# ML in the dating of ice cores

## A GRU Method for Automated Annual Layer Identification

Rasmus Arentoft Nielsen

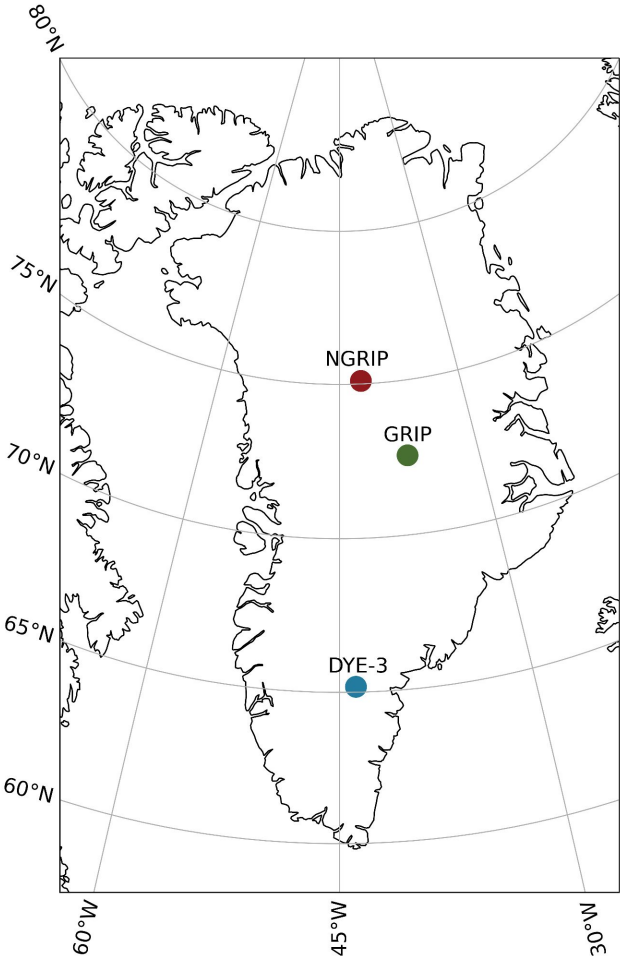
*Supervisors:*
Sune O. Rasmussen
Troels C. Petersen

UNIVERSITY OF COPENHAGEN

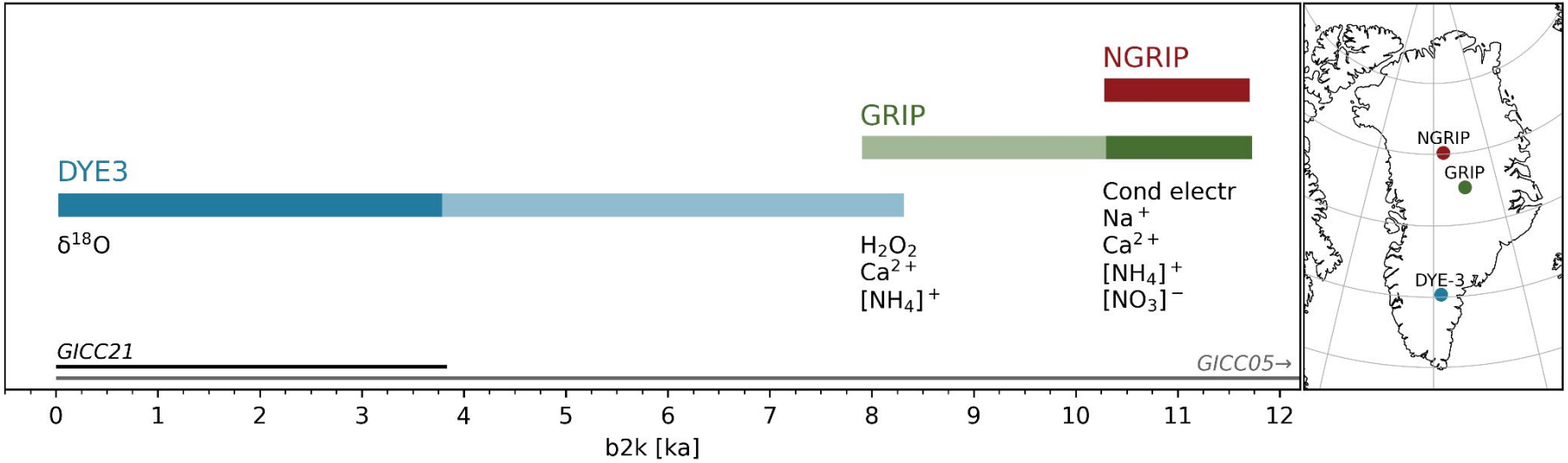
# Outline

- A broad introduction
  Motivation, data and a bit of glaciology
- The model
  GRU model structure and peak detection
- Results
  Examples and overall performance
- Discussion
  What are the limitations of the model?
- Conclusions

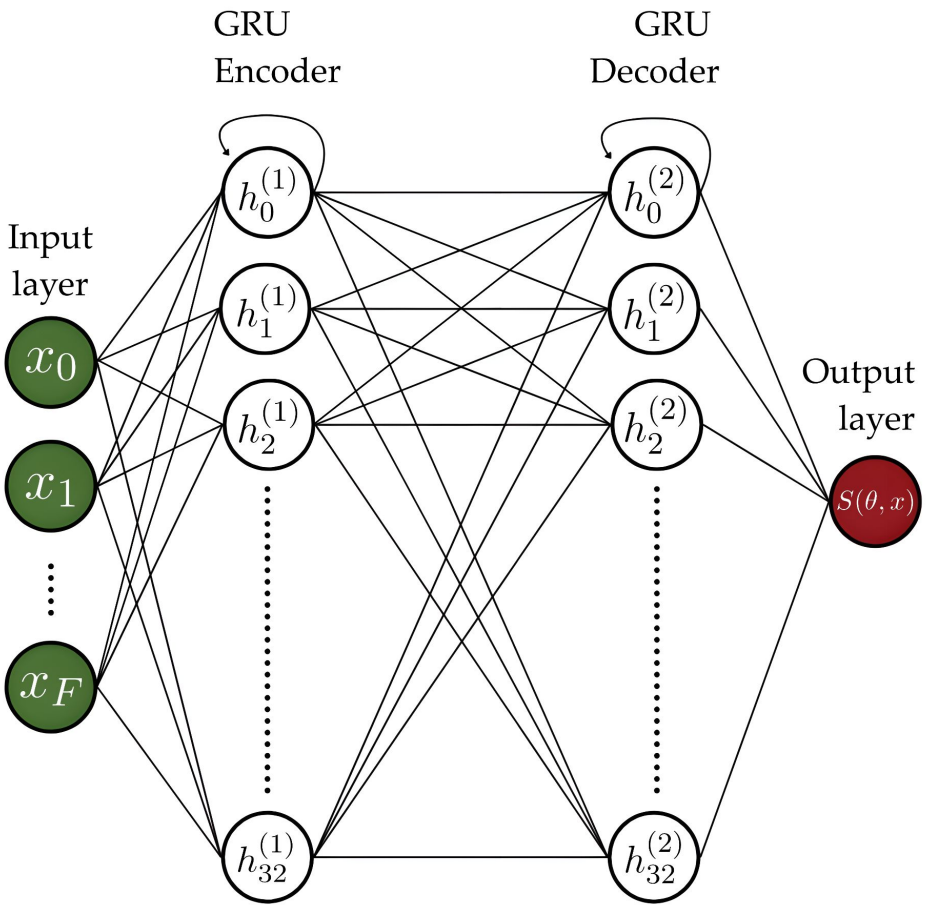
# Introduction



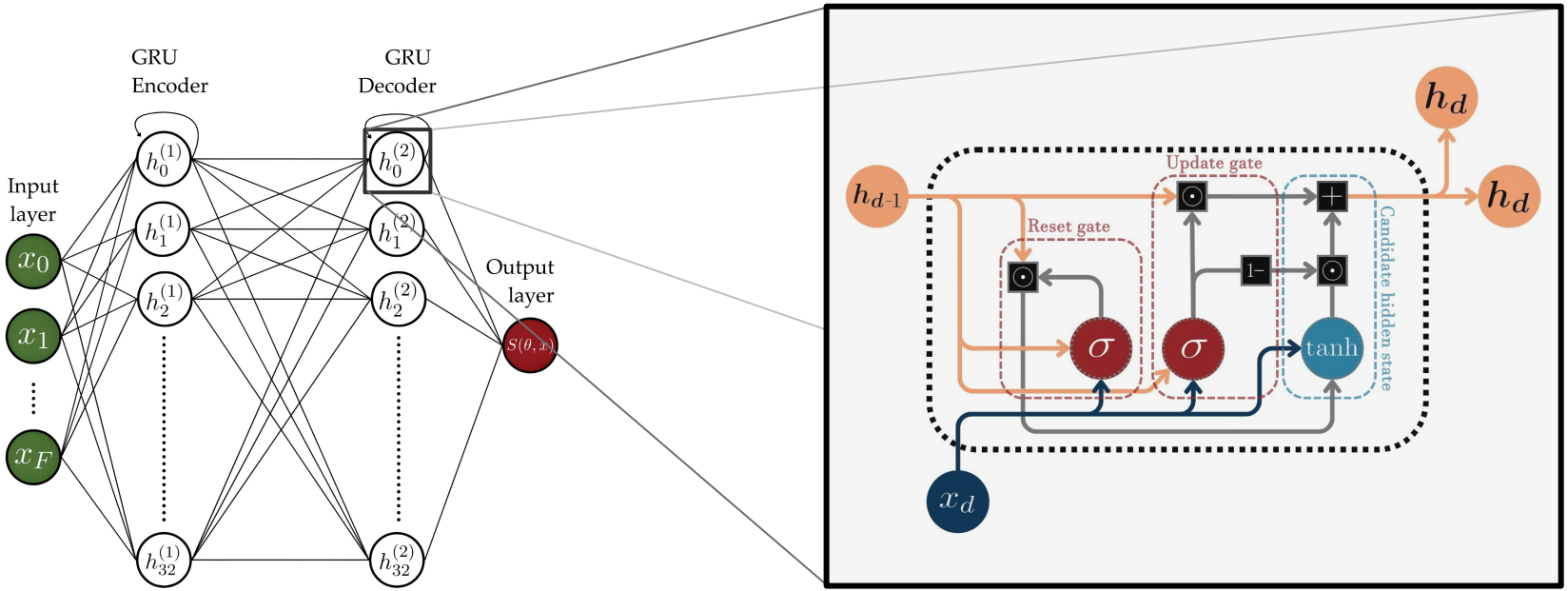
Overview of the temporal ice core data availability and measured features



# GRU Model


GRU
Encoder
GRU
Decoder
Input
layer
Output
layer
$x_0$
$x_1$
$x_F$
$h_0^{(1)}$
$h_1^{(1)}$
$h_2^{(1)}$
$h_{32}^{(1)}$
$h_0^{(2)}$
$h_1^{(2)}$
$h_2^{(2)}$
$h_{32}^{(2)}$
$S(\theta, x)$

# GRU - Looking Inside



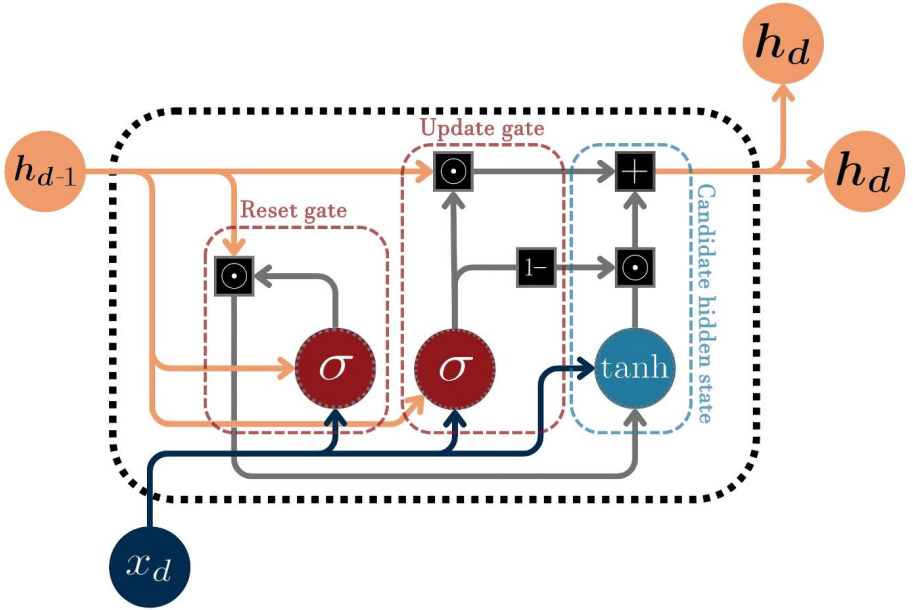
# GRU - Looking Inside

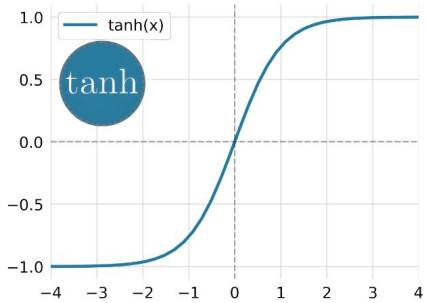
$$\boldsymbol{h}_{\langle d \rangle} = (1 - z_d) \odot \hat{h}_d + z_d \odot \boldsymbol{h}_{\langle d-1 \rangle}$$

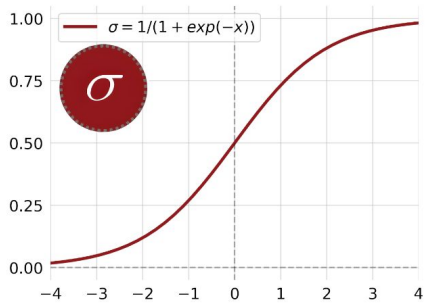
$$\hat{h}_d = \tanh(W_{\hat{h}} x_d + U_{\hat{h}}(r_d \odot h_{d-1}) + b_{\hat{h}})$$

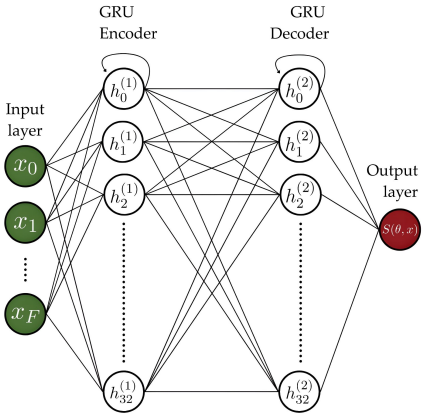
$$z_d = \sigma(W_z x_d + U_z h_{d-1} + b_z)$$

$$r_d = \sigma(W_r x_d + U_r h_{d-1} + b_r)$$

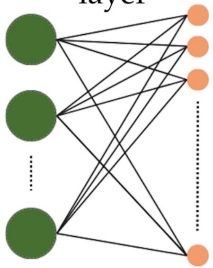


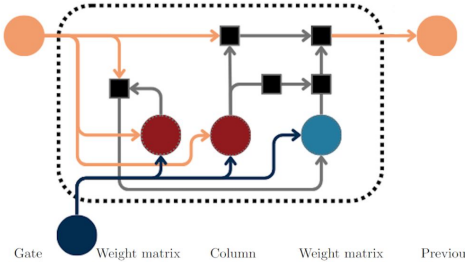





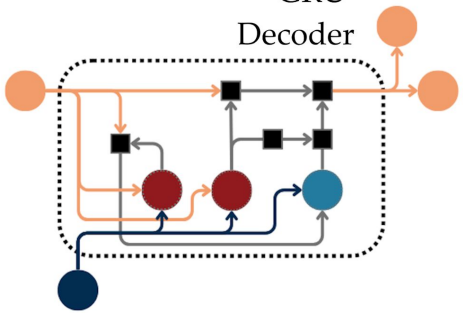

Input layer



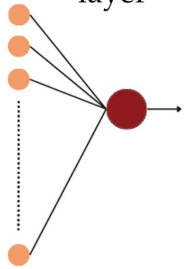
GRU Encoder



GRU Decoder



Output layer



$$\underset{(F \times D)}{\text{Input matrix}}\ \boldsymbol{I} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{F1} & x_{F2} & \cdots & x_{FD} \end{pmatrix}$$

$$\underset{(H \times 1)}{\text{Gate } z,r,\hat{h}} \begin{pmatrix} g_{1,d} \\ g_{2,d} \\ \vdots \\ g_{H,d} \end{pmatrix} = \underset{(H \times F)}{\text{Weight matrix } W} \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1F} \\ w_{21} & w_{22} & \cdots & w_{2F} \\ \vdots & \vdots & \ddots & \vdots \\ w_{H1} & w_{H2} & \cdots & w_{HF} \end{pmatrix} \underset{(F \times 1)}{\text{Column of Input}} \begin{pmatrix} x_{1,d} \\ x_{2,d} \\ \vdots \\ x_{F,d} \end{pmatrix} + \underset{(H \times H)}{\text{Weight matrix } U} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1H} \\ u_{21} & u_{22} & \cdots & u_{2H} \\ \vdots & \vdots & \ddots & \vdots \\ u_{H1} & u_{H2} & \cdots & u_{HH} \end{pmatrix} \underset{(H \times 1)}{\text{Previous hidden state}} \begin{pmatrix} h_{1,d-1} \\ h_{2,d-1} \\ \vdots \\ h_{H,d-1} \end{pmatrix} + \underset{(H \times 1)}{\text{Bias vector(s)}} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_H \end{pmatrix}$$

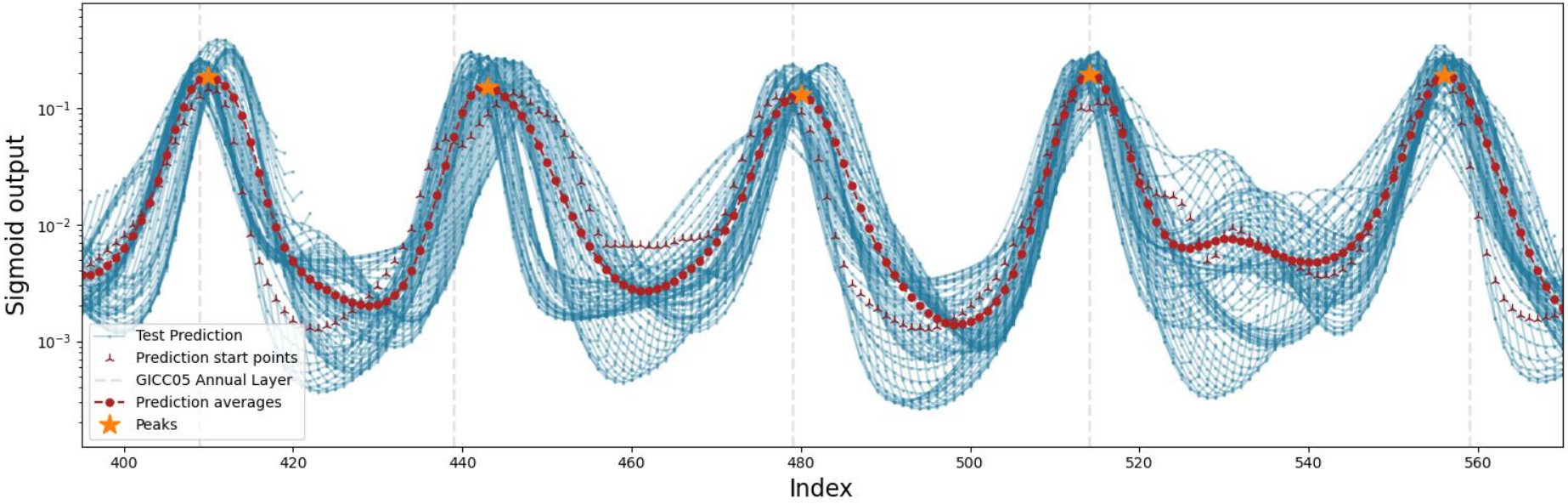
$$\underset{(H \times 1)}{\text{Gate } z,r,\hat{h}} \begin{pmatrix} g_{1,d} \\ g_{2,d} \\ \vdots \\ g_{H,d} \end{pmatrix} = \underset{(H \times 2H)}{\text{Weight matrix } W} \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1,2H} \\ w_{21} & w_{22} & \cdots & w_{2,2H} \\ \vdots & \vdots & \ddots & \vdots \\ w_{H1} & w_{H2} & \cdots & w_{H,2H} \end{pmatrix} \underset{(2H \times 1)}{\text{Summary } c} \begin{pmatrix} c_{1,d} \\ c_{2,d} \\ \vdots \\ c_{2H,d} \end{pmatrix} + \underset{(H \times H)}{\text{Weight matrix } U} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1H} \\ u_{21} & u_{22} & \cdots & u_{2H} \\ \vdots & \vdots & \ddots & \vdots \\ u_{H1} & u_{H2} & \cdots & u_{HH} \end{pmatrix} \underset{(H \times 1)}{\text{Previous hidden state}} \begin{pmatrix} h_{1,d-1} \\ h_{2,d-1} \\ \vdots \\ h_{H,d-1} \end{pmatrix} + \underset{(H \times 1)}{\text{Bias vector(s)}} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_H \end{pmatrix}$$

$$\vec{S} = \underset{(D' \times 2H)}{\text{Output matrix}} \begin{pmatrix} o_{11} & x_{12} & \cdots & x_{1H} \\ o_{21} & x_{22} & \cdots & x_{2H} \\ \vdots & \vdots & \ddots & \vdots \\ x_{D'1} & x_{D'2} & \cdots & x_{D'H} \end{pmatrix} \underset{(2H \times 1)}{\text{Weight vector}} \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_H \end{pmatrix} + b = \underset{(D' \times 1)}{\text{Sigmoid output}} (S_1\ S_2 \cdots S_{D'})$$

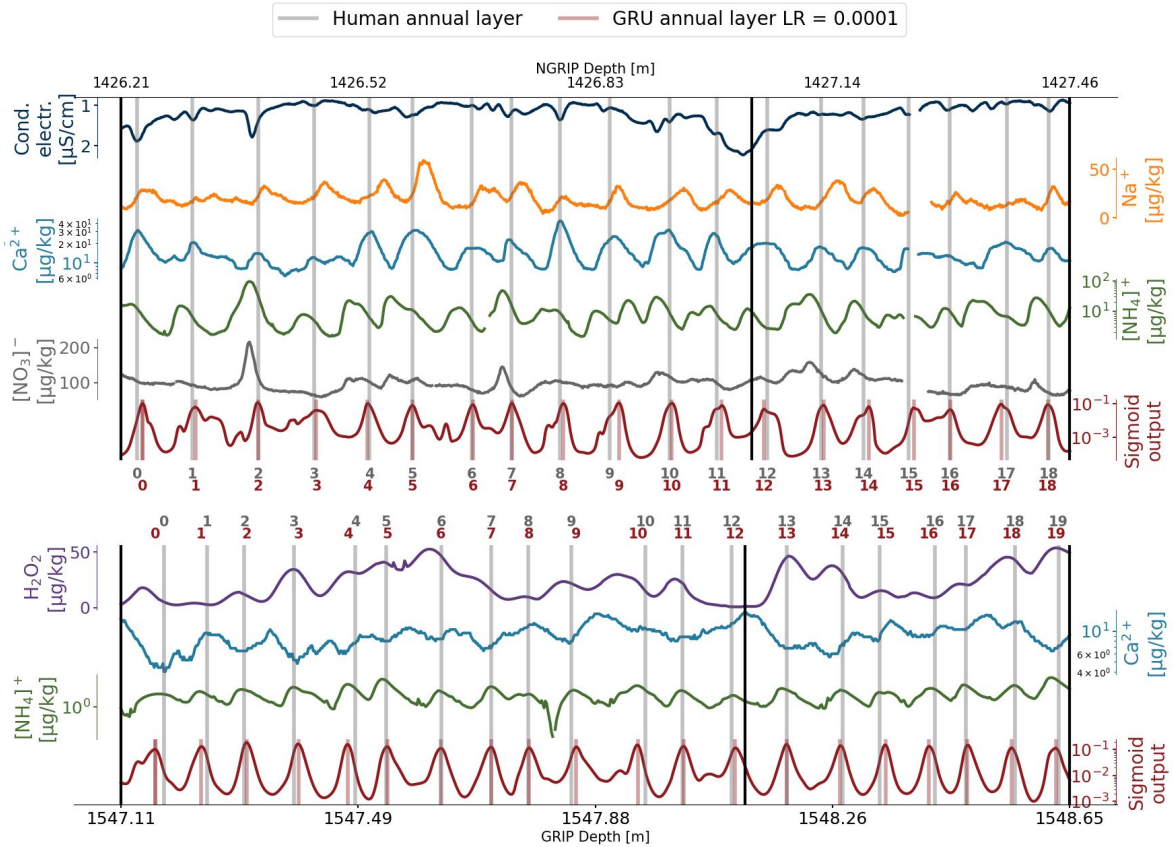
# GRU Model - Setup

| | |
|---|---|
| Model type | Bidirectional GRU Encoder-Decoder |
| Encoder | 32 neurons, return_sequences = False |
| Decoder | 32 neurons, return_sequences = True |
| Optimizer | Adam, loss = ‘binary_crossentropy’ (BCE) |
| Window size | 1-2 years (35-140 samples for GRIP/NGRIP) in GICC05 timescale |
| Learning rate | Typically [5e-4, 1e-4, 5e-5] |

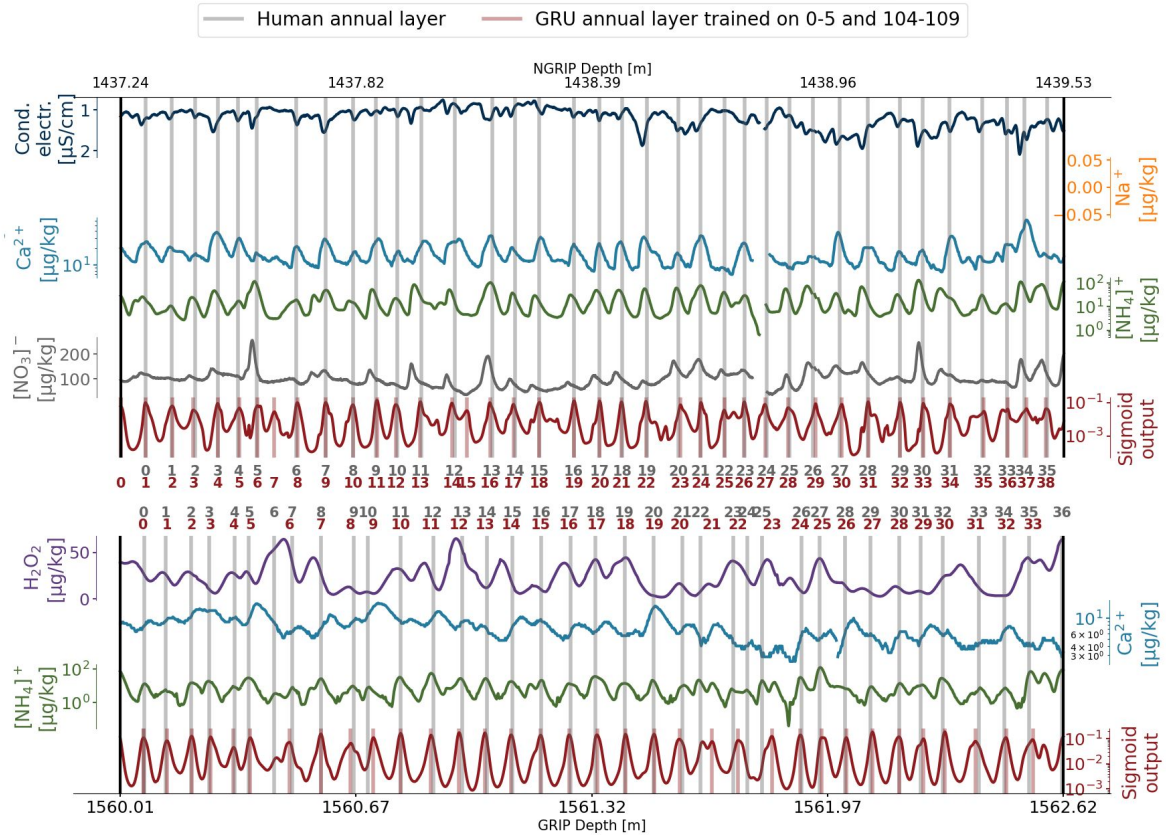
# GRU Model - Making predictions



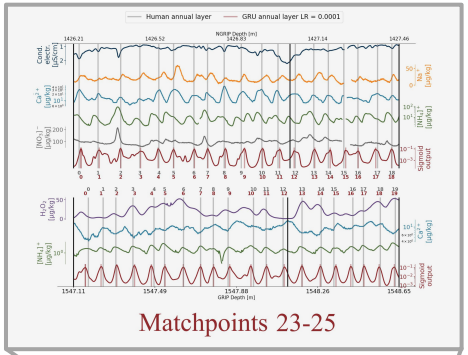
# Results



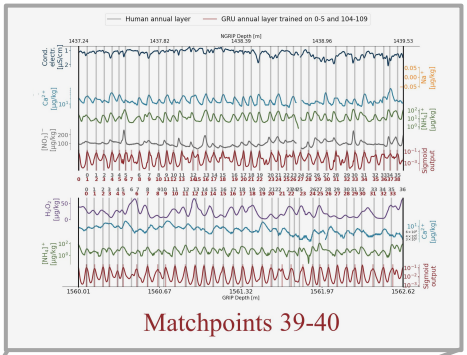

Matchpoints 23-25




# Matchpoints 39-40

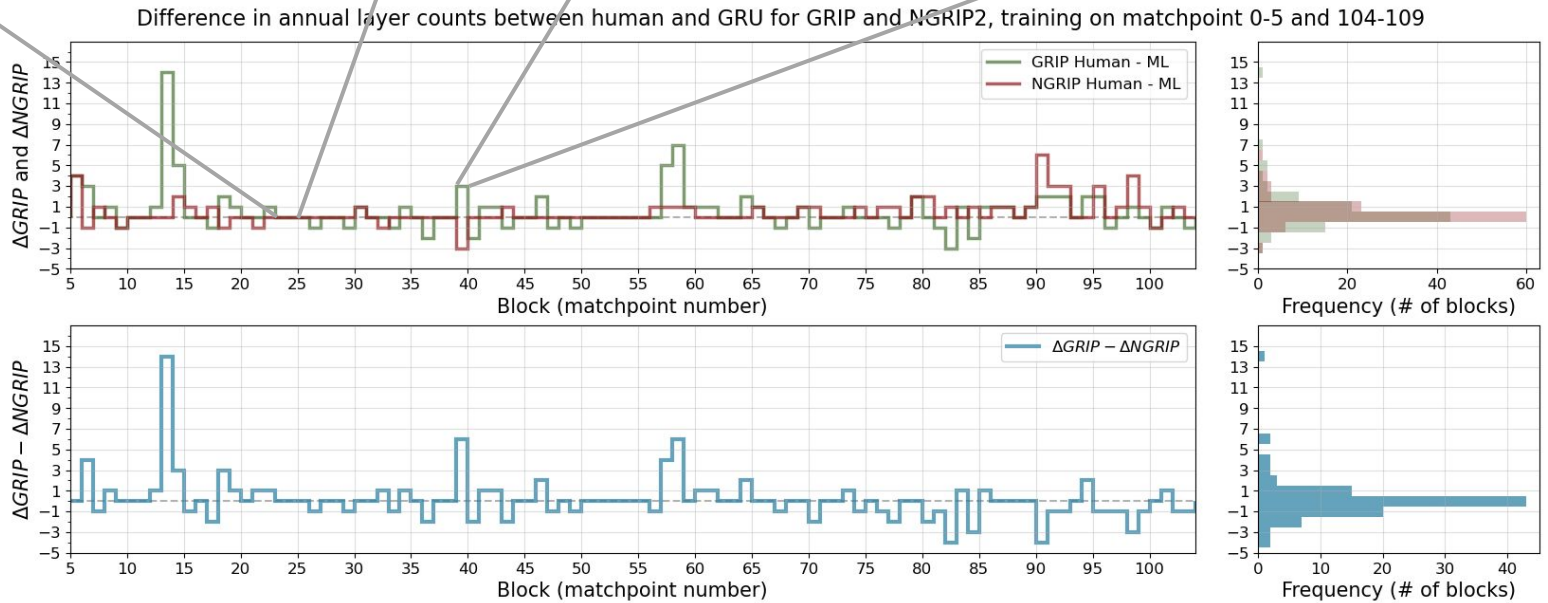



Matchpoints 23-25

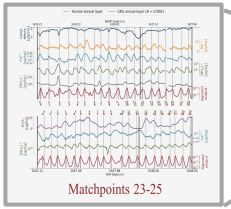


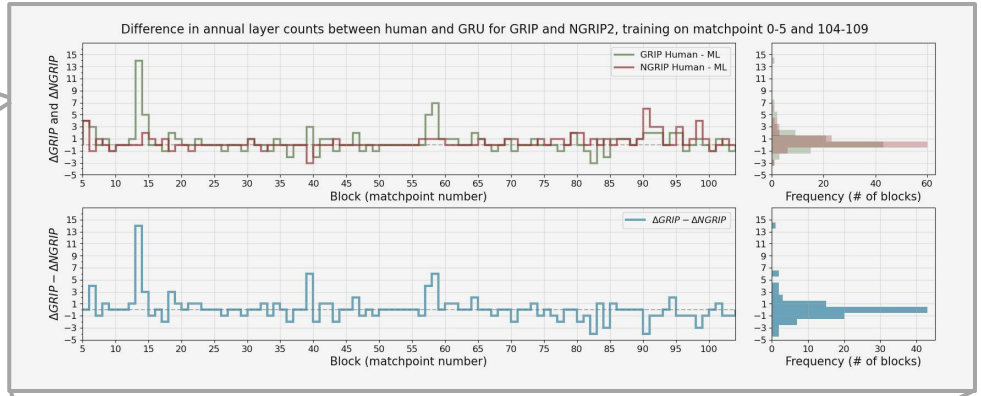

Matchpoints 39-40

Within each block, 89.9% and 79.8% of GRU counts are within 1 of the GICC05 count for NGRIP and GRIP, respectively.



# Results




| Run | GRU counting | GICC counting | Difference | Summary |
|---|---|---|---|---|
| NGRIP $K = 2$ | 1395 | 1394 | 0.07% | Figure 20 |
| NGRIP $K = 3$ | 1354 | 1394 | 2.87% | Figure 29 |
| GRIP $K = 2$ | 1353 (1382) | 1395 | 3.01% (0.93%) | Figure 20 |
| GRIP $K = 3$ | 1355 (1384) | 1395 | 2.86% (0.78%) | Figure 29 |
| NGRIP TOE | 1215 | 1259 | 3.49% | Figure 31 |
| GRIP TOE | 1204 (1233) | 1259 | 4.36% (2.06%) | Figure 31 |
| DYE-3 $K = 2$ | 3724 | 3814 | 2.39% | Figure 33 |
| DYE-3 $K = 3$ | 3749 | 3814 | 1.70% | Figure 33 |
| DYE-3 'TOE' | 3439 | 3414 | 0.73% | Figure 37 |

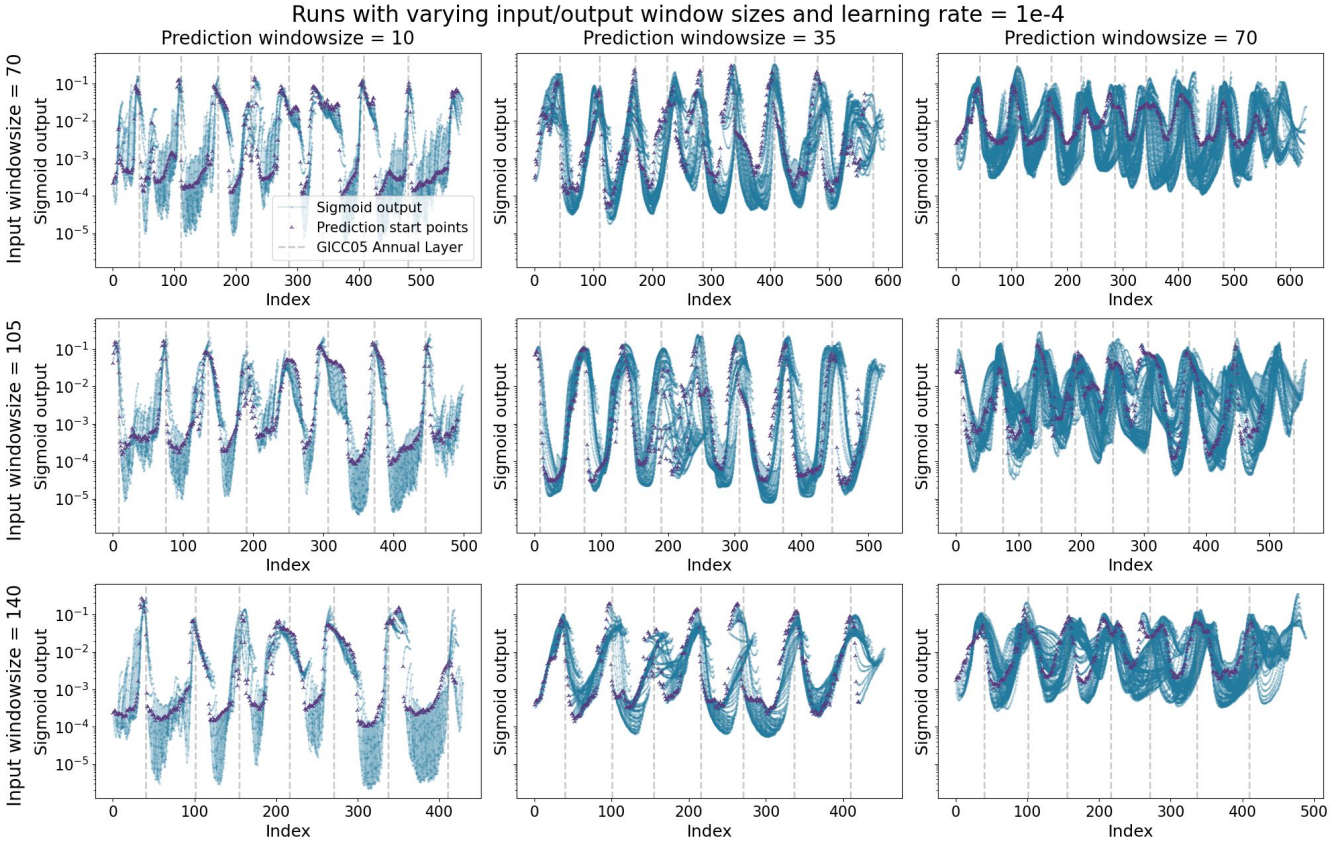
**Table 4:** Summary of the overall model performance for each run. Note that this is merely differences for the total count in the existing timescales versus the model predictions. A better overview of the performance can be found in the figures referenced in the rightmost column.

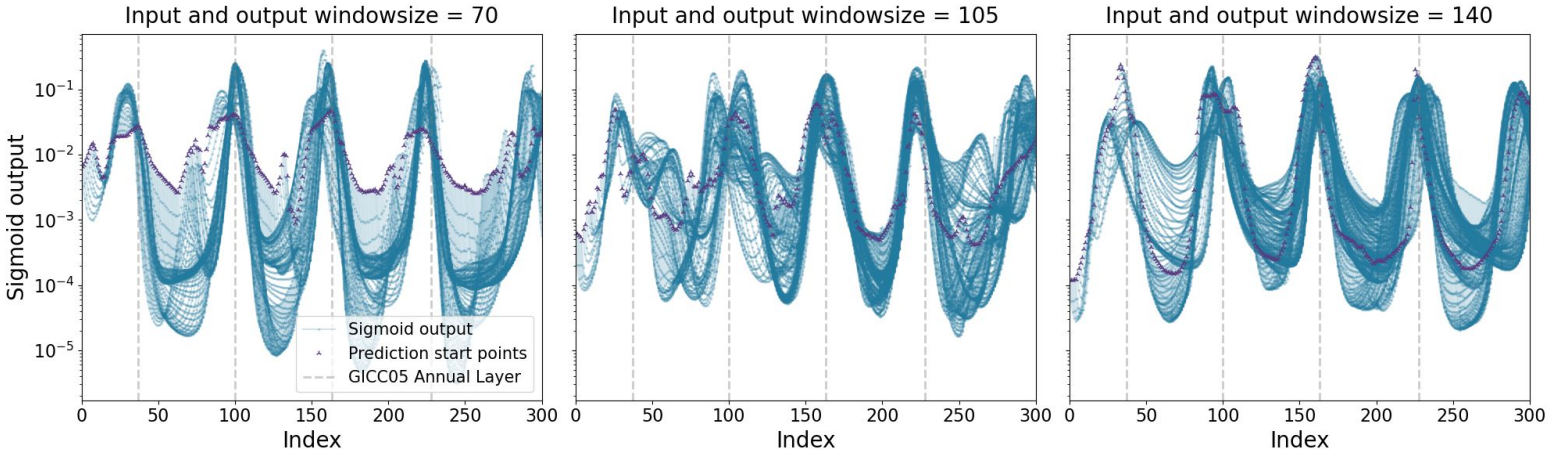
# Discussion

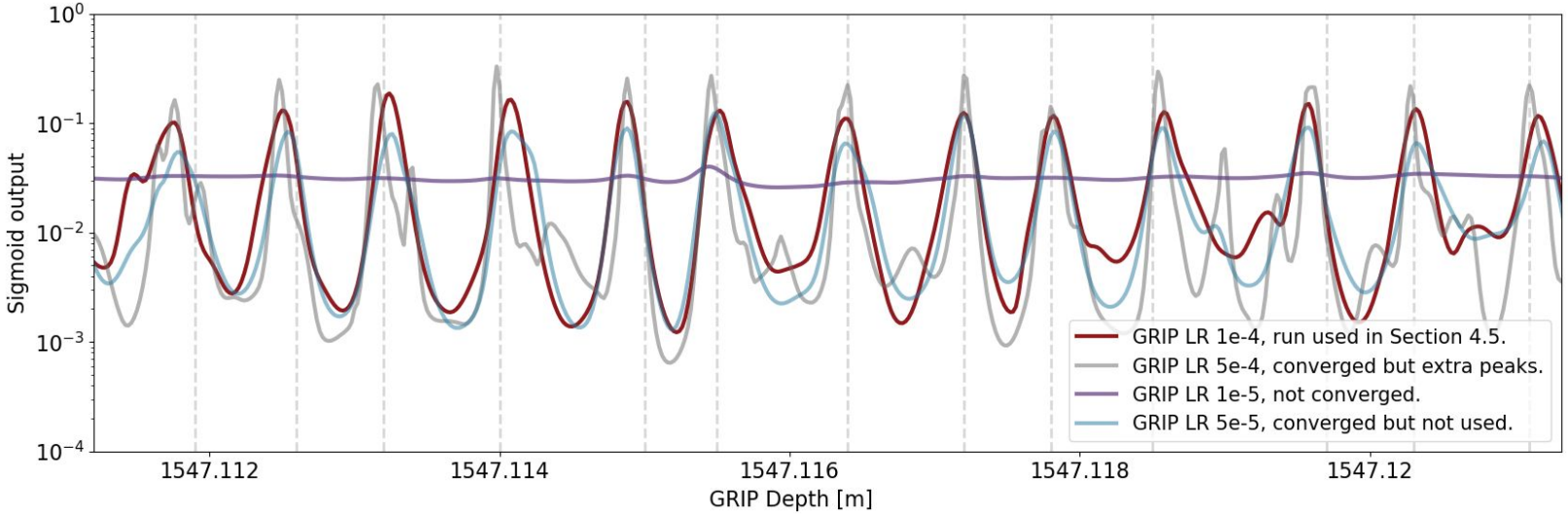
- Training and evaluation is done using manually identified annual layers
- Model assumes equidistant time-series input, but the data is a depth-series.
- Peak detection is sensitive to used parameter values.
- Training and testing on shorter sections of cores with similar characteristics is recommended.

# Conclusion

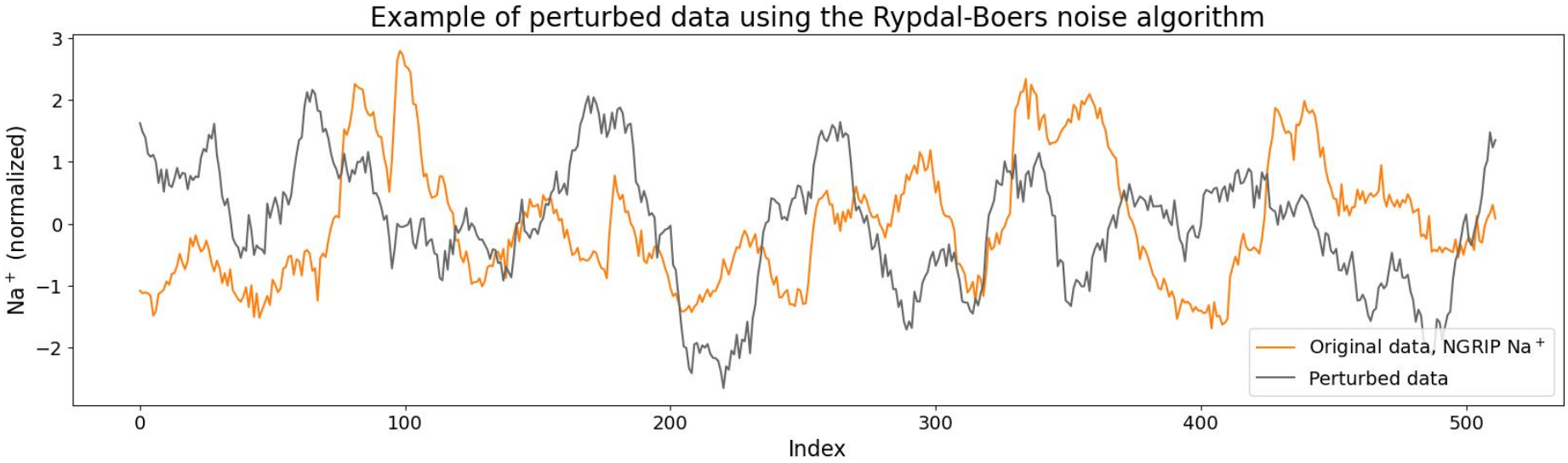
# Appendix A - Hyperparameter tuning



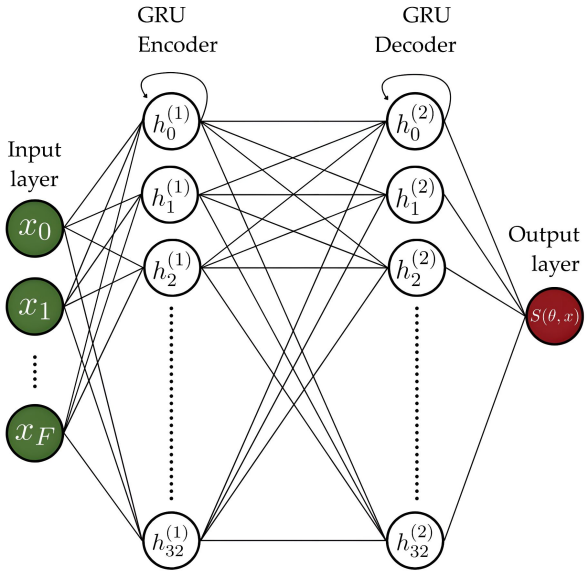

Input and output windowsize = 70
Input and output windowsize = 105
Input and output windowsize = 140
Sigmoid output
Index
Sigmoid output
Prediction start points
GICC05 Annual Layer

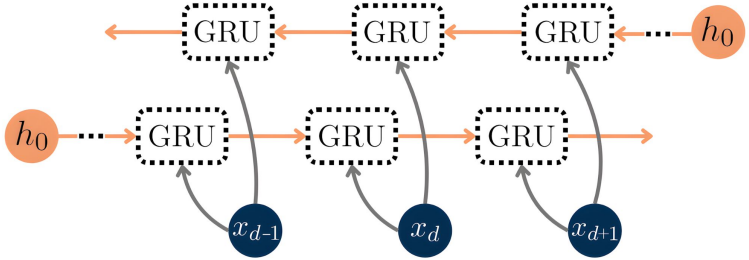

Sigmoid output
$10^{0}$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-4}$
1547.112
1547.114
1547.116
1547.118
1547.12
GRIP Depth [m]
GRIP LR 1e-4, run used in Section 4.5.
GRIP LR 5e-4, converged but extra peaks.
GRIP LR 1e-5, not converged.
GRIP LR 5e-5, converged but not used.

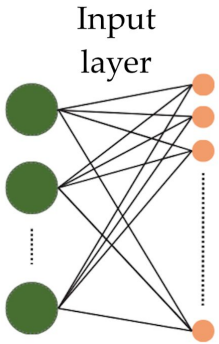
# Appendix B - Perturbed data for feature importance

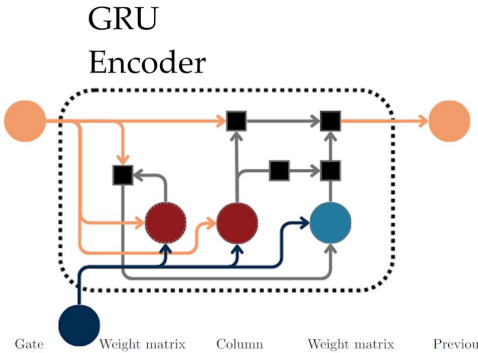


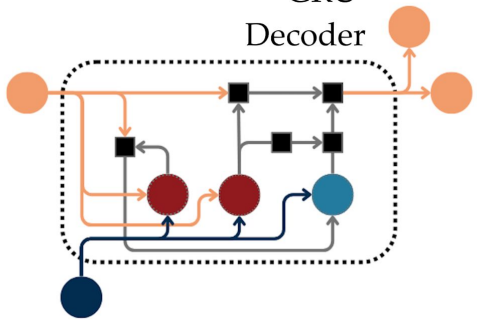
# Appendix C - Model

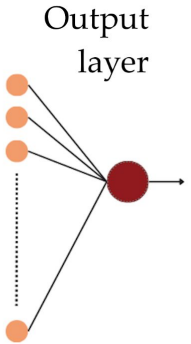


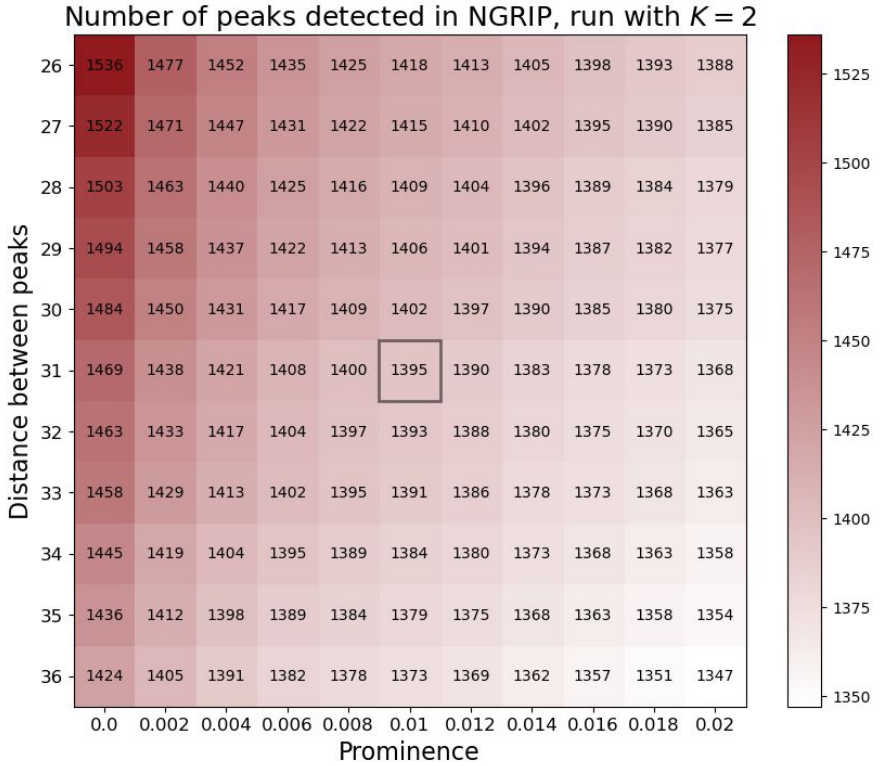



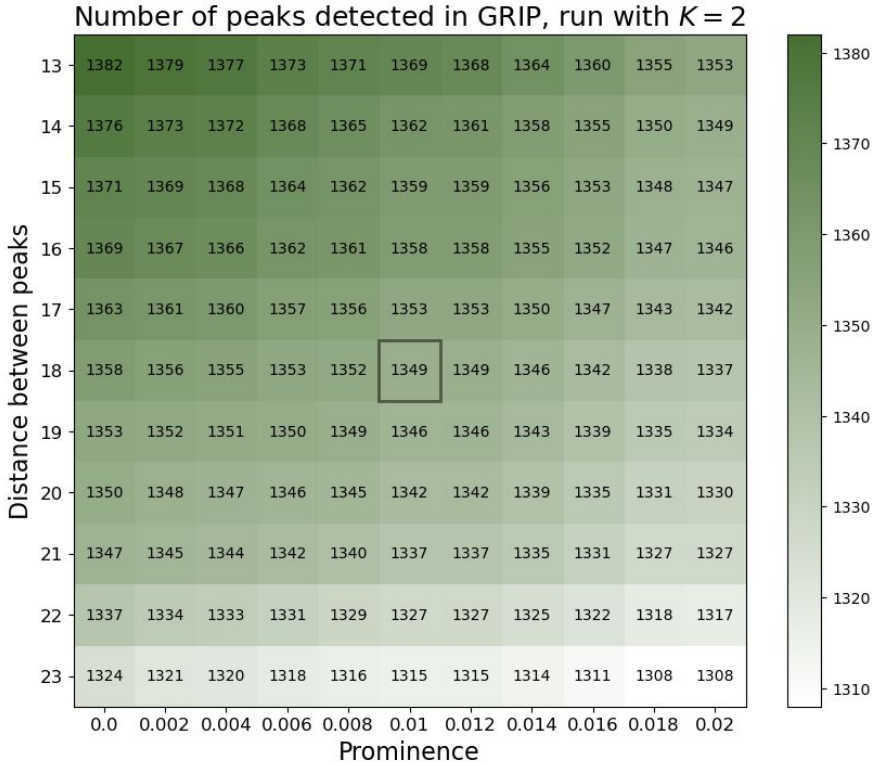



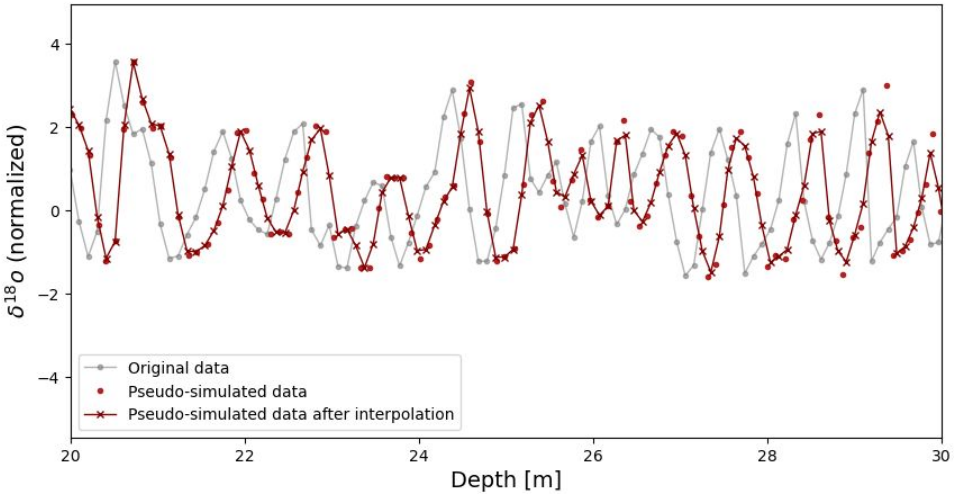



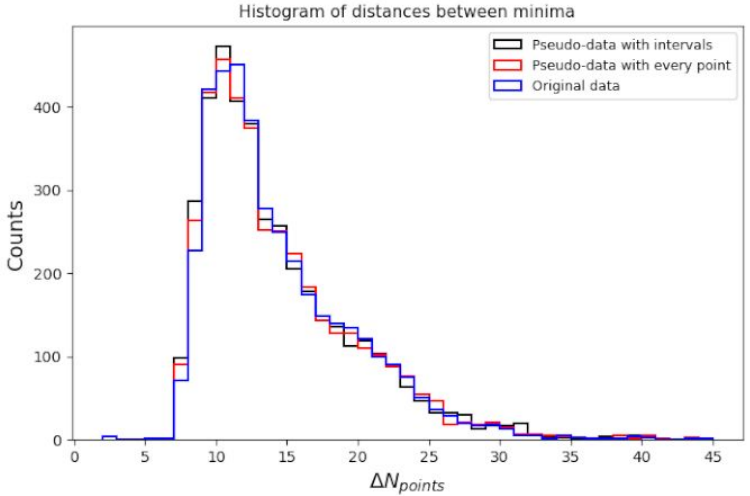
# Appendix D - Sensitivity of peak detection



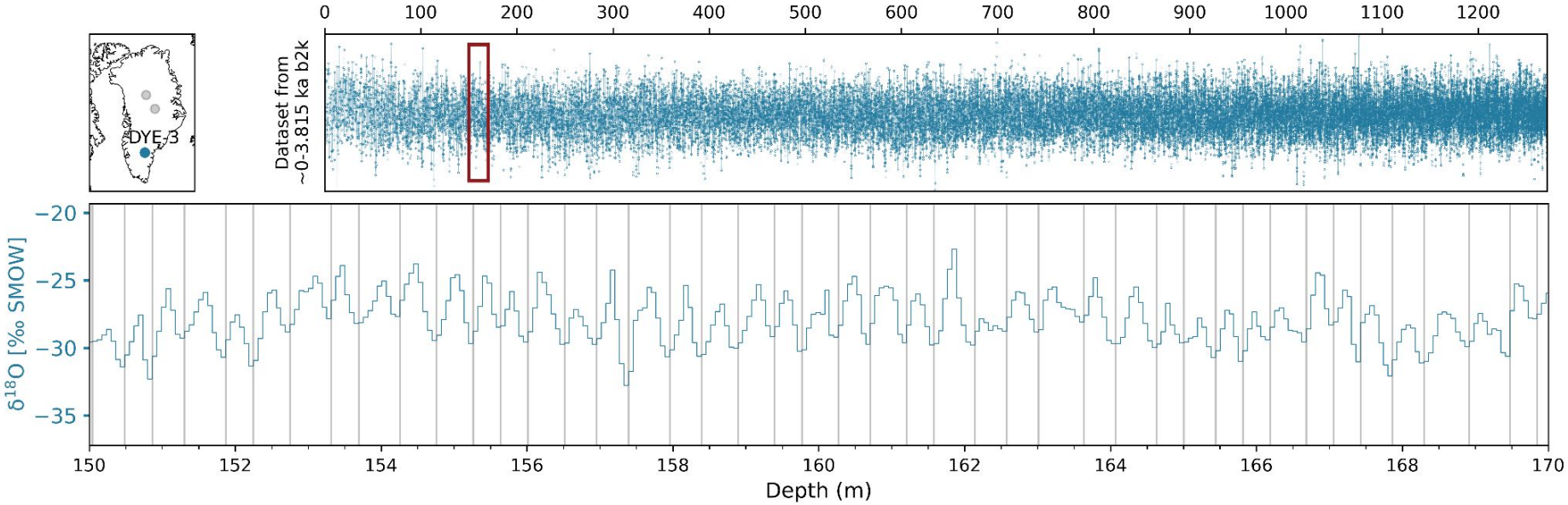



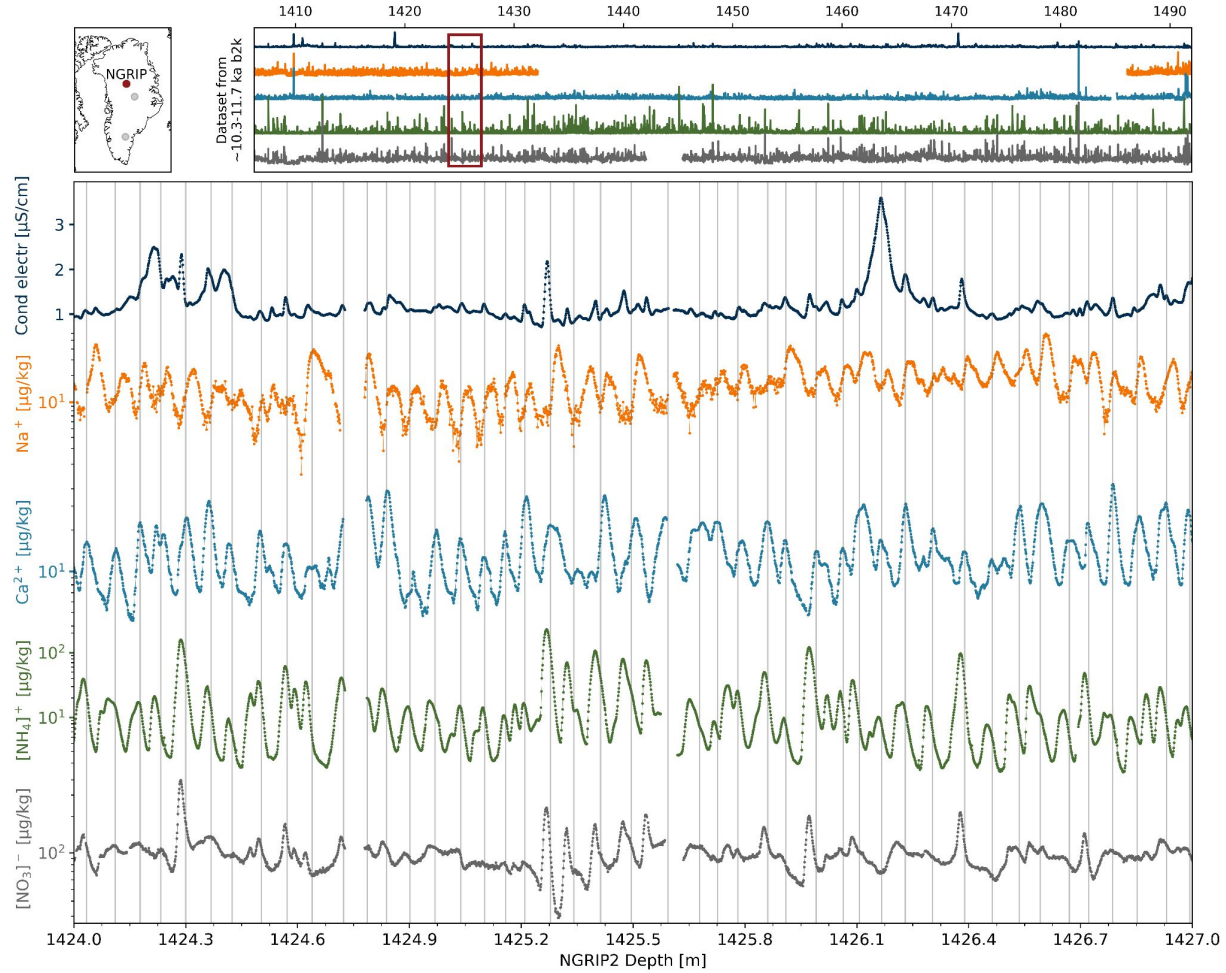
# Appendix E - Pseudo-data

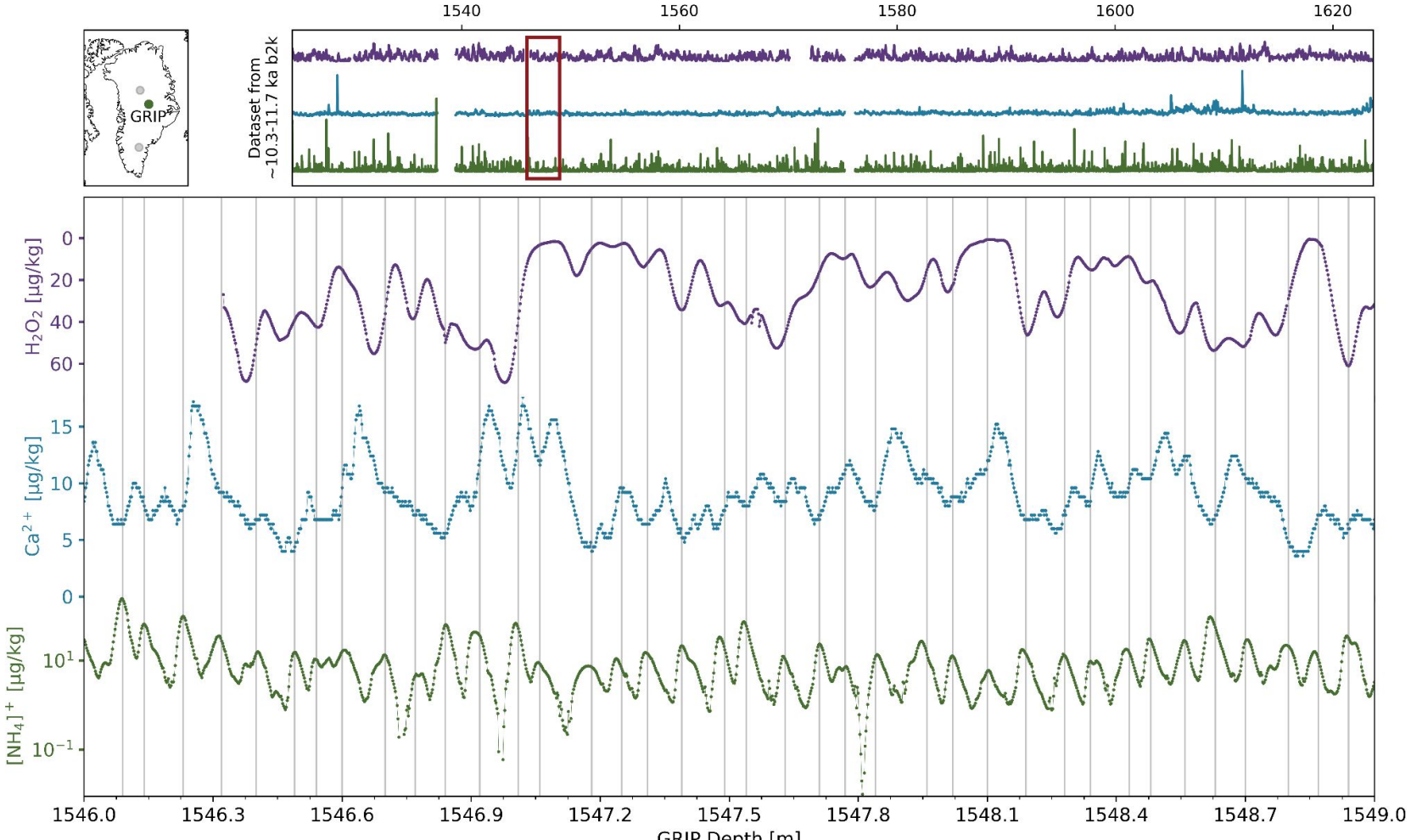


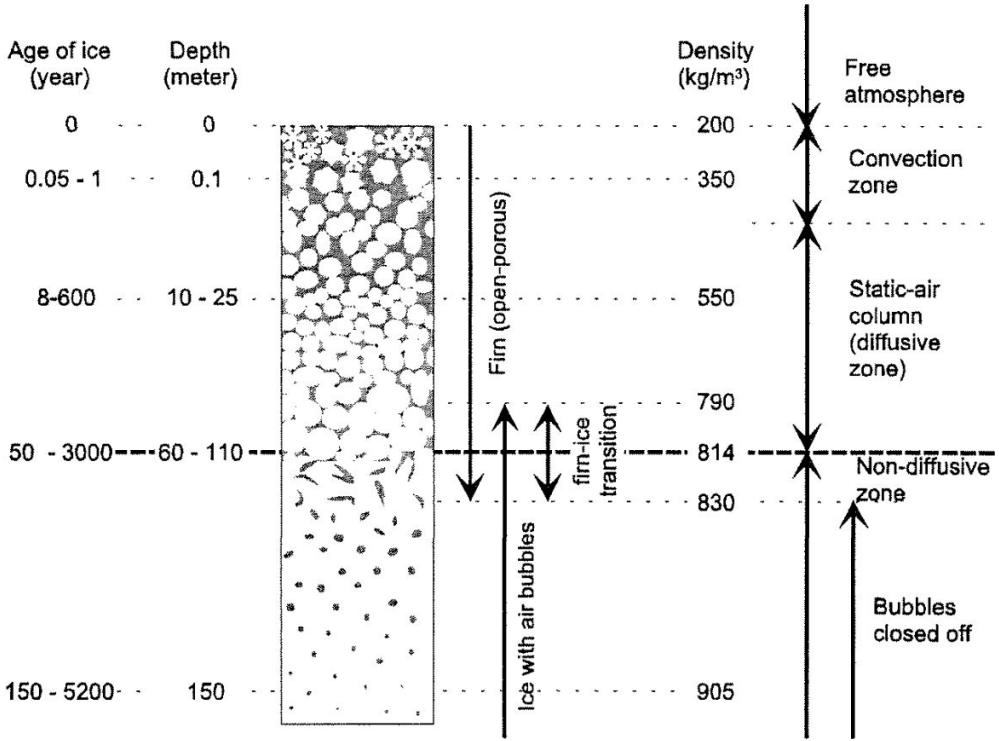
# Appendix F - Data




NGRIP
Dataset from ~10.3-11.7 ka b2k
1410
1420
1430
1440
1450
1460
1470
1480
1490
Cond electr [µS/cm]
Na+ [µg/kg]
Ca2+ [µg/kg]
[NH4]+ [µg/kg]
[NO3]− [µg/kg]
1424.0
1424.3
1424.6
1424.9
1425.2
1425.5
1425.8
1426.1
1426.4
1426.7
1427.0
NGRIP2 Depth [m]


GRIP
Dataset from ~10.3-11.7 ka b2k
1540
1560
1580
1600
1620
$H_2O_2$ [µg/kg]
0
20
40
60
$Ca^{2+}$ [µg/kg]
15
10
5
0
$[NH_4]^+$ [µg/kg]
$10^1$
$10^{-1}$
1546.0
1546.3
1546.6
1546.9
1547.2
1547.5
1547.8
1548.1
1548.4
1548.7
1549.0
GRIP Depth [m]

# Appendix G: Glaciology



# Appendix H: Annual Layer Thicknesses