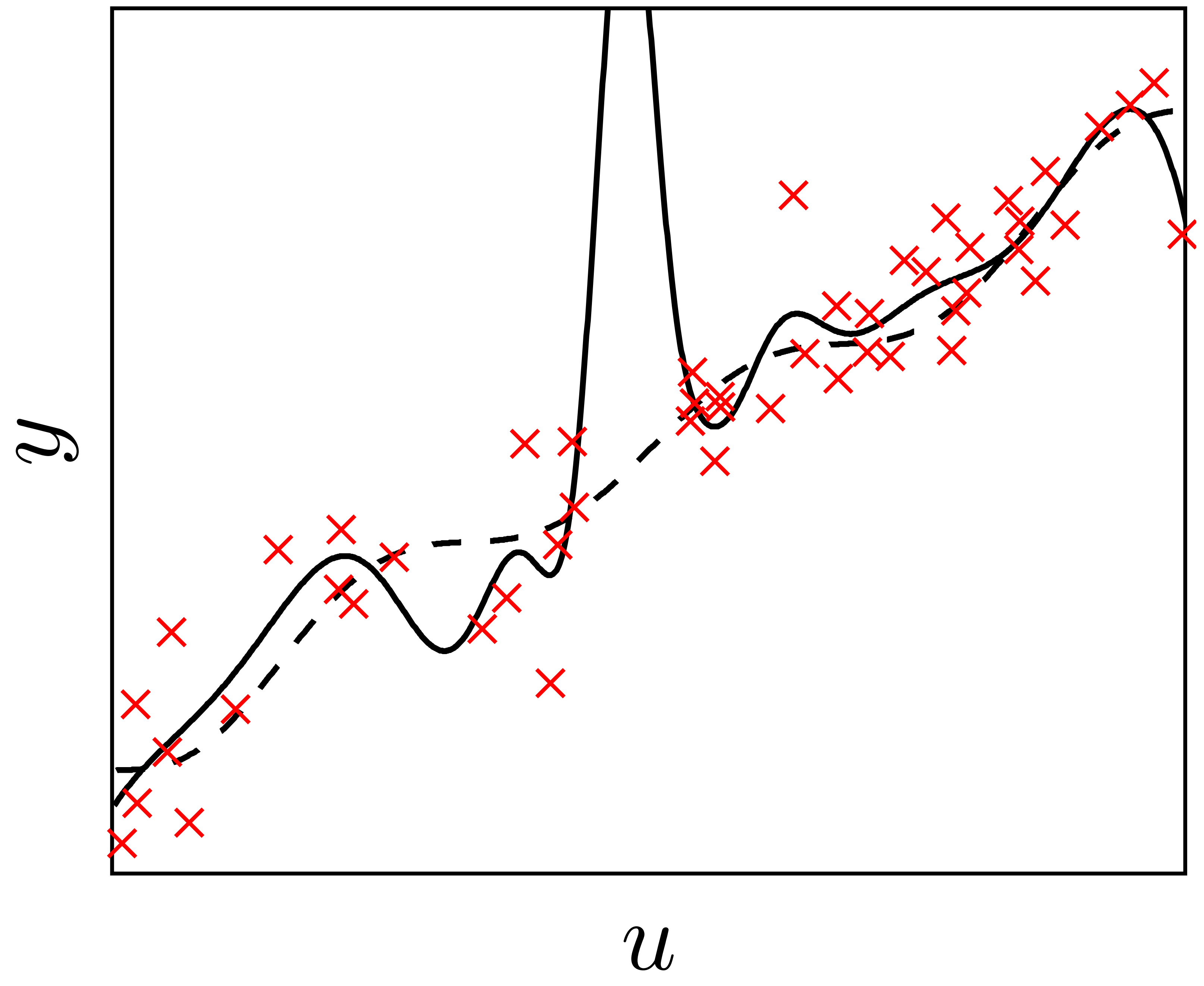
Constraint

Adapted from Neumann (2013)

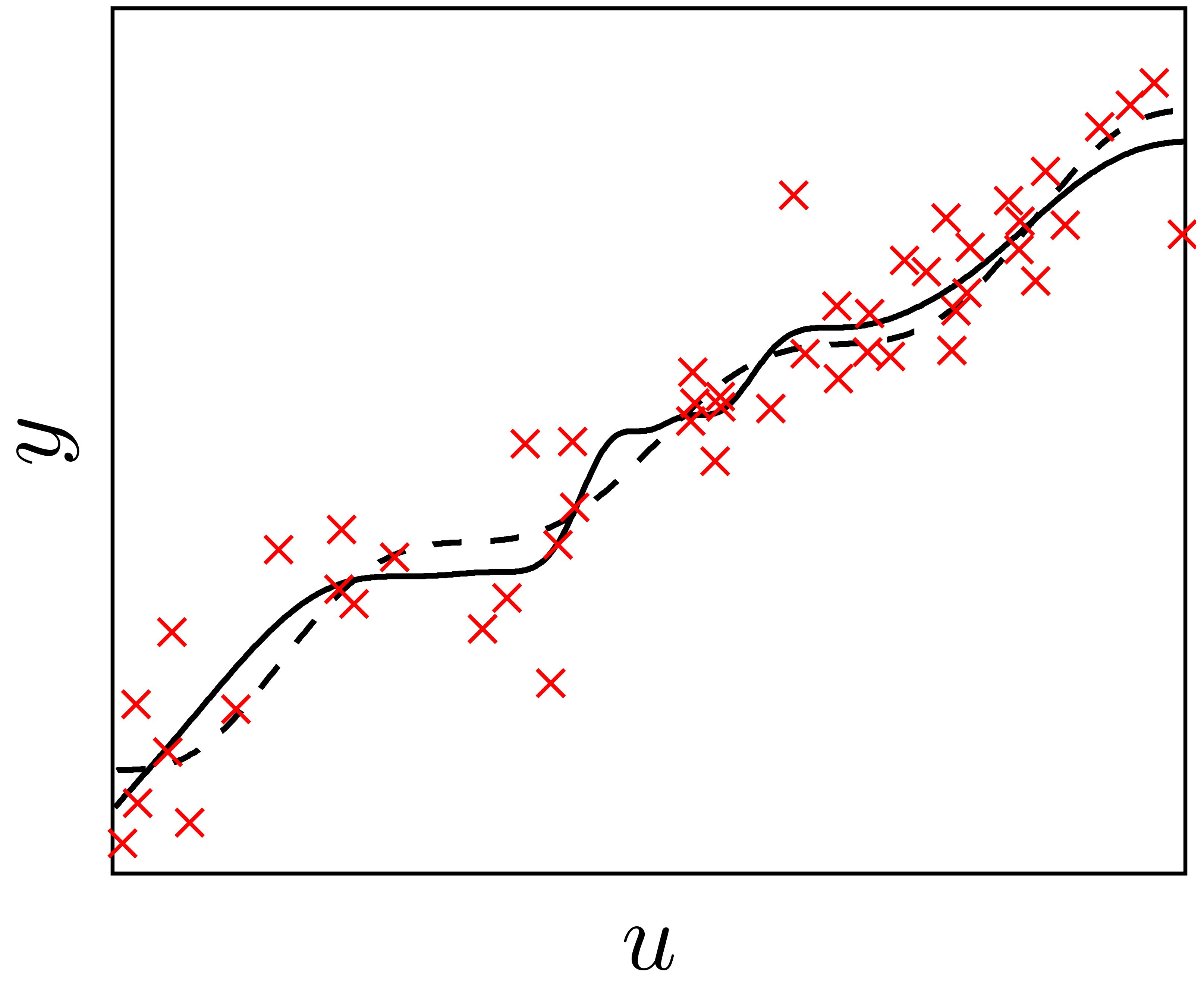
Context: Neumann, Rolf and Steil (2012)



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 $u(1), \ldots, u(t-1), u(t)$
Output $y(1), \ldots, y(t-1), y(t)$

Adapting constraint to time series case

Static Case
(Feedforward Neural Network)

Input u
Output y

↓

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

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

↓

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

vs.

Dynamic Case

(Recurrent Neural Network)

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

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

↓

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

bound

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

network output

coefficients

*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):

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

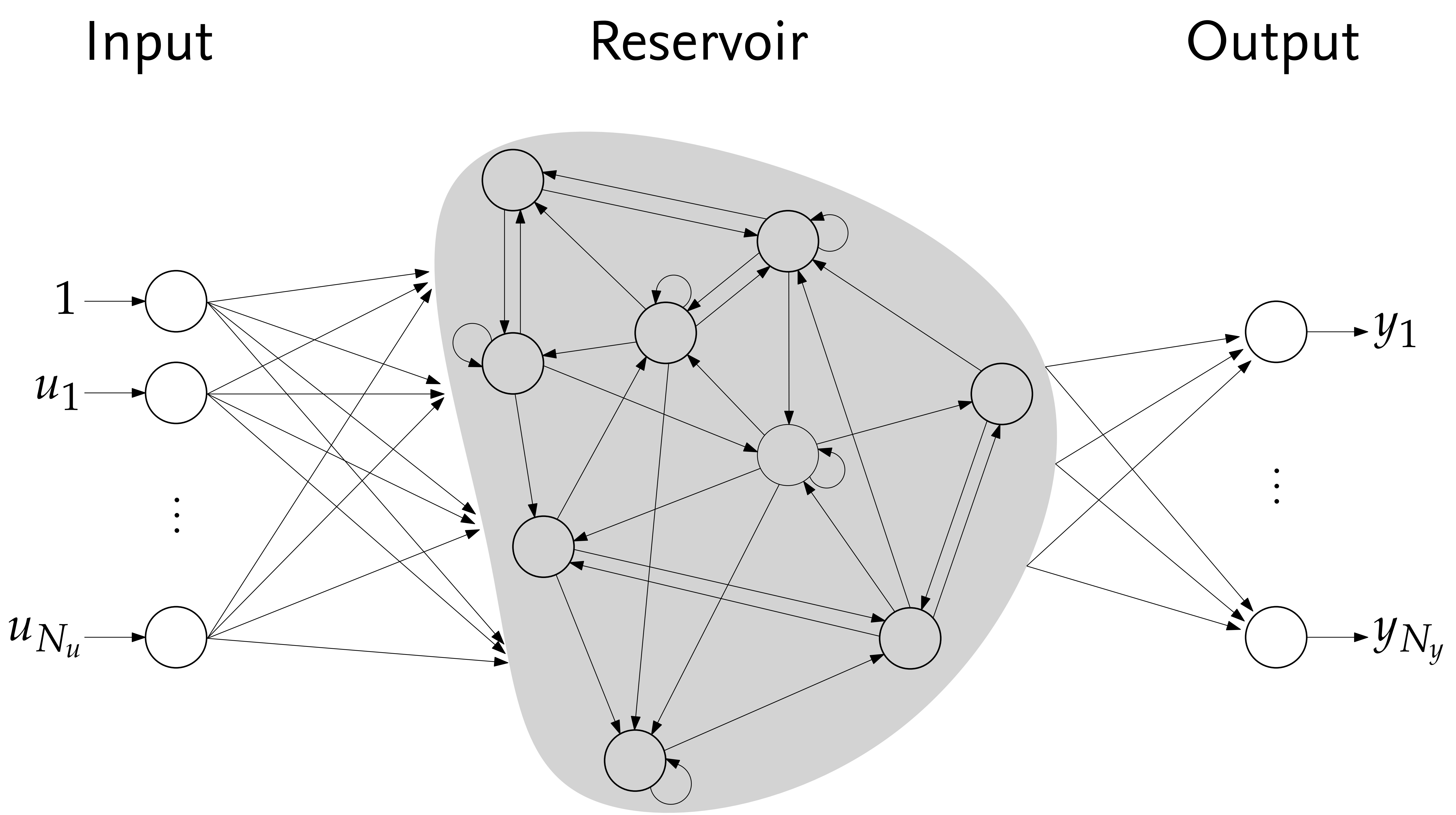
- Difference Quotient of arbitrary order

Constraint Definition:

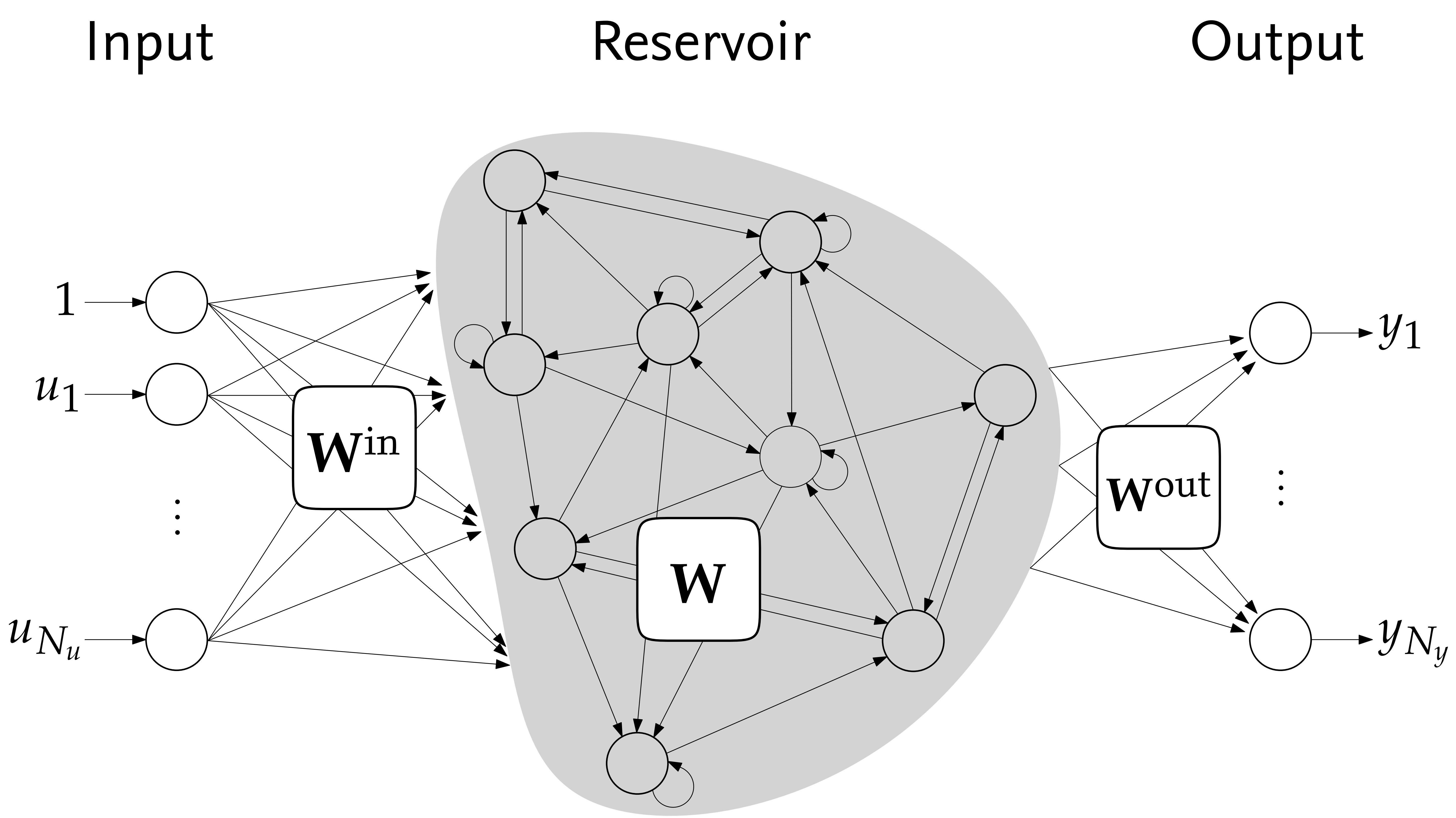
$$\sum_{h=0}^{H} \gamma_h y(t-h) \leq c$$

Outline

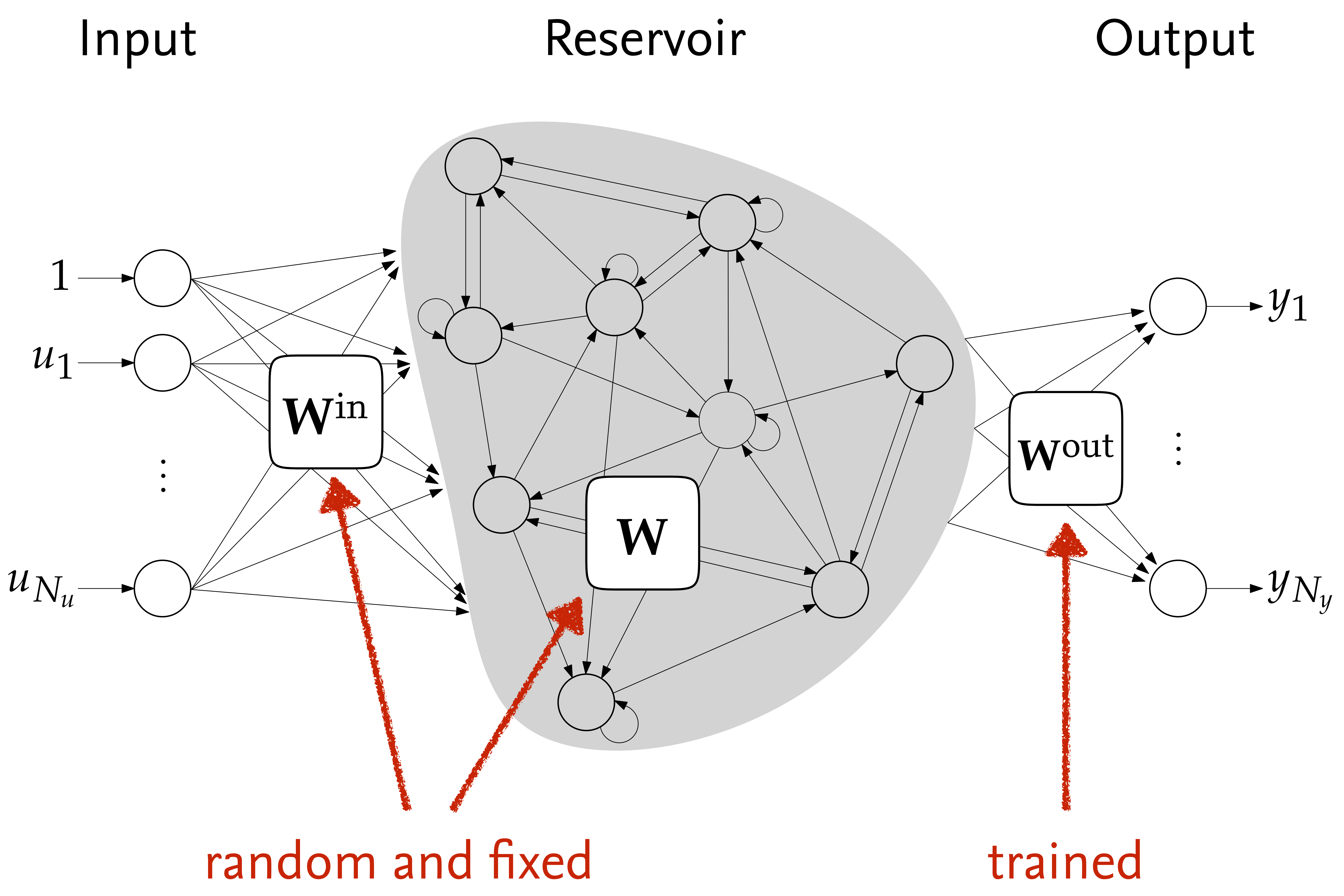
The Echo State Network (1/3)

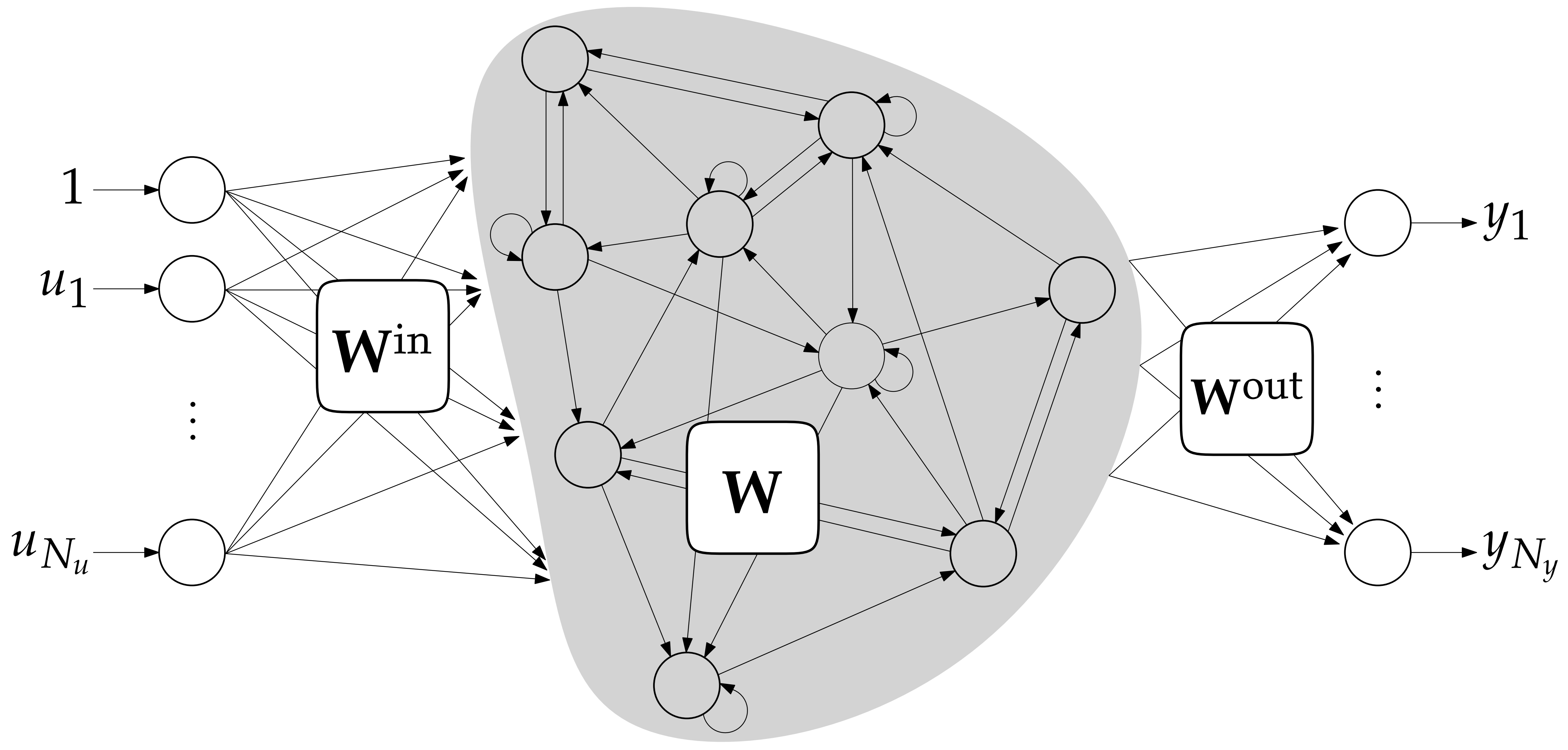


The Echo State Network (1/3)

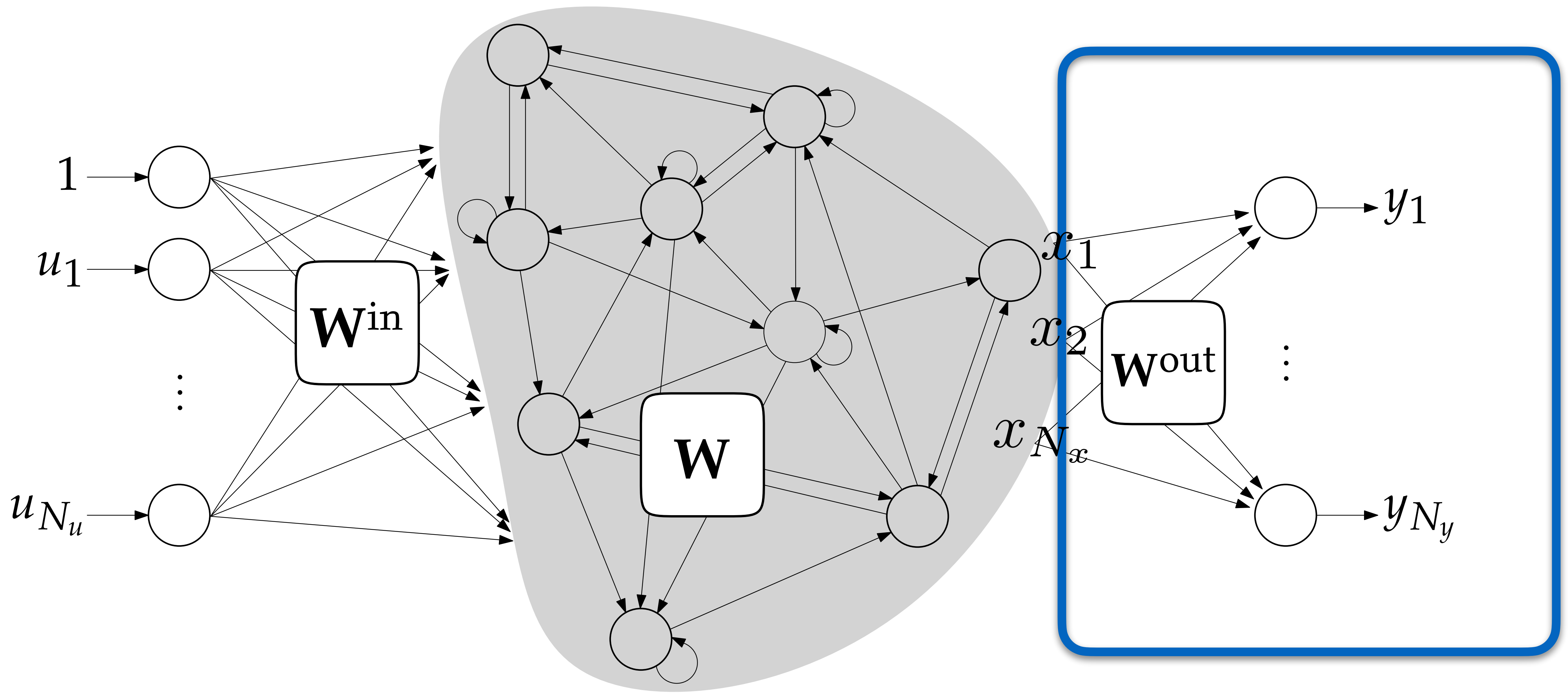


The Echo State Network (1/3)




1
u_1
u_{N_u}
$\mathbf{W}^{\text{in}}$
$\mathbf{W}$
$\mathbf{W}^{\text{out}}$
y_1
y_{N_y}

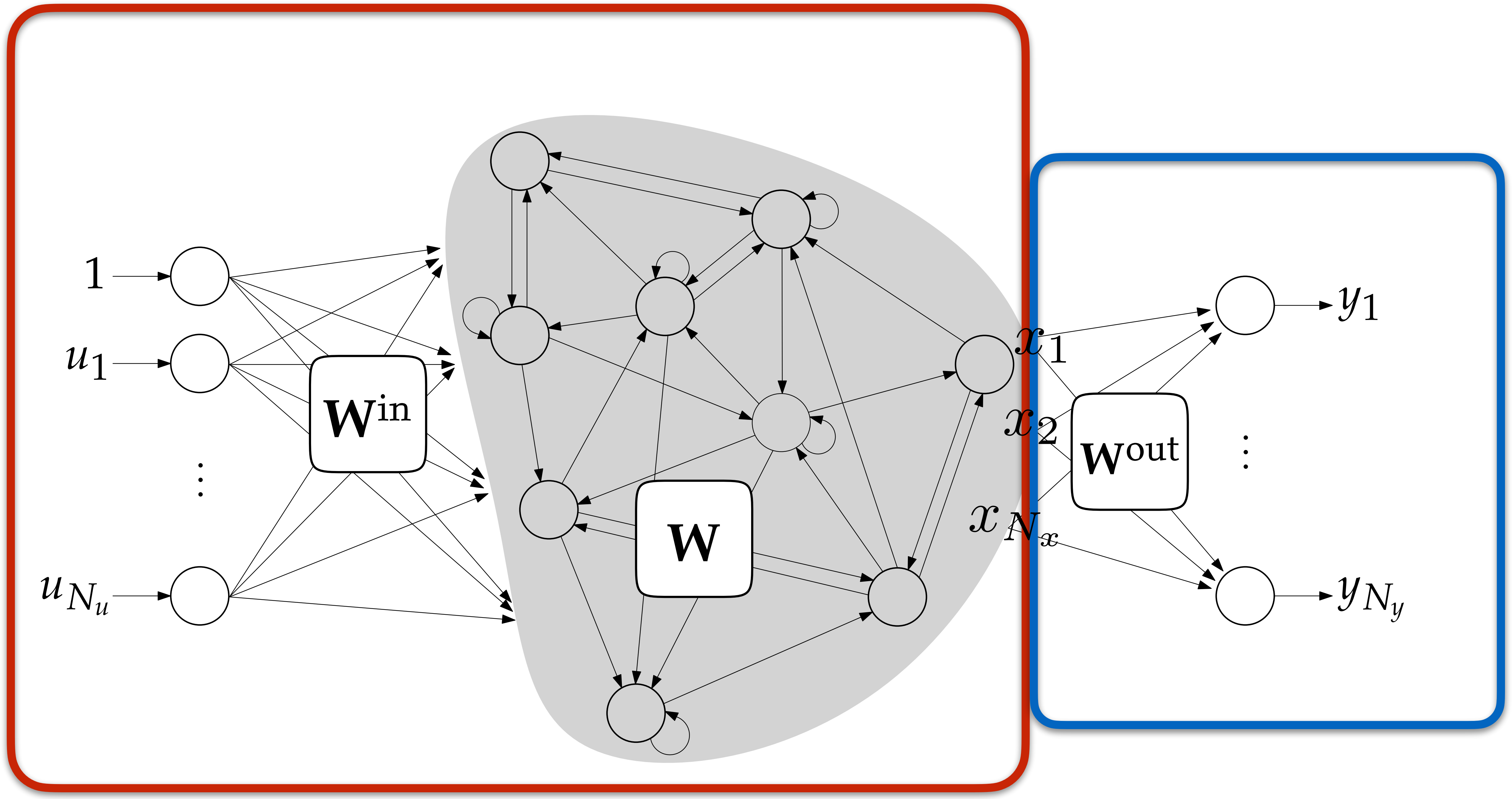
$$\mathbf{y}(t) = \mathbf{W}^{\text{out}}\mathbf{x}(t)$$



- feedforward
- linear read-out

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

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



- high-dimensional
- non-linear
- recurrent

- feedforward
- linear read-out

The Echo State Network (3/3)

- Minimize Training Error:

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

The Echo State Network (3/3)

- Minimize Training Error:

$$\epsilon_{\mathrm{RMSE}} = \left|\left|\mathbf{Y} - \mathbf{Y}^{\mathrm{target}}\right|\right|^2 = \left|\left|\mathbf{W}^{\mathrm{out}}\mathbf{X} - \mathbf{Y}^{\mathrm{target}}\right|\right|^2$$

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

The Echo State Network (3/3)

- Minimize Training Error:

$$\epsilon_{\text{RMSE}} = \left|\left|\mathbf{Y} - \mathbf{Y}^{\text{target}}\right|\right|^2$$
$$= \left|\left|\mathbf{W}^{\text{out}}\mathbf{X} - \mathbf{Y}^{\text{target}}\right|\right|^2$$
$$= \left|\left|\mathbf{X}^T(\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\right|\right|^2$$

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

The Echo State Network (3/3)

- Minimize Training Error:

$$\epsilon_{\text{RMSE}} = \left|\left|\mathbf{Y} - \mathbf{Y}^{\text{target}}\right|\right|^2$$
$$= \left|\left|\mathbf{W}^{\text{out}}\mathbf{X} - \mathbf{Y}^{\text{target}}\right|\right|^2$$
$$= \left|\left|\mathbf{X}^T(\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\right|\right|^2$$

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

- Equivalent to a Linear Least Squares Problem:

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

The Echo State Network (3/3)

- Minimize Training Error:

$$\epsilon_{\text{RMSE}} = \left|\left|\mathbf{Y} - \mathbf{Y}^{\text{target}}\right|\right|^2$$
$$= \left|\left|\mathbf{W}^{\text{out}}\mathbf{X} - \mathbf{Y}^{\text{target}}\right|\right|^2$$
$$= \left|\left|\mathbf{X}^T(\mathbf{W}^{\text{out}})^T - (\mathbf{Y}^{\text{target}})^T\right|\right|^2$$

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

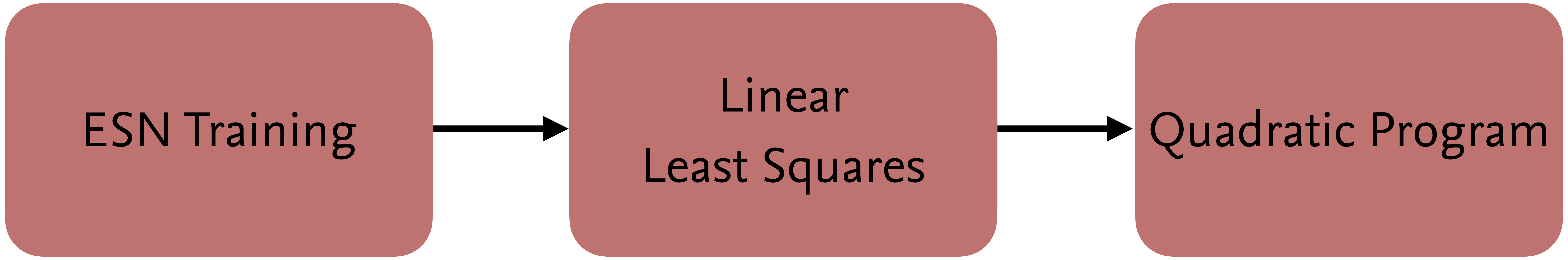
- Equivalent to a Linear Least Squares Problem:

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

Linear Least Squares:

$$\min \left|\left|\mathbf{A}\mathbf{X} - \mathbf{b}\right|\right|^2$$

Reduction to Quadratic Program



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

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

$$\min \ \mathbf{xQx} \quad \text{s.t.} \ \mathbf{Gx} \leq \mathbf{h}$$

Constraints* to QP constraints

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

*simplified to one-dimensional output

Constraints* to QP constraints

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

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

*simplified to one-dimensional output

Constraints* to QP constraints

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

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

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

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

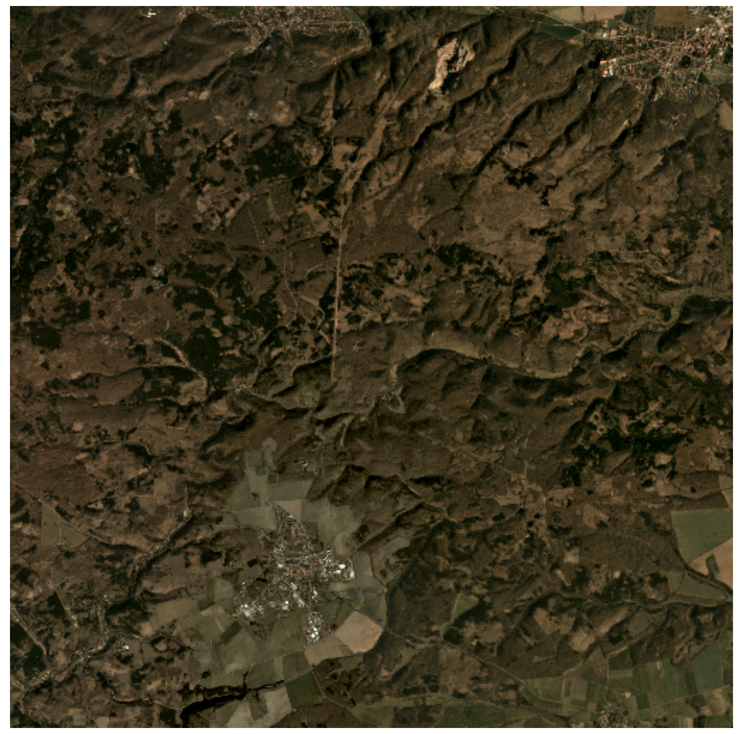
*simplified to one-dimensional output

Outline

Satellite Image Time Series of the Harz (monthly)

Jan 2020

Feb 2020



March 2020

April 2020

May 2020

June 2020

July 2020

Aug 2020

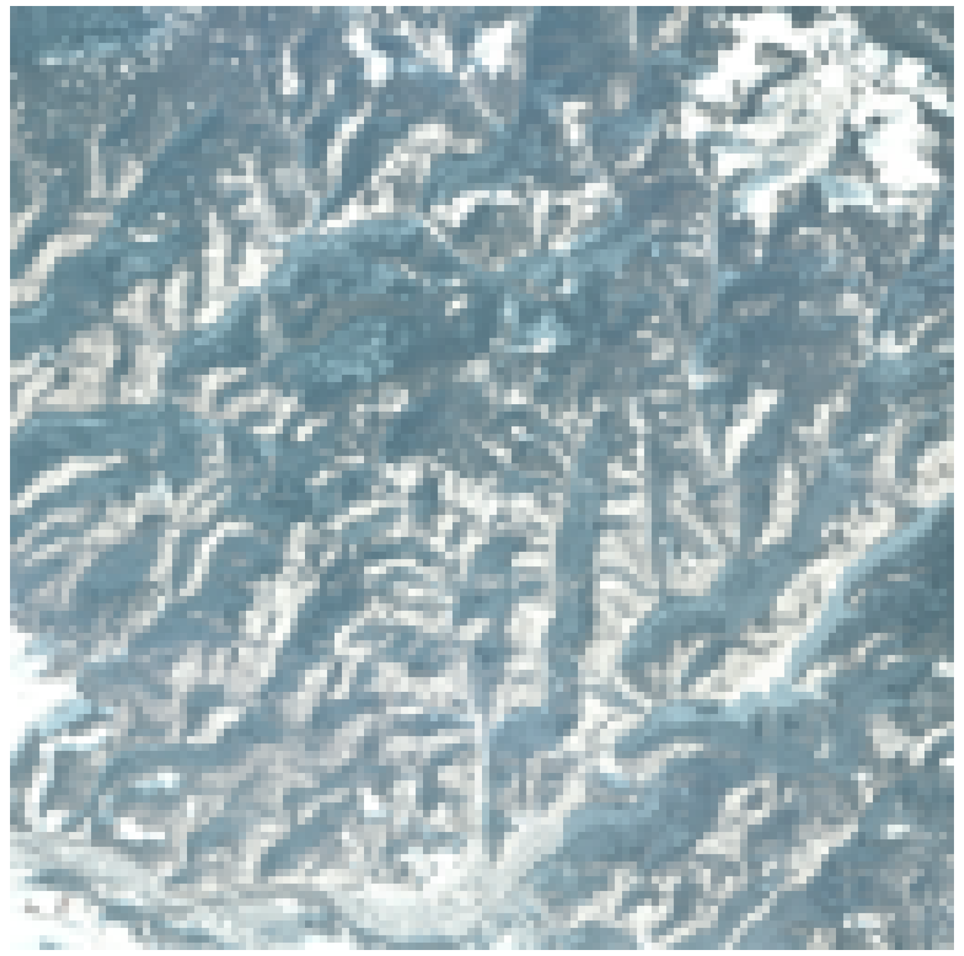
Sep 2020

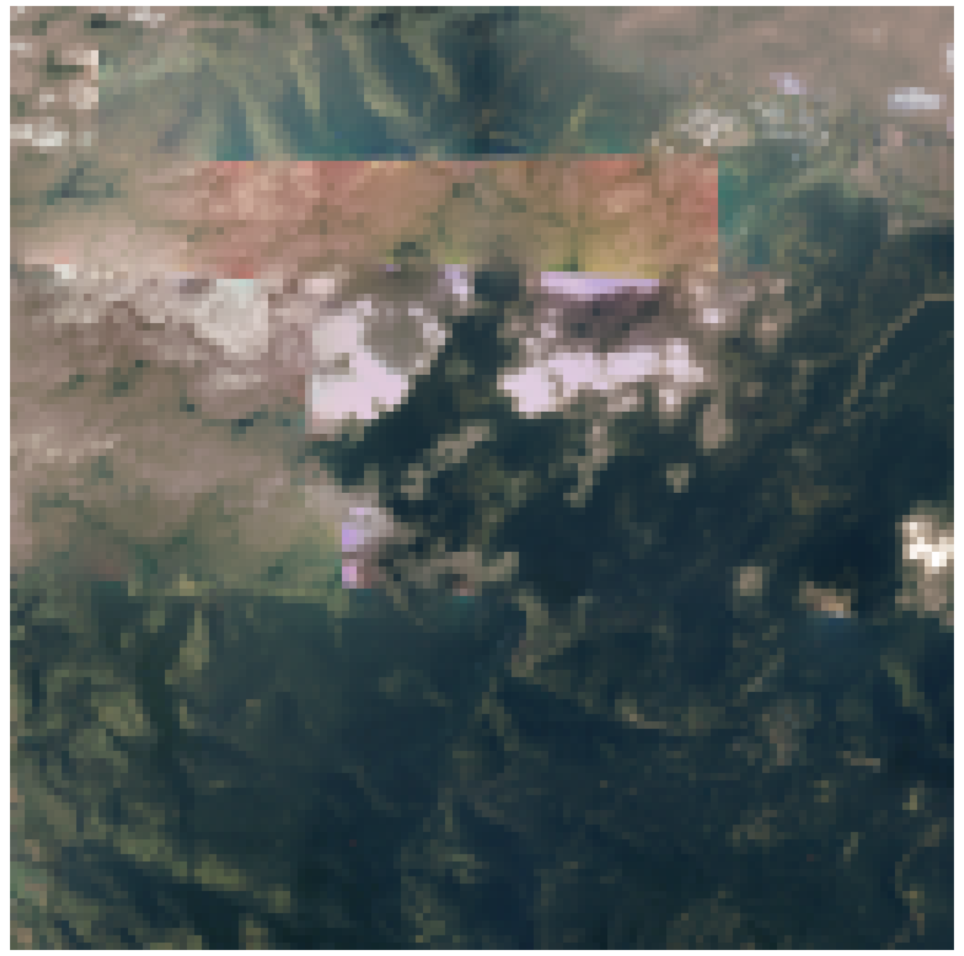
Oct 2020

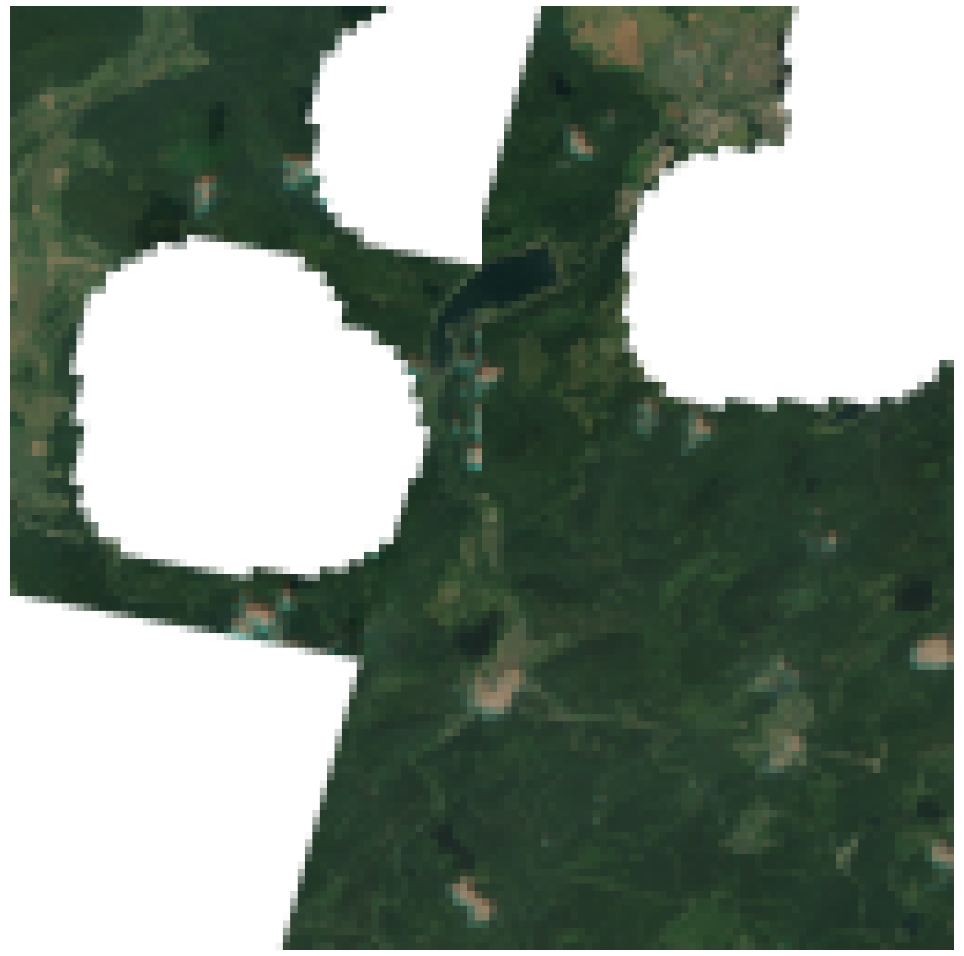
Nov 2020

Dec 2020

Noisy Images



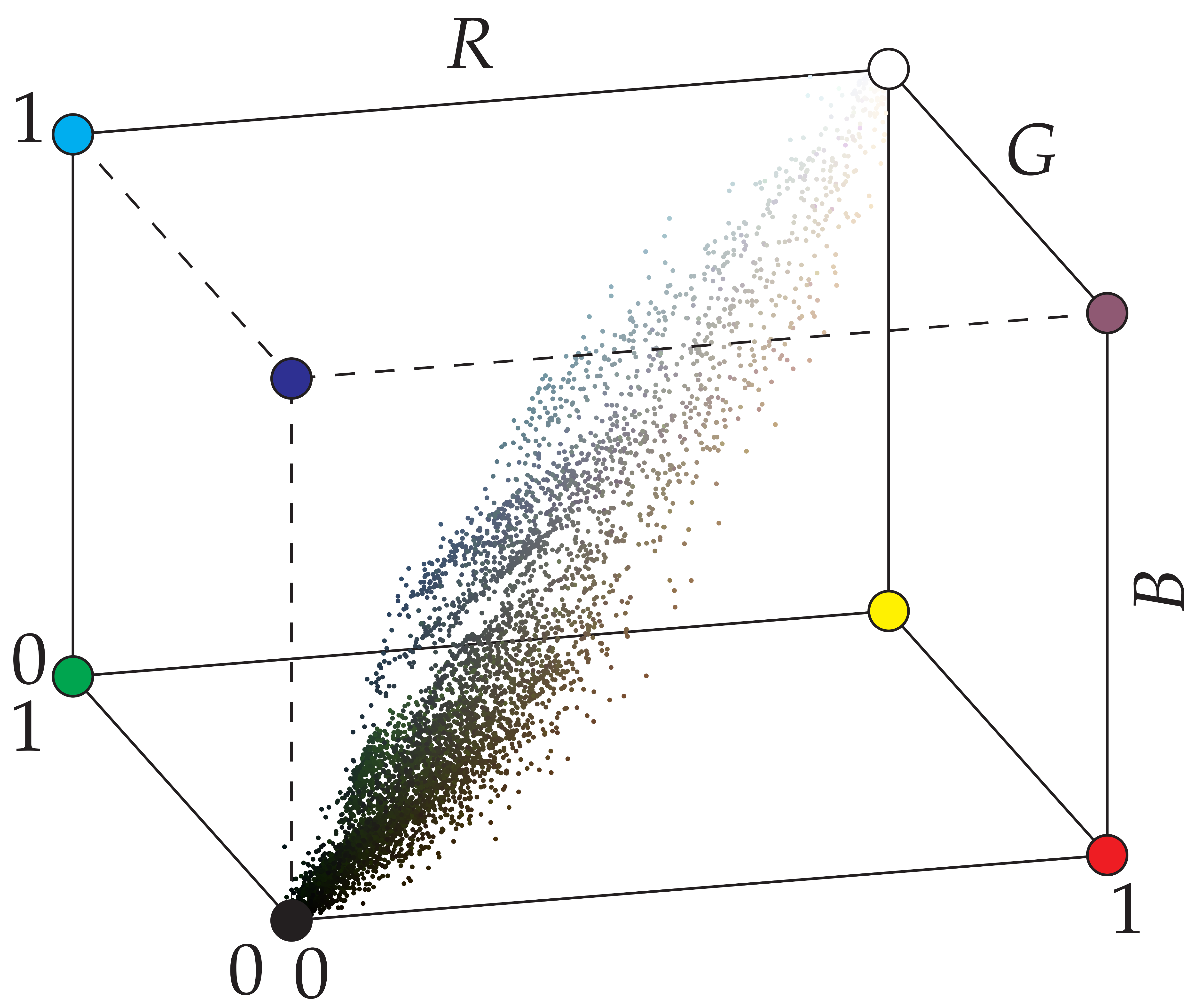




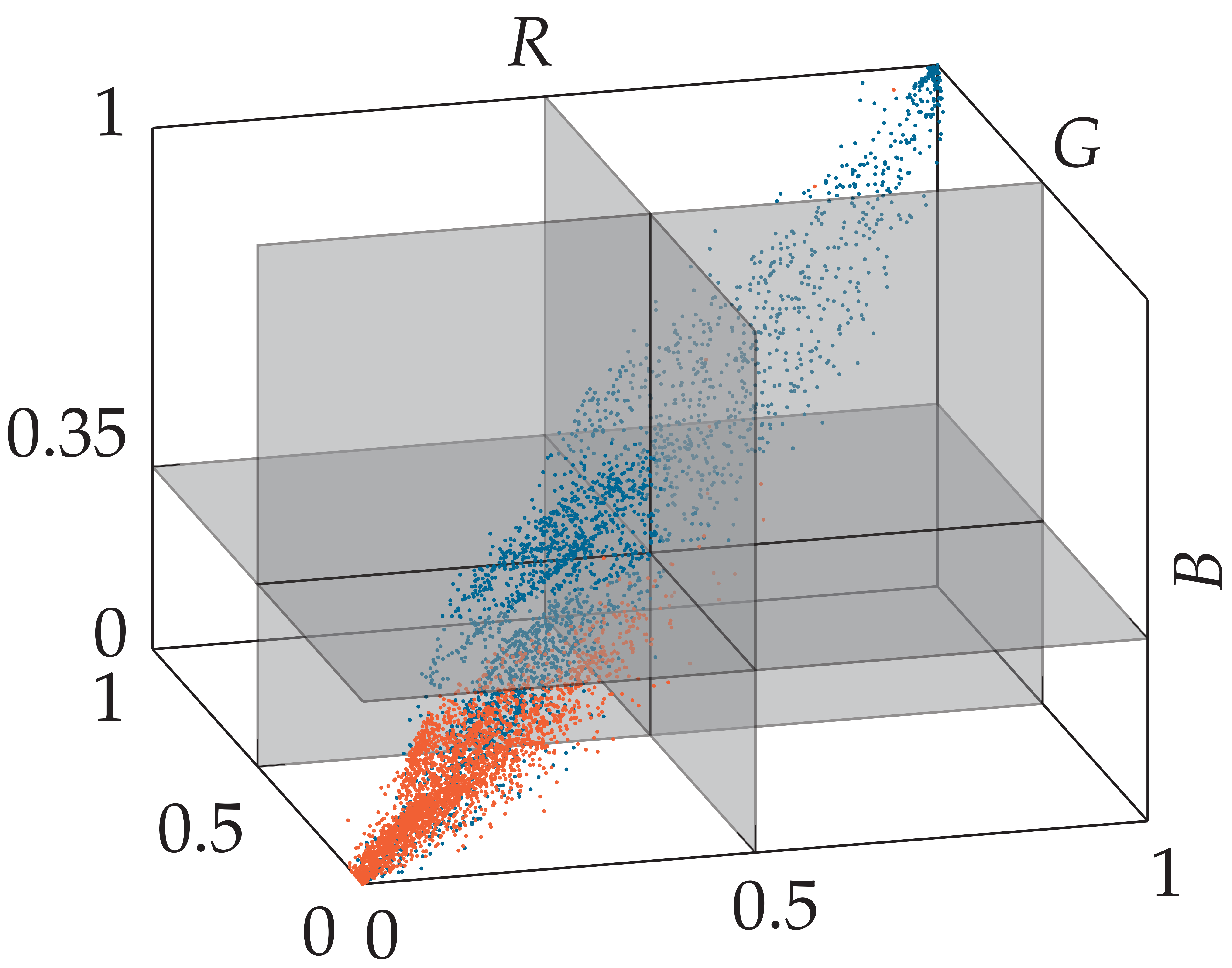
Training Task

- 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



- Not noisy image pixels
- Noisy image pixels

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

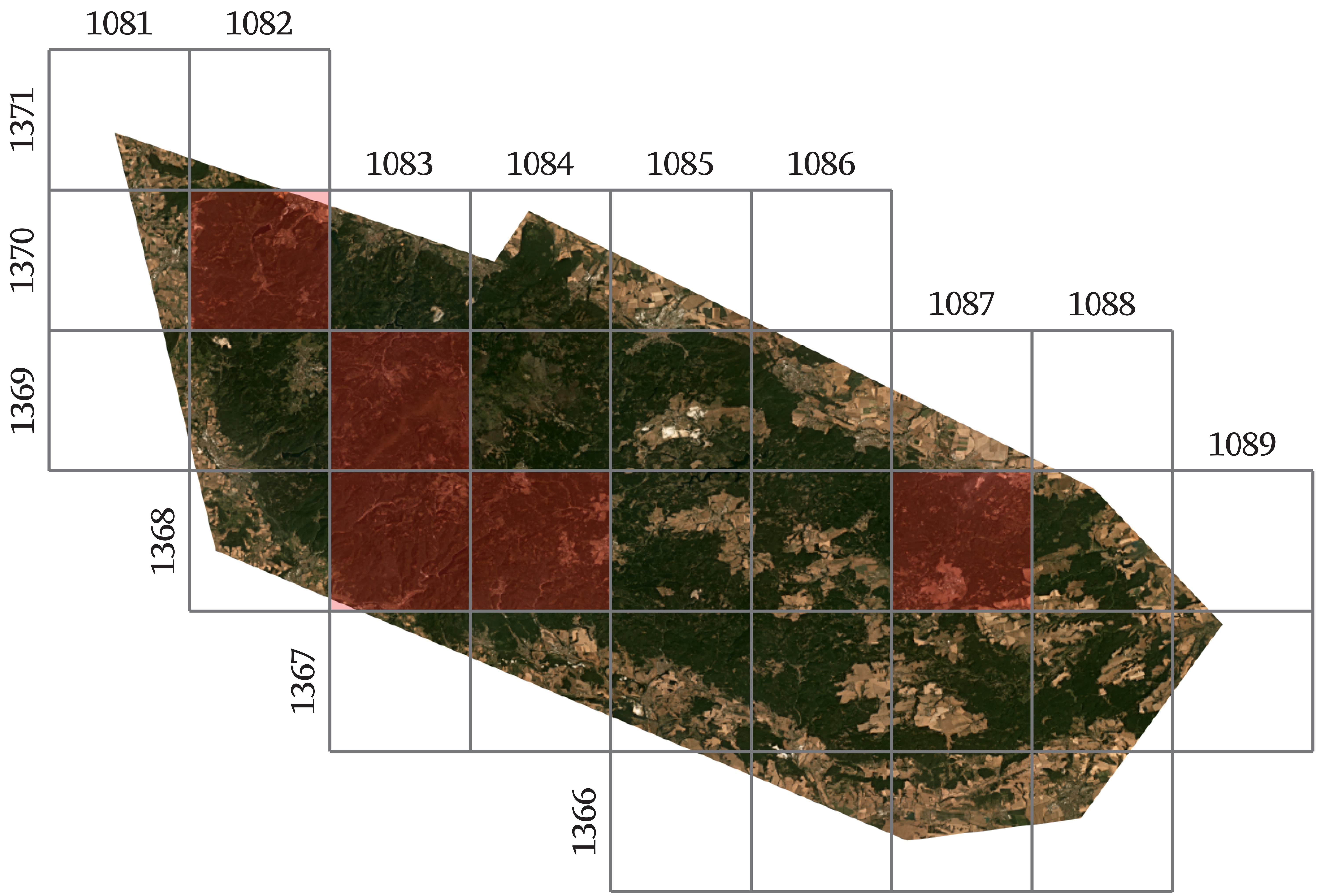
$$|y_R(t) - y_R(t-1)| \leq 0.05$$

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

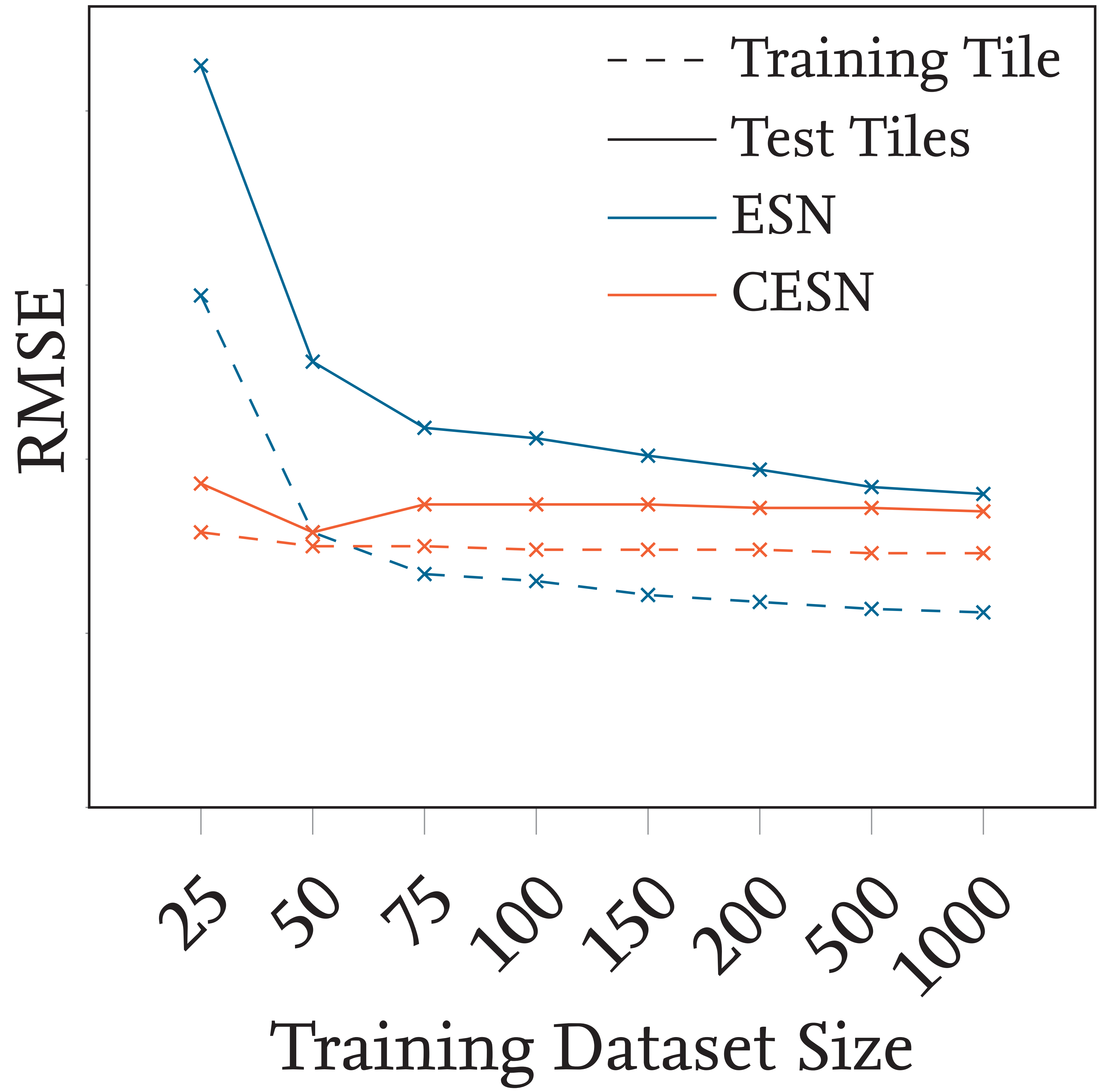
$$|y_B(t) - y_B(t-1)| \leq 0.05$$

→ 9 constraints at each time step in total

Training Method

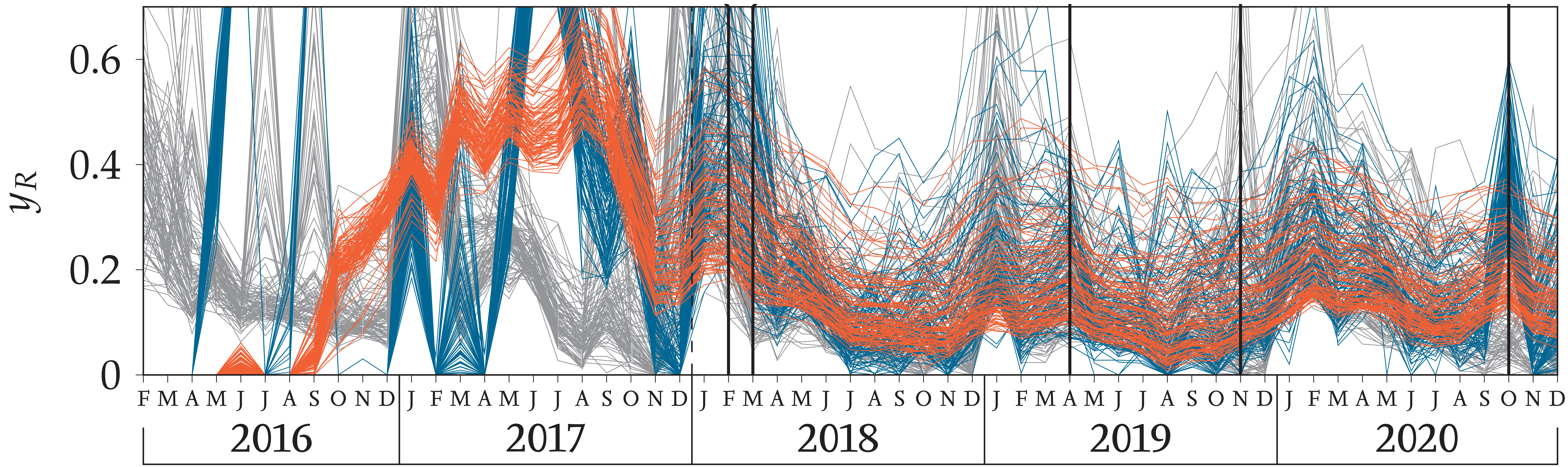


Results (1/2)



Training Tile
Test Tiles
ESN
CESN
RMSE
25
50
75
100
150
200
500
1000
Training Dataset Size

Results (2/2)



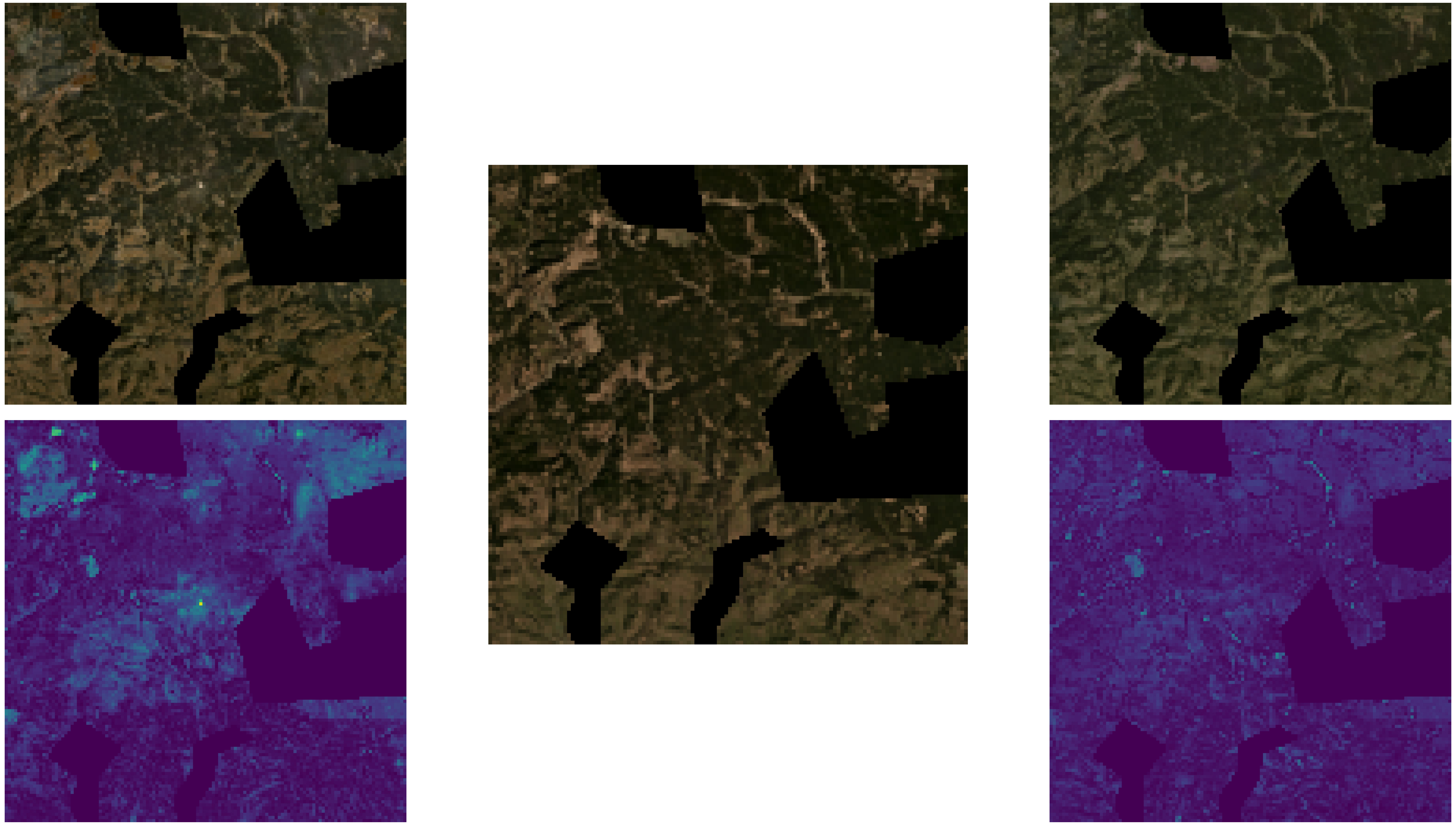
0.6
0.4
0.2
0
y_R
F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D
2016
2017
2018
2019
2020

Prediction Examples (1/3)

ESN | Target | CESN

Prediction

Error



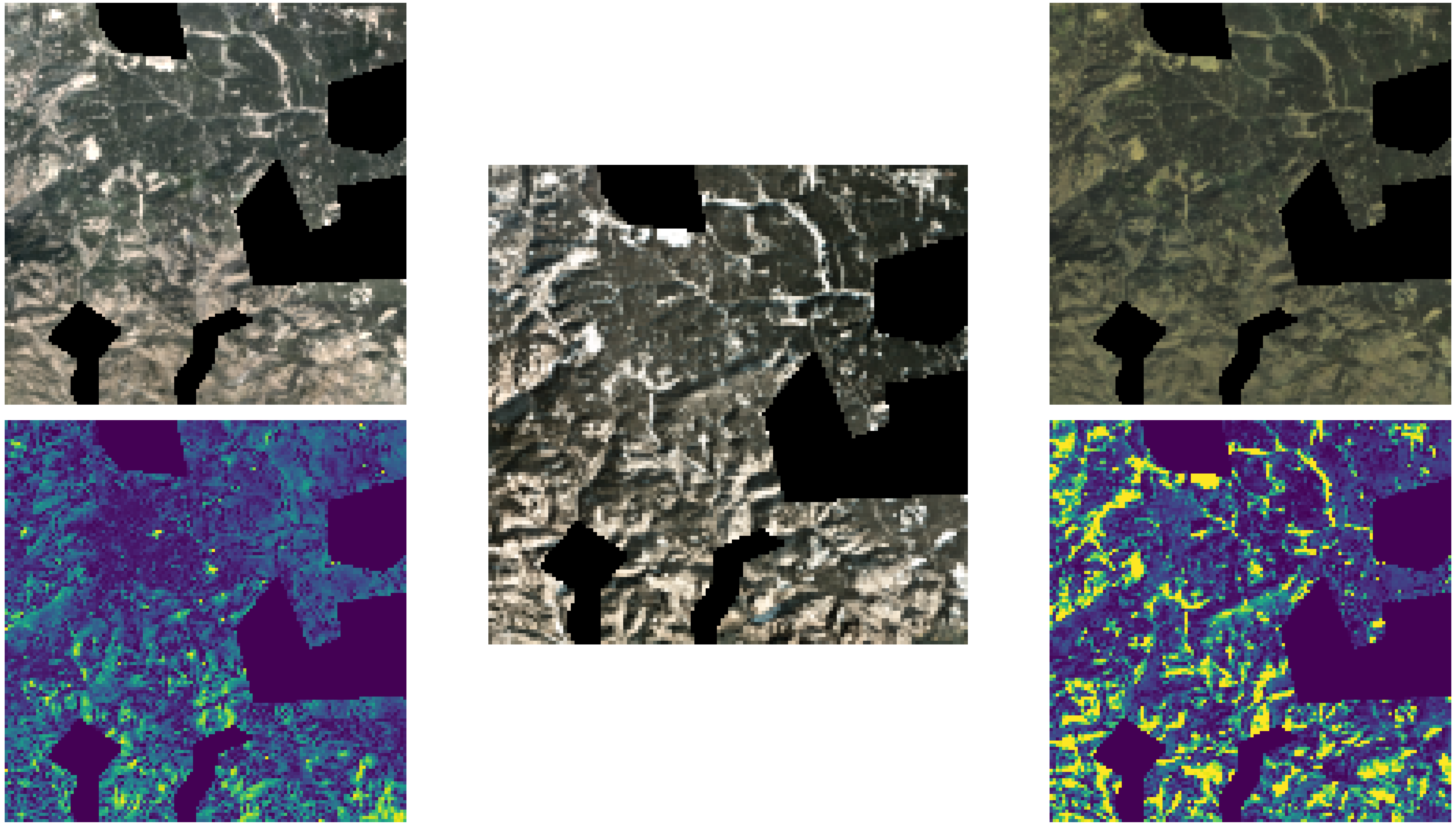
Technische Universität Braunschweig

Prediction Examples (2/3)

ESN | Target | CESN

Prediction

Error

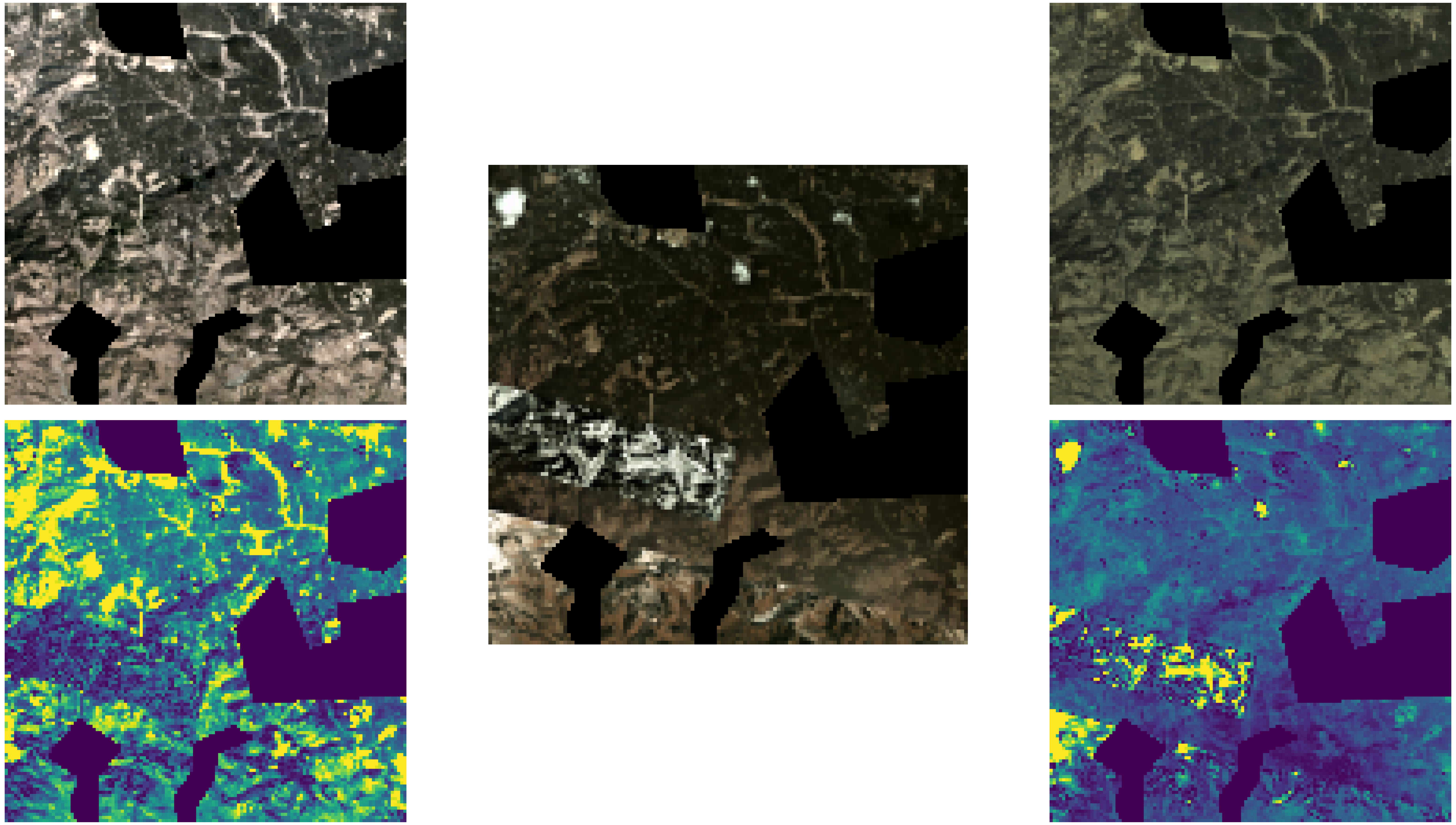


Prediction Examples (3/3)

ESN | Target | CESN

Prediction

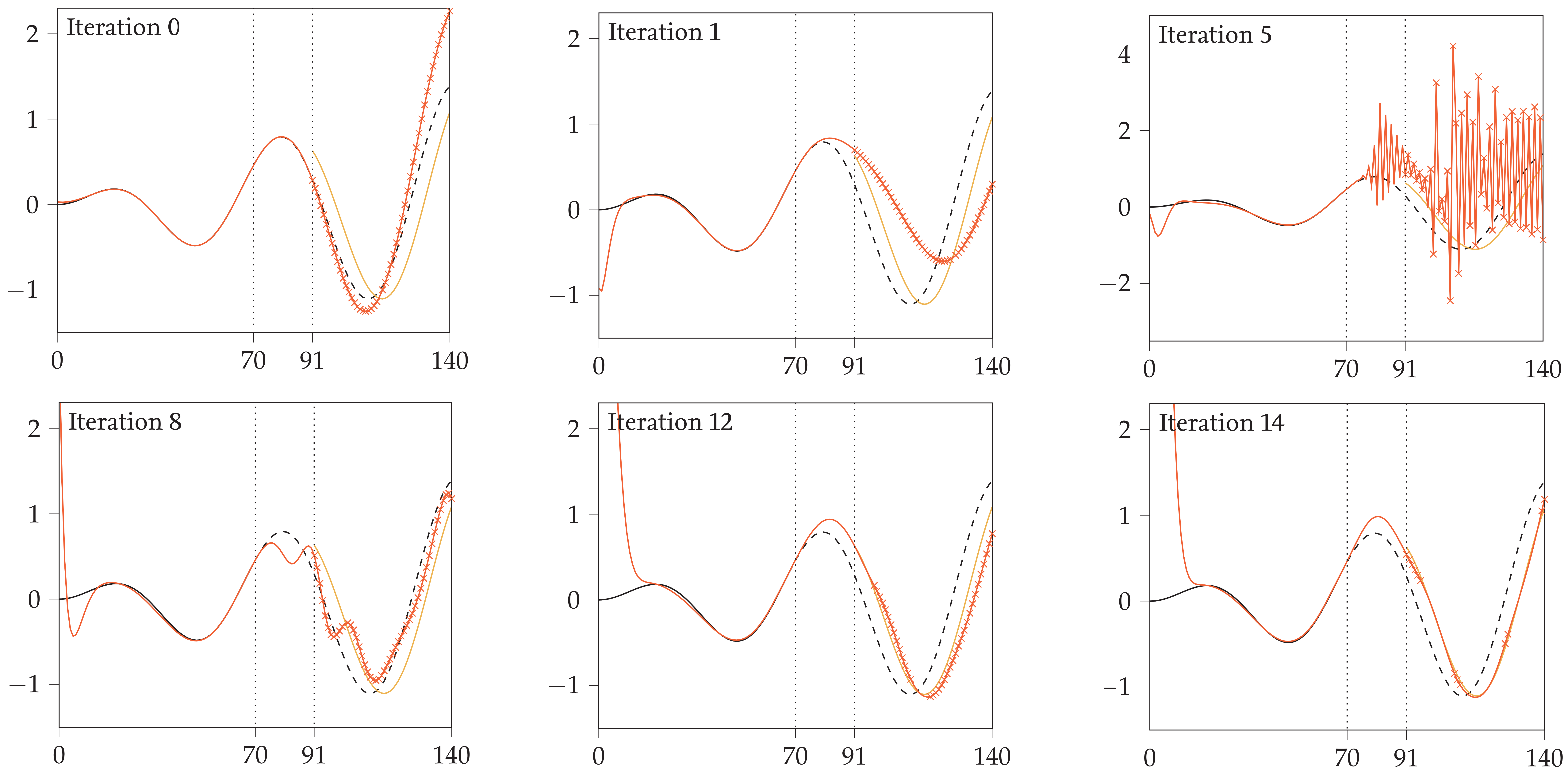
Error



Outline

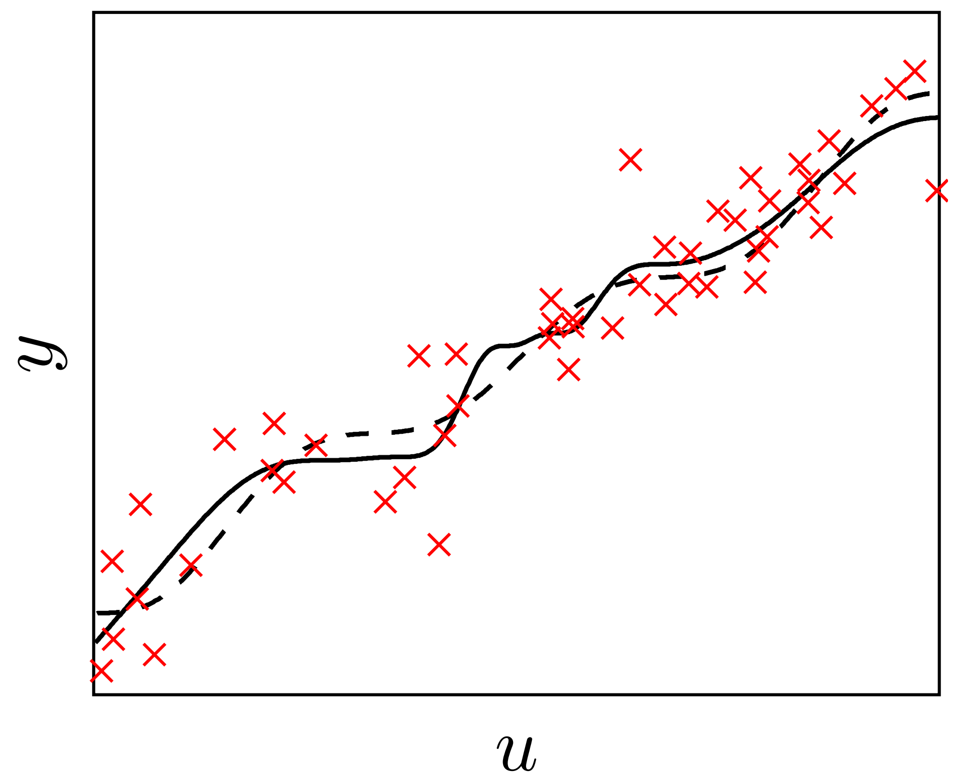
1. Constraint Definition
2. Embedding Constraints into the Neural Network
3. Example: Forecasting of Satellite Images
4. Future Work & Conclusion

Future Work: Recursive Predictions

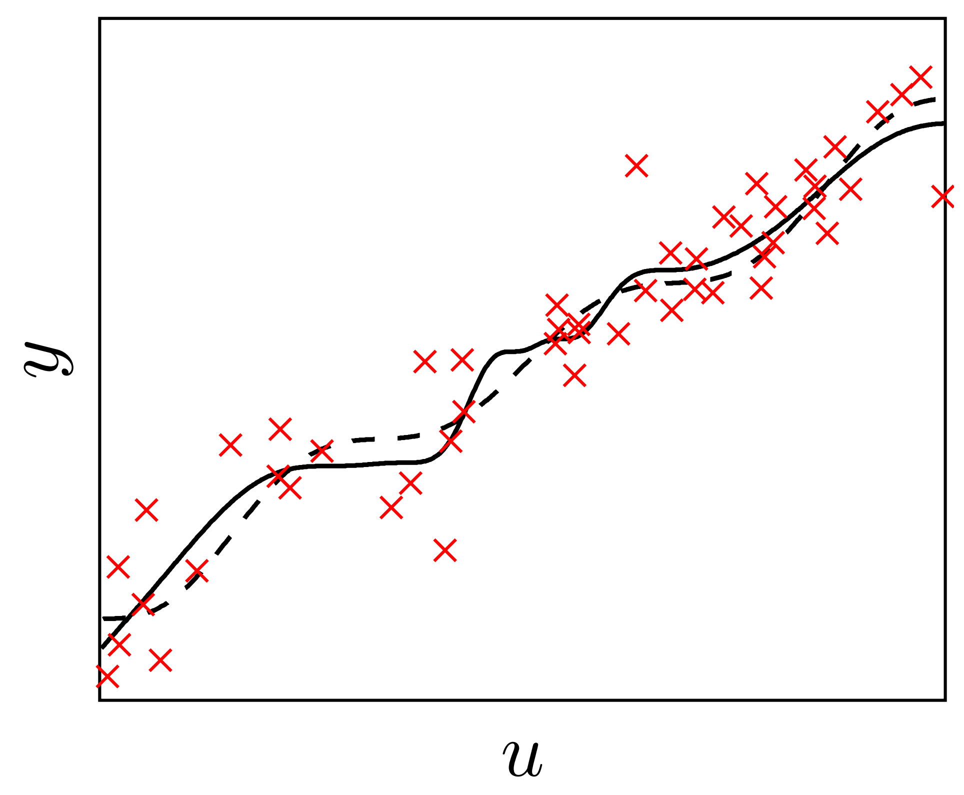


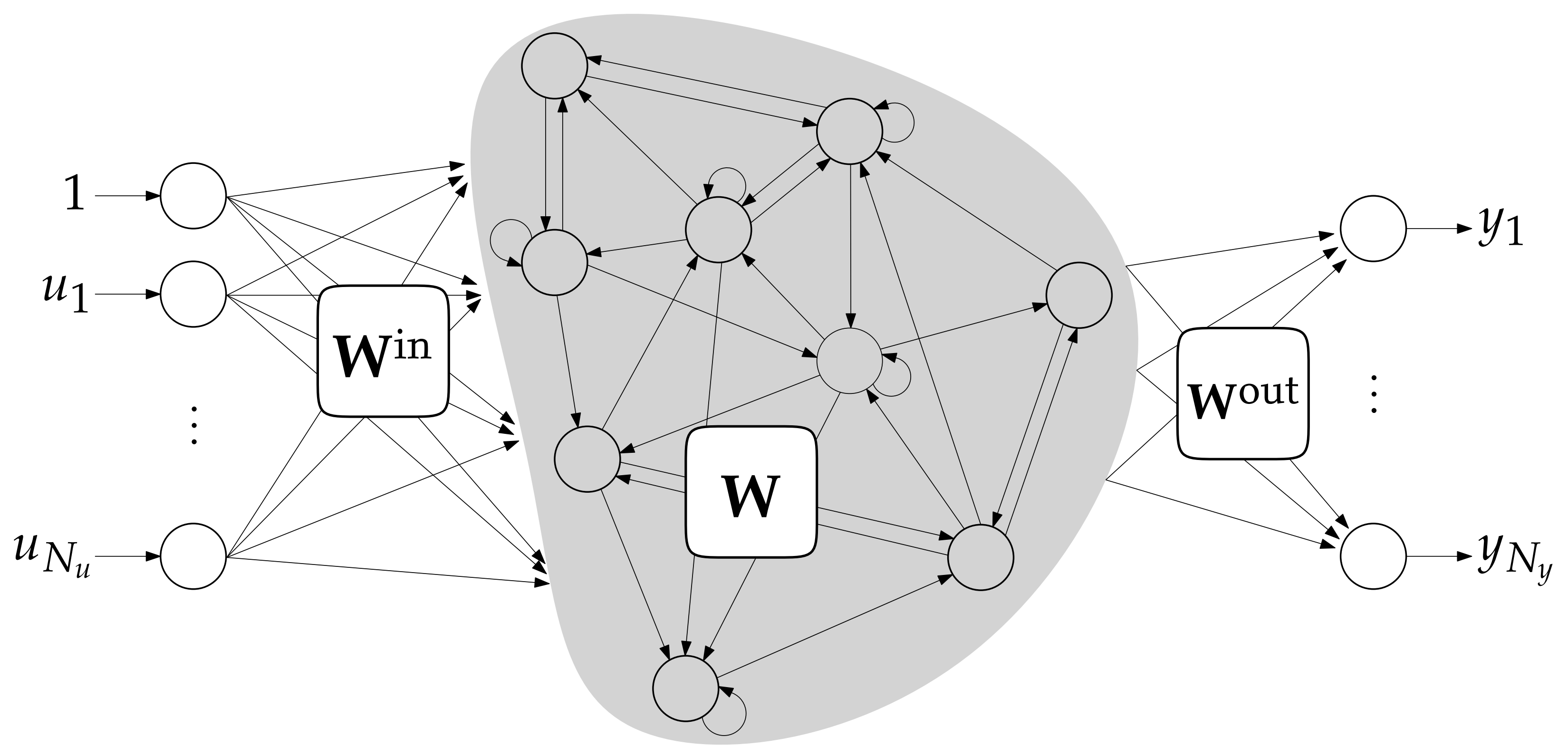
Conclusion

Conclusion

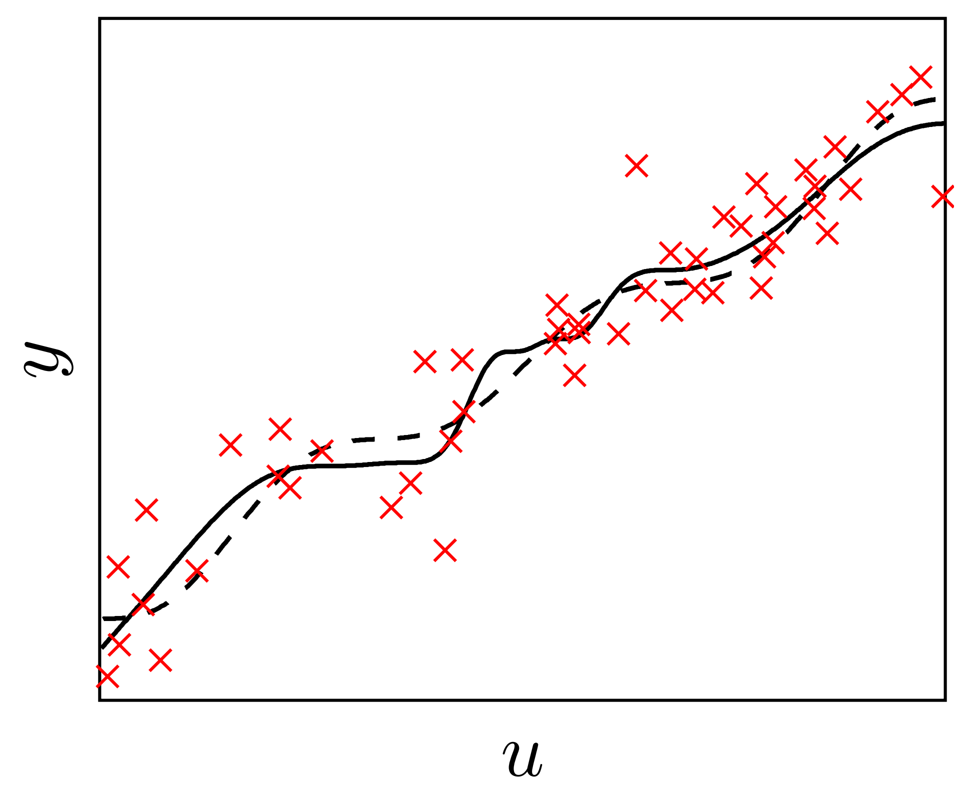


Conclusion

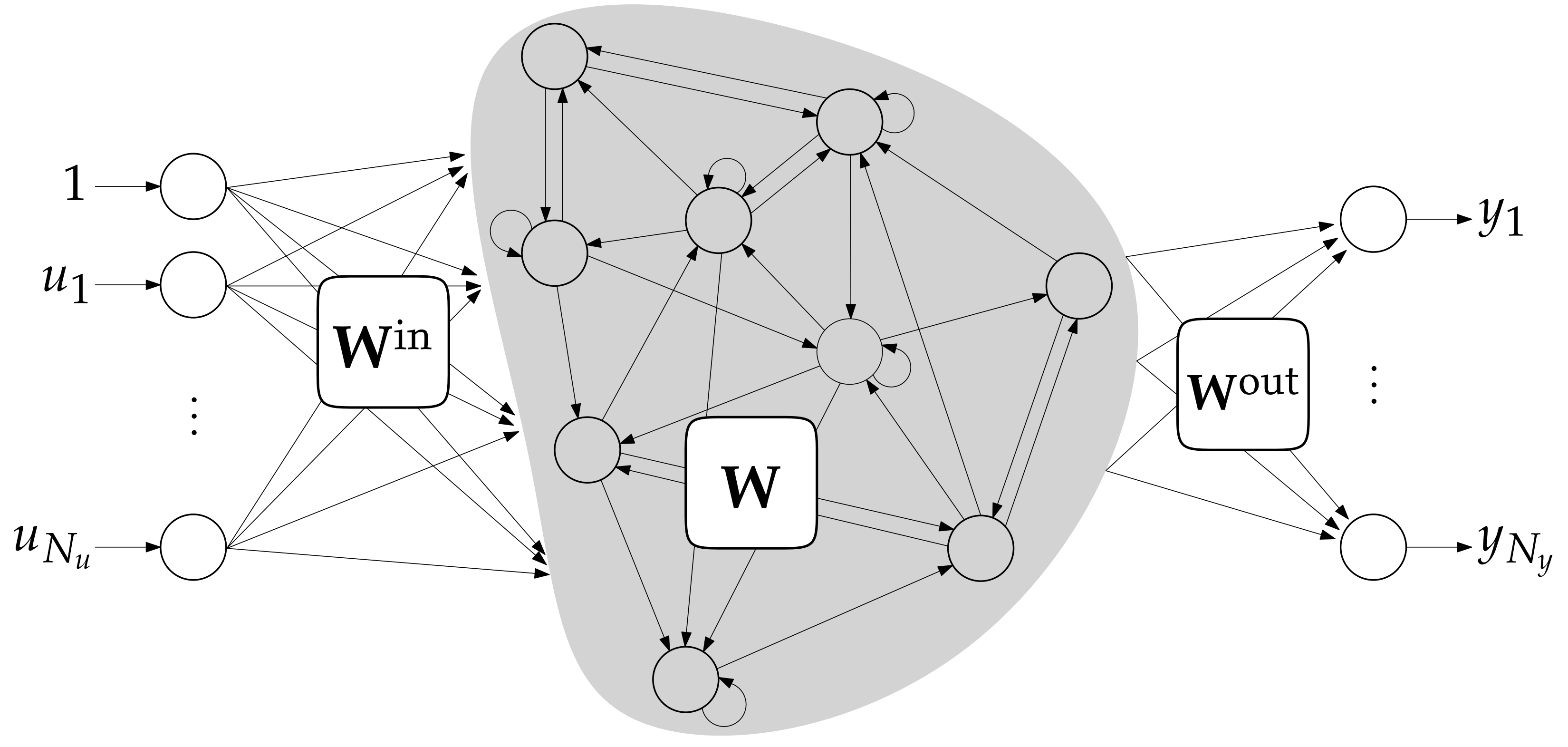




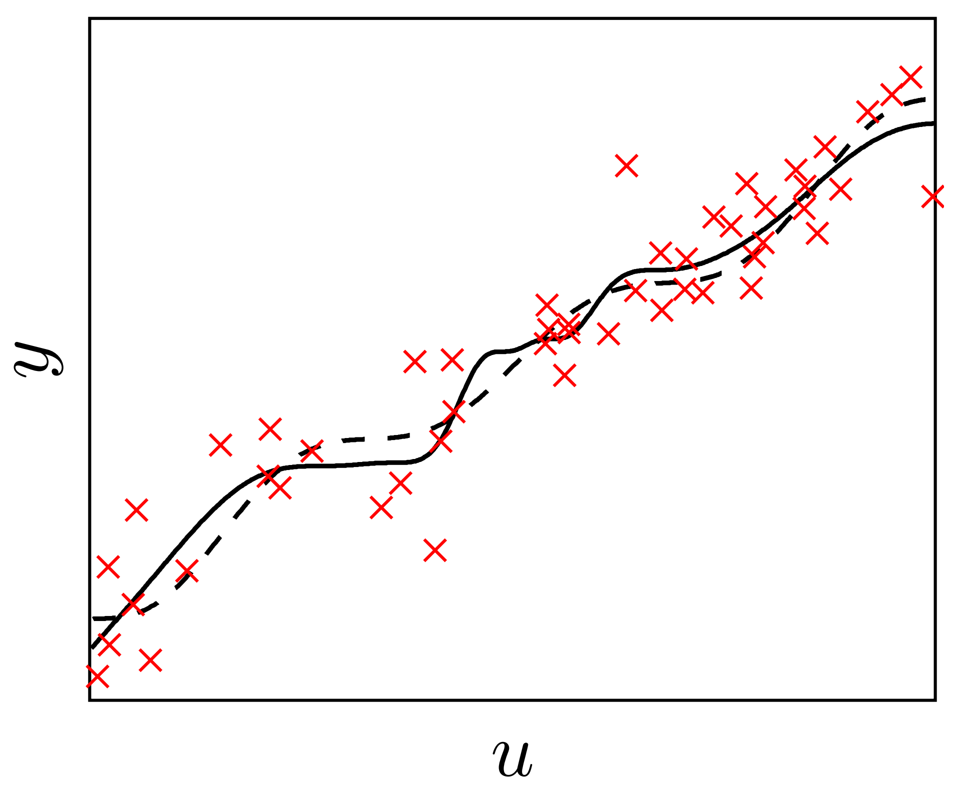
Conclusion



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

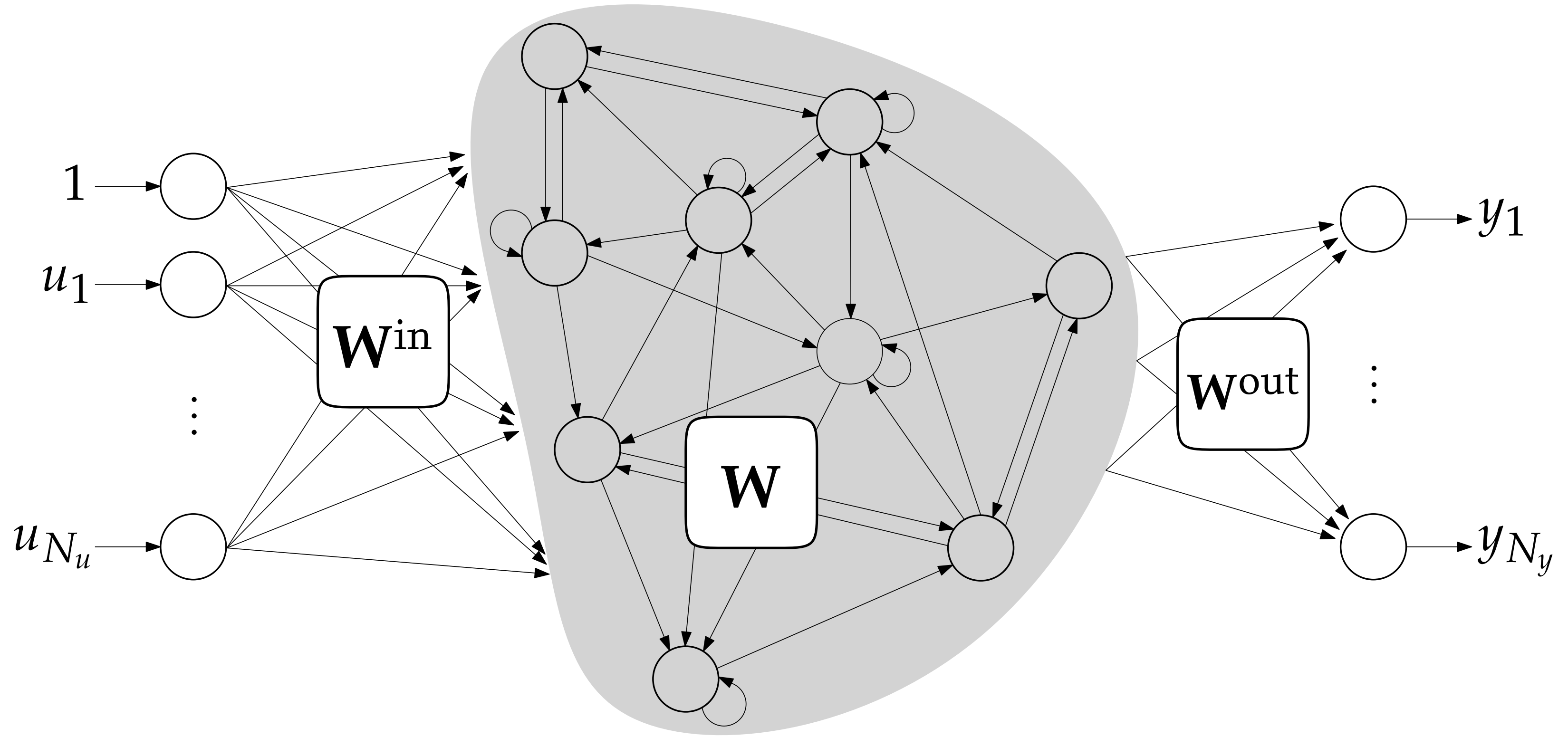


Conclusion

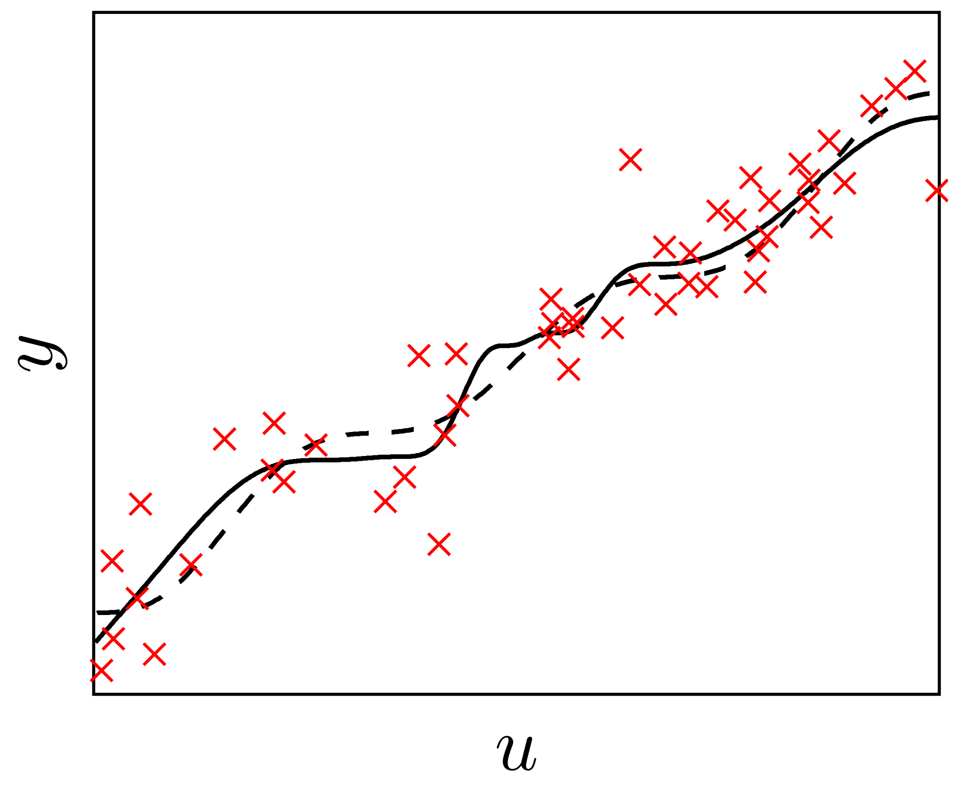


$$\sum_{h=0}^{H} \gamma_h y(t-h) \leq c$$

ESN Training → Linear Least Squares → Quadratic Program

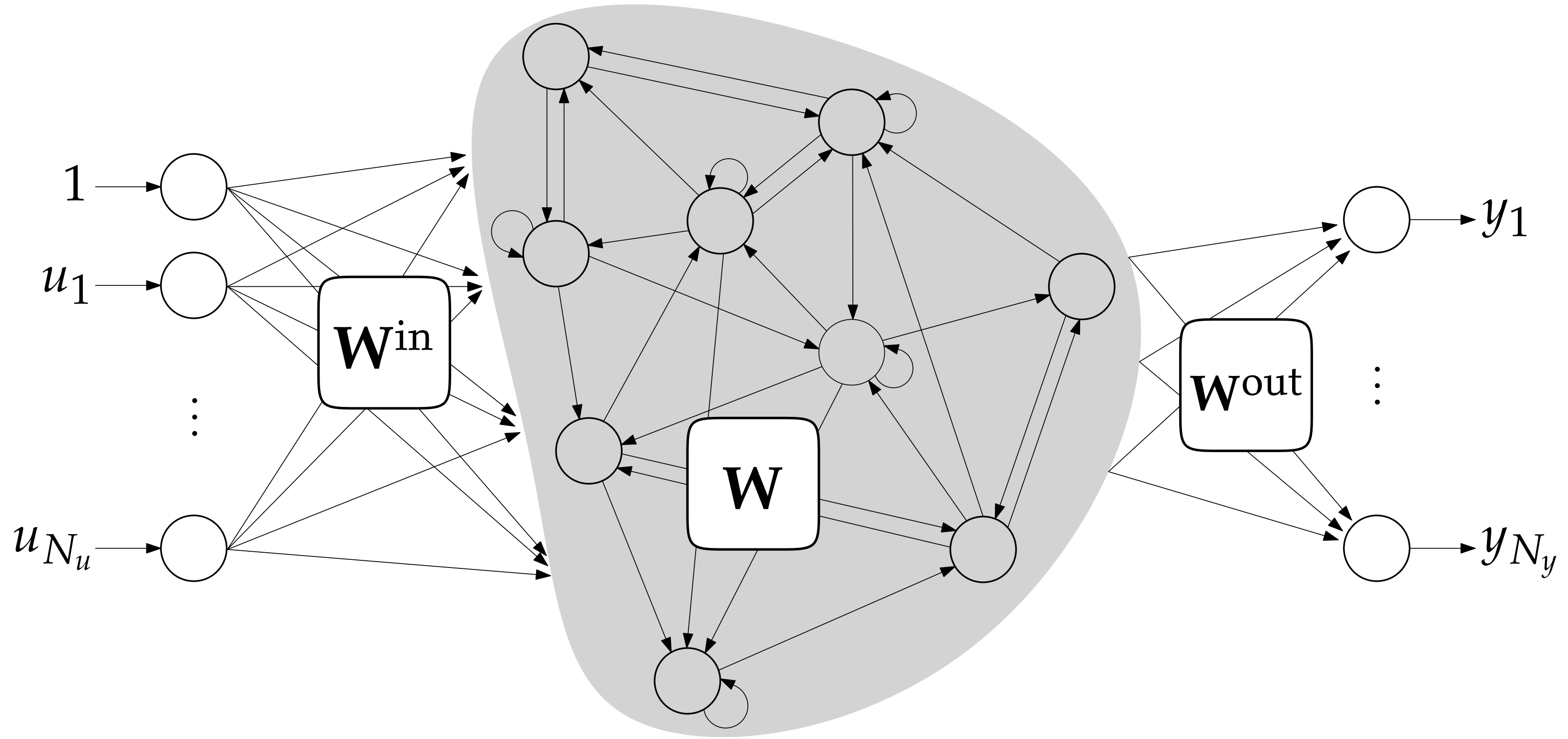


Conclusion

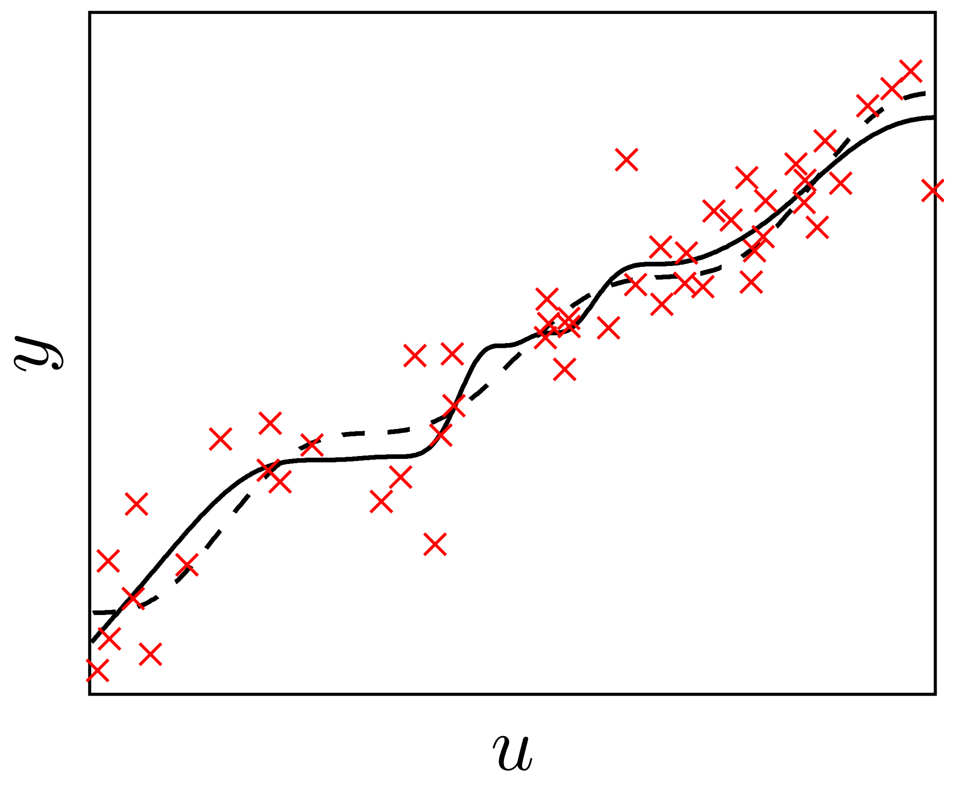


$$\sum_{h=0}^{H} \gamma_h y(t-h) \leq c$$

ESN Training → Linear Least Squares → Quadratic Program

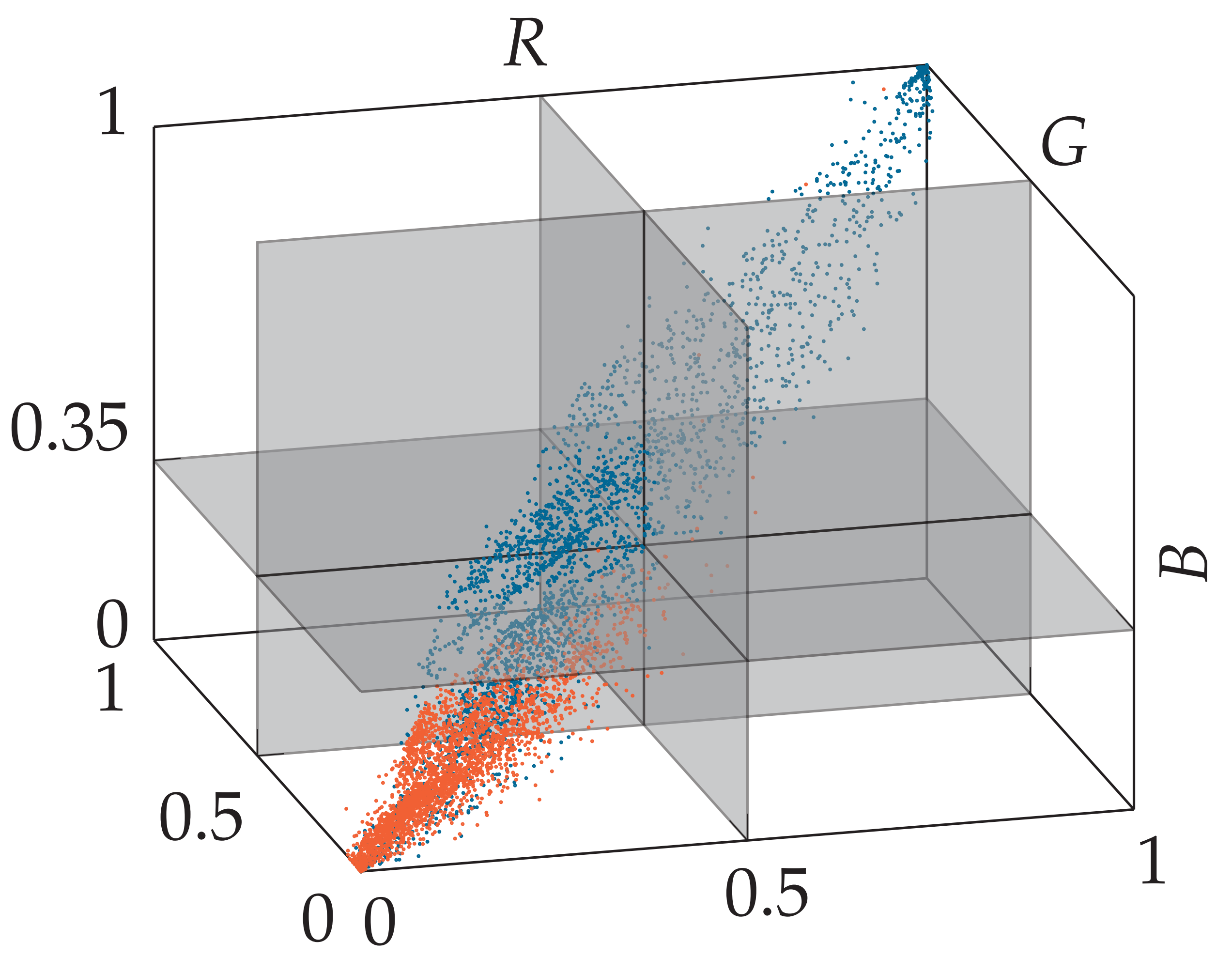


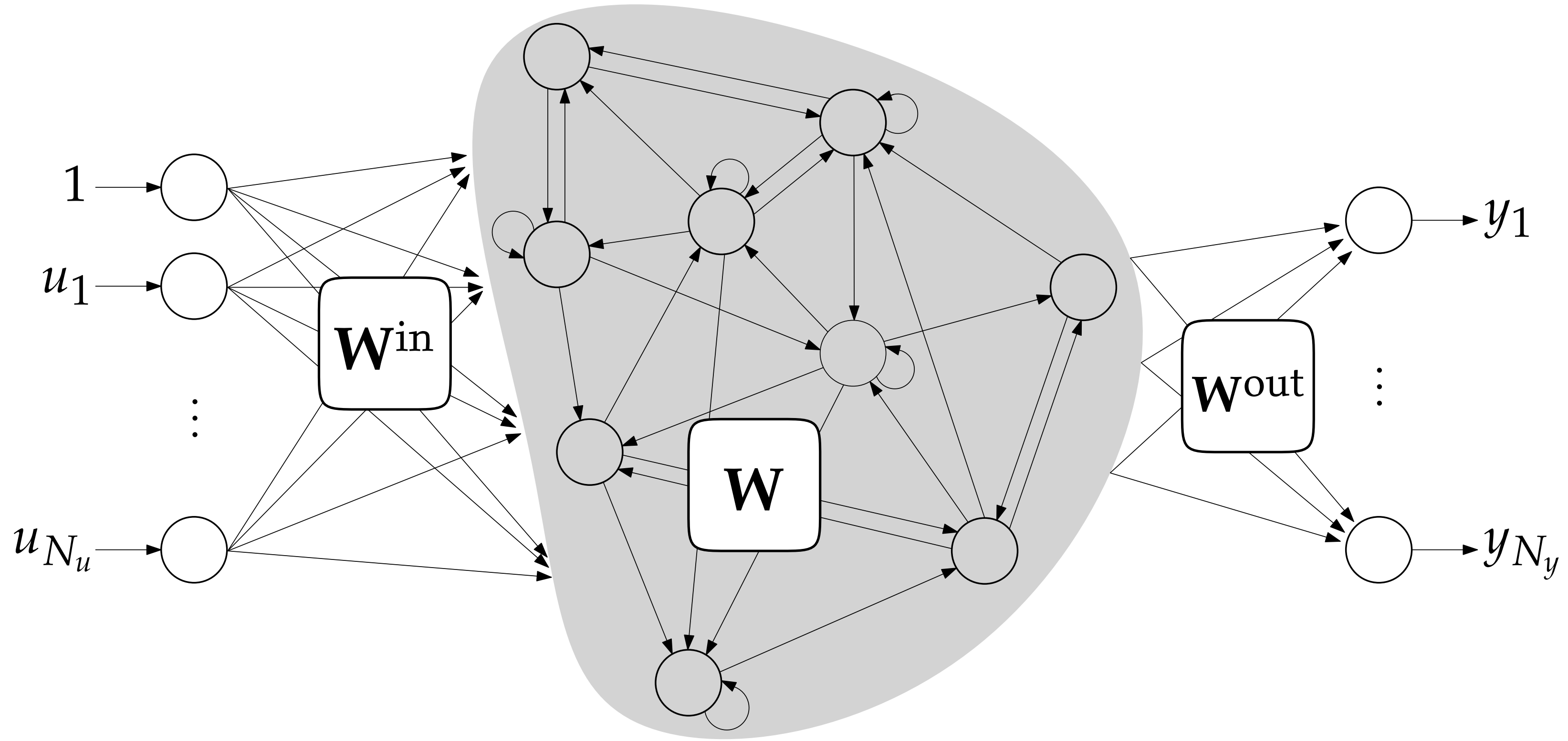
Conclusion



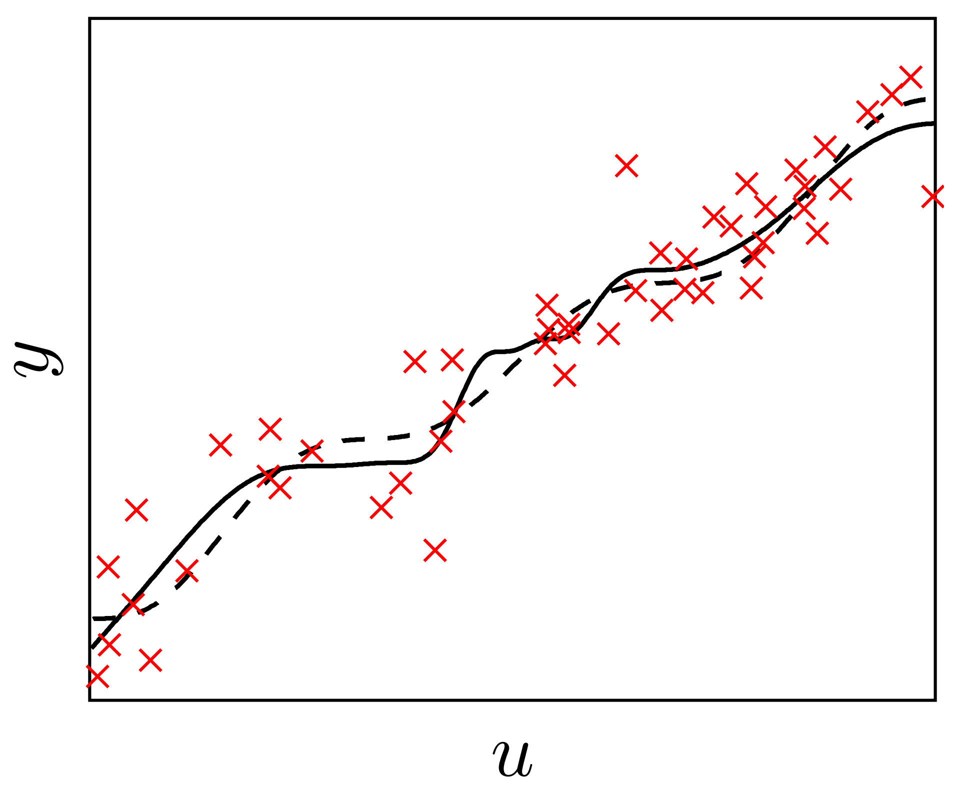
$$\sum_{h=0}^{H} \gamma_h y(t-h) \leq c$$

ESN Training → Linear Least Squares → Quadratic Program





Conclusion



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

ESN Training → Linear Least Squares → Quadratic Program