Application of Deep learning (DL) in the Gulf of St. Lawrence and Estuary

François Roy[(1)], Gregory Smith[(1)], Jean-Philippe Paquin[(1)], Simon St-Onge Drouin[(2)], Charlie Hébert-Pinard[(1)]

1 – Environment and Climate Change Canada (ECCC)

2- Department of Fisheries and Oceans (DFO) Canada

Topics

- Context and objectives
  - Gulf of St. Lawrence (GSL) circulation, tides, flow separation, challenges
- Datasets (ocean model)
- Machine Learning (ML) models
  - LST_{bt}, Anemoi
  - List of training/inference experiments
- ML results
  - 24h ML forecast compared to « ocean model truth »
- Conclusions and next steps

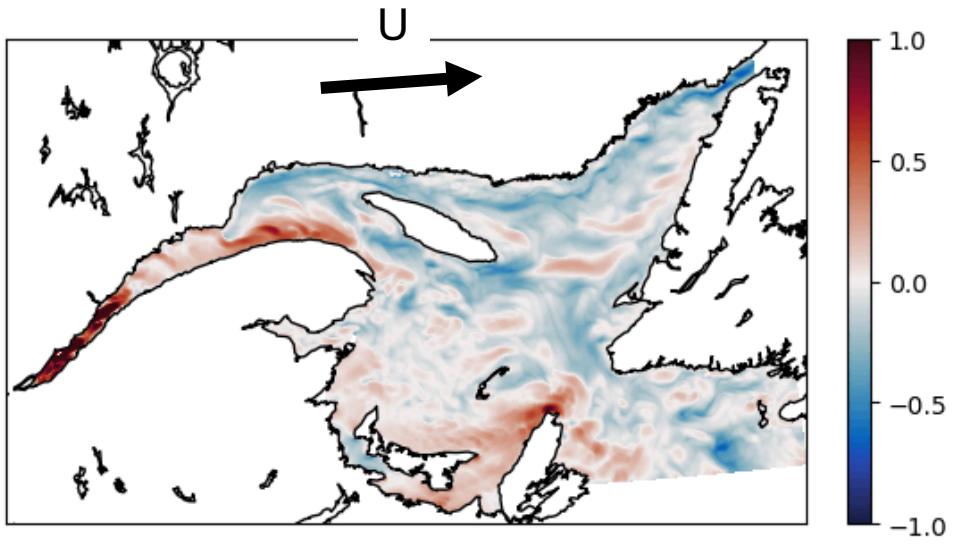
Context and objectives

- GFD ML applications with hourly time steps and tides, not common
- The coastal environment of the Gulf of St. Lawrence is relatively complex, ML training design is not straightforward
  - tides, estuarine circulation, strong seasonal cycles, evolving stratification at different time scales, steep topography features, sea ice, internal waves, …
- Explore how ML techniques can emulate sea surface height and currents in a coastal environment, including tides
- Explore and test barotropic/baroclinic separation to address the issue of limited information from 3D outputs (footprints of HF in geophysical fields ➔ tide aliasing …)

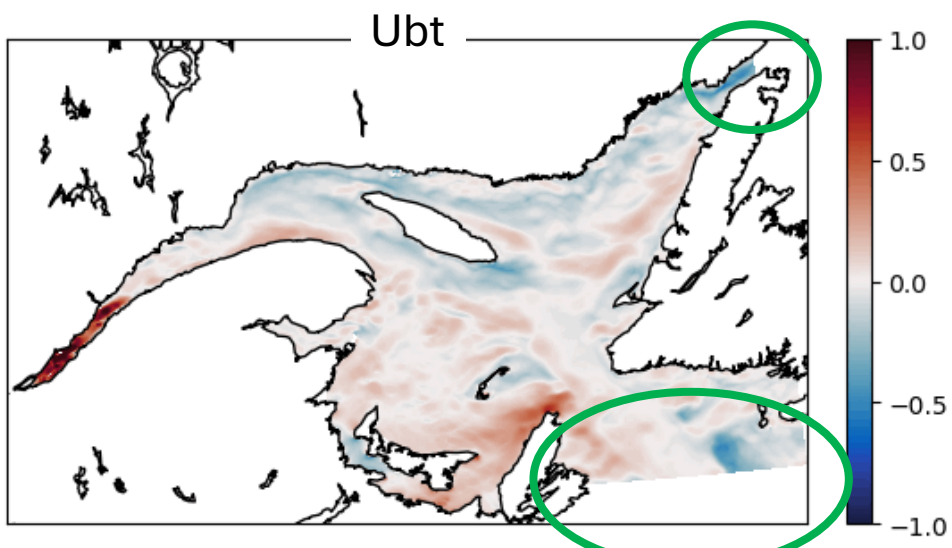
Physical model outputs:
training dataset (hourly)

A
Surface current (U)

Total



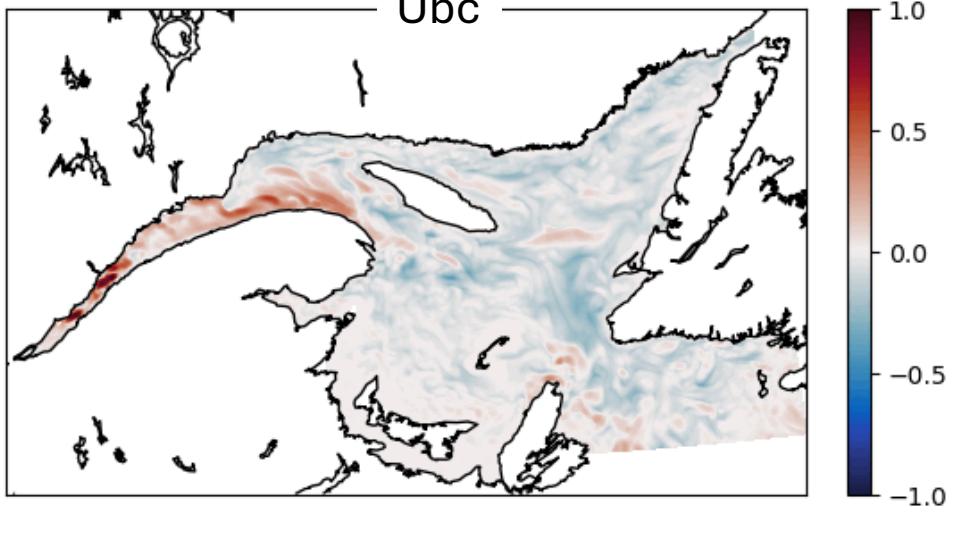
Tides, Storm surge f(SSH_{BDY}, Ubt_{BDY}, Vbt_{BDY}), remote



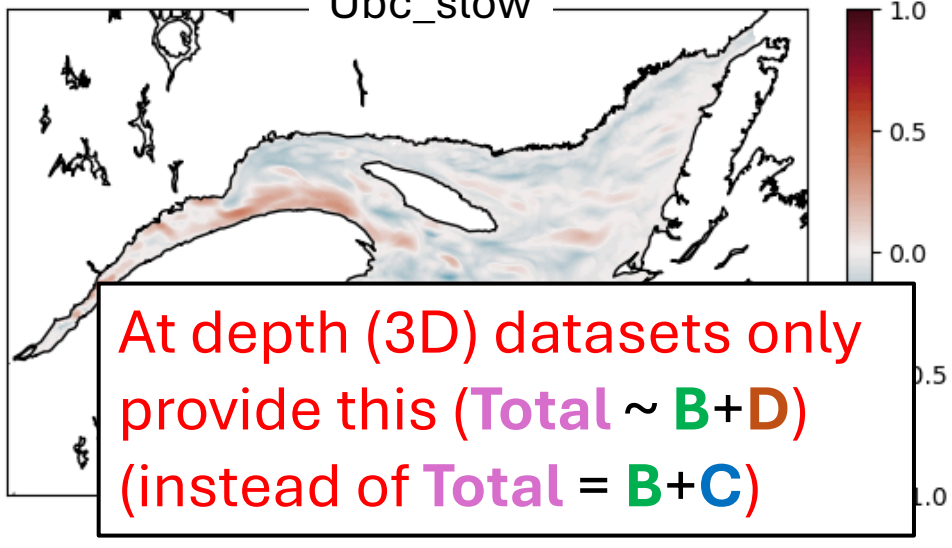
B
Barotropic component
(U avg over water column, z)

C
Surface baroclinic residual (U)
(**A** – **B**)

Tides (local rectification), f(stratification, bathymetry), Surface winds more locally driven



Slow baroclinic circulation f(large scale density flow)

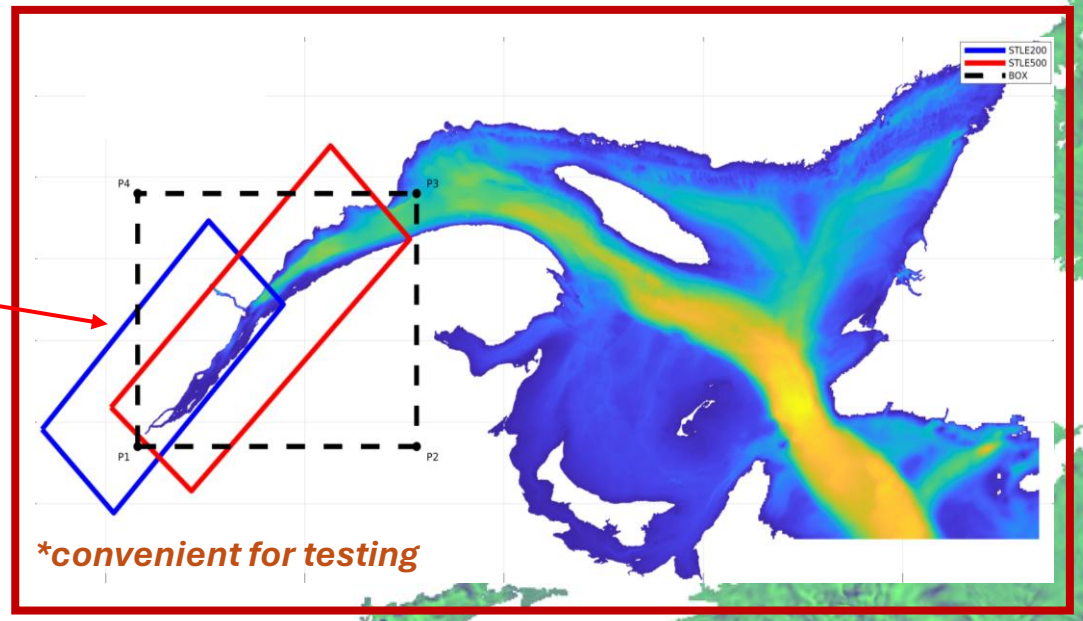


D
Surface baroclinic residual (**A** – **B**)**
**From daily avg outputs

At depth (3D) datasets only provide this (Total ~ B+D) (instead of Total = B+C)

Datasets (ocean models)

Gulf of St. Lawrece (GSL): domain of interest*



*convenient for testing

Future downscaling applications No shown today

CIOPS-E (1/36 ~2km)

Canadian **I**ce-**O**cean
Prediction **S**ystem
East (operational)

Paquin, JP., Roy, F., Smith, G.C. *et al.* A new high-resolution Coastal Ice–Ocean Prediction System for the East Coast of Canada. *Ocean Dynamics* 74, 799–826 (2024). https://doi.org/10.1007/s10236-024-01634-7

30 years hindcast dataset

Internal wave surface footprint from steric height

Coverage

- Ocean model (dataset)
  - CIOPS-E (1/36, ~2km) (1994-2024)
    - Hourly: 2D surface and vertically integrated variables
    - Daily mean: other 3D variables (not yet used)

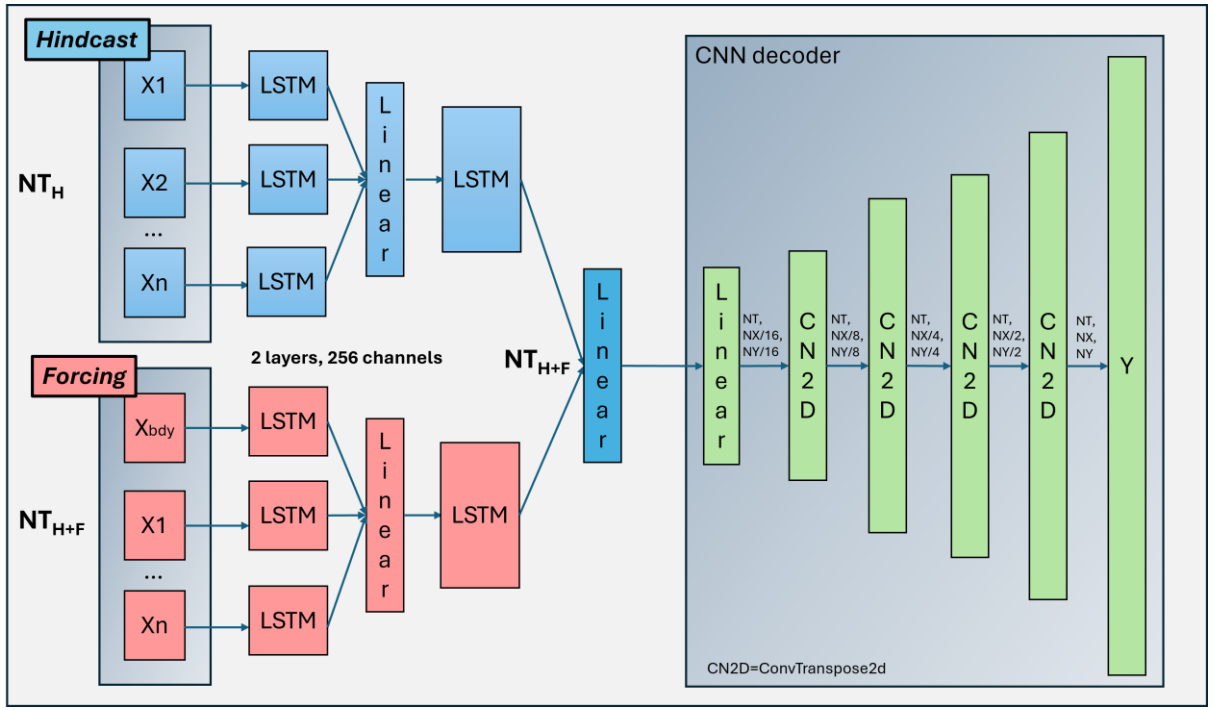
- Daily forecast with ML emulators (2022-2023)
  - 24h surface forecast (including tides) ➔ Gulf of St. Lawrence (GSL)

- Training/inference setup
  - 1994-2016 (training, 23 years)
  - 2017-2021 (validation, 5 years)
  - 2022-2023 (test, 2 years)

ML models

LST_{bt} *(for barotropic)*

(in house)



In house

LSTM+CNN decoder

Time resolution focused (degraded inputs)

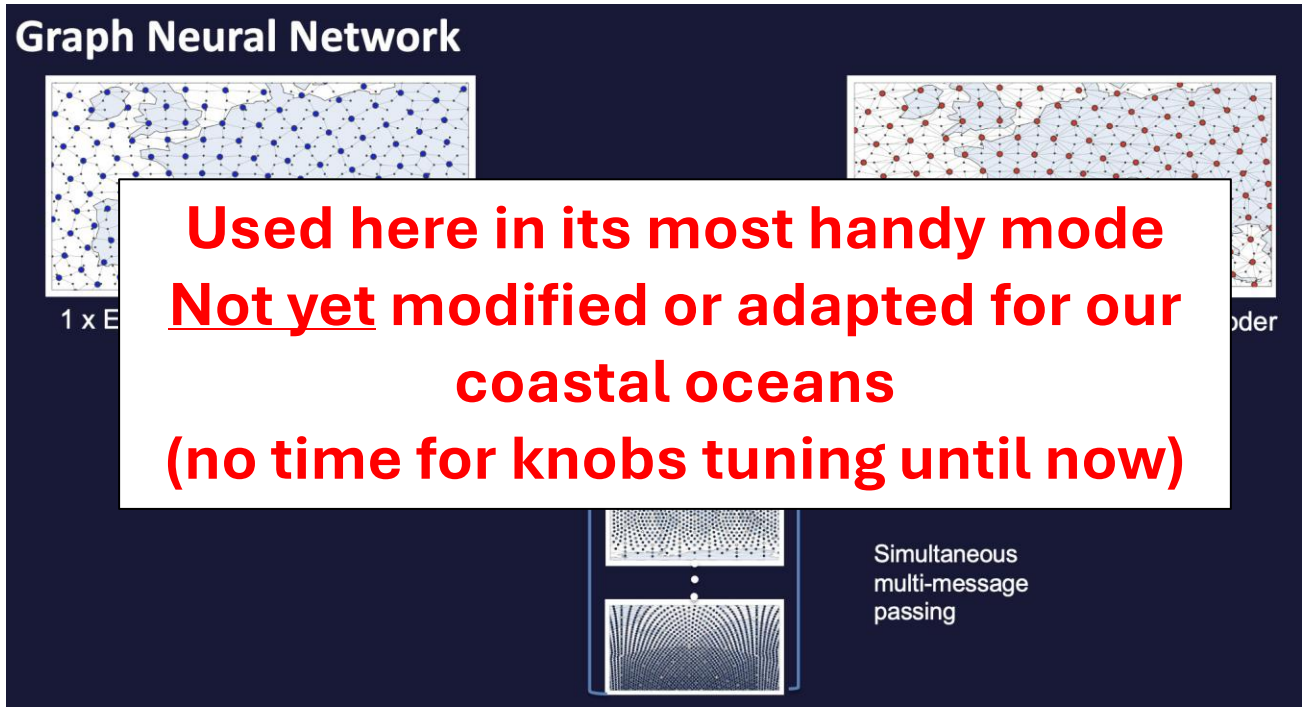
Not auto-regressive (friendly)

Heavy on memory (model size)

single-gpu

Anemoi

(ECMWF)



Community supported

Encoder – (GNN, Graphtransformer, Transformer w flash attention)) - Decoder

Spatial resolution focused

Auto-regressive

Lighter on memory

multi-gpu

ML models (predictors and targets)

LST_{bt}

Forcing (**h=-95 to 24**):
- *Wind stress equiv (U_w^2, V_w^2)*
- *St. Lawrence river runoff*
- *SSH (BDY)*

Predictors (**h=-95 to 0**):
- SSH,
- Ubt, Vbt
- Ubc, Vbc
- Ubc_slow, Vbc_slo
- Rho_a (density and
- RhoU_bt, RhoV_bt (barotropic density transport)

Target (individually) (**h=-95 to 24**):
- SSH
- Ubt, Vbt
- U, V

Anemoi

Forcing (**h=-95 to 24**):
- *Wind stress equiv (U_w^2, V_w^2)*
- *St. Lawrence river runoff*
- *SSH (not yet implemented)*

or & target) (**h=-95 to 0,1,…**):

slow

- RhoU_bt, RhoV_bt

Diagnostic (target) (**h=0 to 24**):
- U, V (out of error growth path)

Variables chosen based on the momentum equation

$$\frac{D\mathbf{u}}{Dt} + \frac{1}{\rho_*}\nabla_h p' + g\nabla_h \eta = \mathbf{R}_u + \mathbf{F}_u$$

ML models time-stepping, 96h pre-conditioning, 24h forecast

LST_{bt}

Forward model:

$Y_{t=-95...+24} = F(X_{F,t=-95...+24}, X_{P,t=-95...0})$

Loss $= f(Y_{t=-95...+24}, \mathbf{MT})$

(custom in house)

Forward model weights coverage: -95h to +24h

No
(no

to-regressive
chanics
Rollout = 6

P: **P**redictors in (LST_{bt}), **P**rognostic (Anemoi)
D: diagnostic
F: forcing
t: time
MT: model truth

Anemoi

Forward model (**multistep=96**):

$\mathbf{Y}_{P,D,t=+1} = F(X_{F,t=-95...0}, X_{P,t=-95...0})$

$\mathbf{Y}_{P,D,t=+2} = F(X_{F,t=-93...+2}, \mathbf{Y}_{P,t=+2})$

Forward model weights coverage: -95h to +1h

...for growth by design (for prognostic variables)

...

$Y_{P,D,t=+6} = F(X_{F,t=-89...+5}, Y_{P,t=+5})$

Loss $= f(Y_{P,D,t=+1}, Y_{P,D,t=+2}, \ldots Y_{P,D,t=+6}, \mathbf{MT})$

(huber loss)

Training experiments

- Experimental setup
  - 1994-2016 (training, 23 years)
  - 2017-2021 (validation, 5 years)
  - 2022-**2023** (inference, 2 years)

- 4 experiments: *(**apple to apple** comparison is **not really the goal** for now ➔ exploration phase)*

  1- **LST_{bt}** (1 GPU, 1 variable at a time)

  2- Anemoi: GraphTransformer (1 processor layer) (**GTR**) (16 GPU)

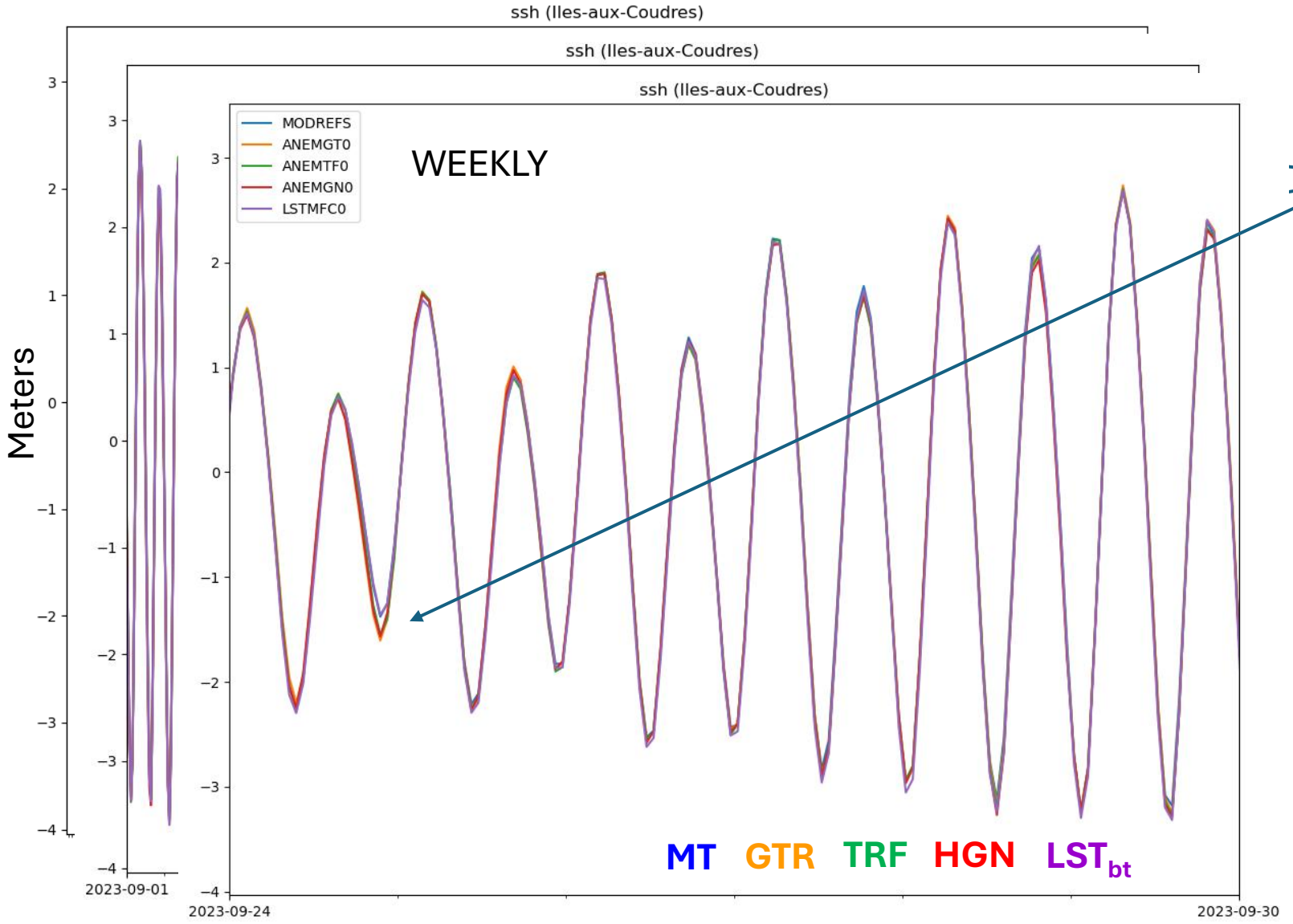
  3- Anemoi: Transformer w/ flash attention (1 processor layer) (**TRF**) (16 GPU)

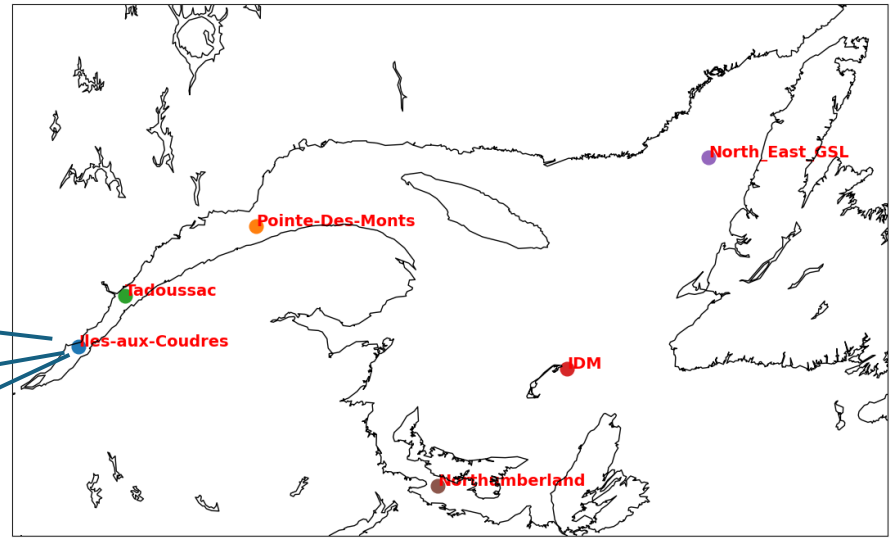
  4- Anemoi: GNN (hierarchical: 3 processor layers) (**HGN**) (32 GPU)
     - *No attention mechanism, but large scale spread through multi-mesh*

- Compared to CIOPS-E run « model truth » (**MT**)

ML results (end-to-end 0-23h forecasts, SSH)

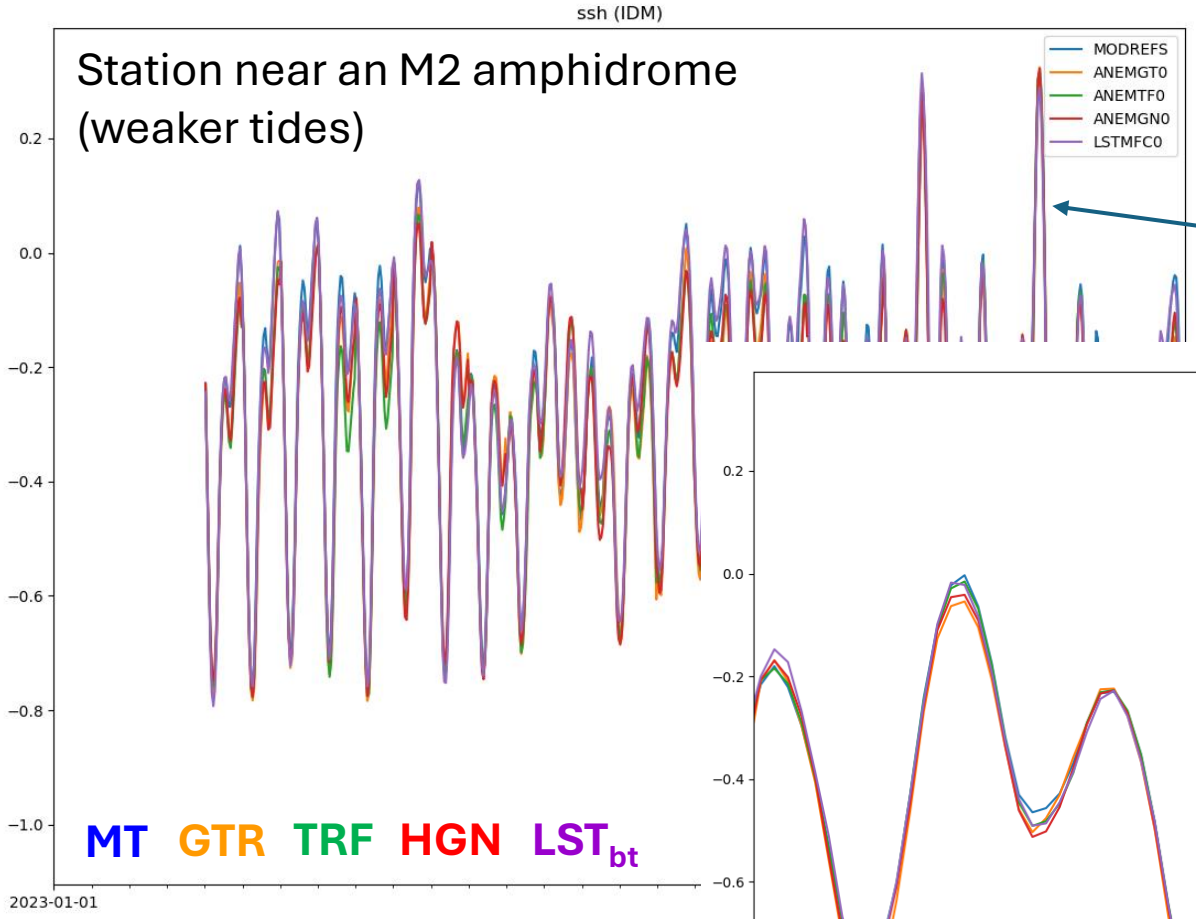
So can ML models emulate tides?

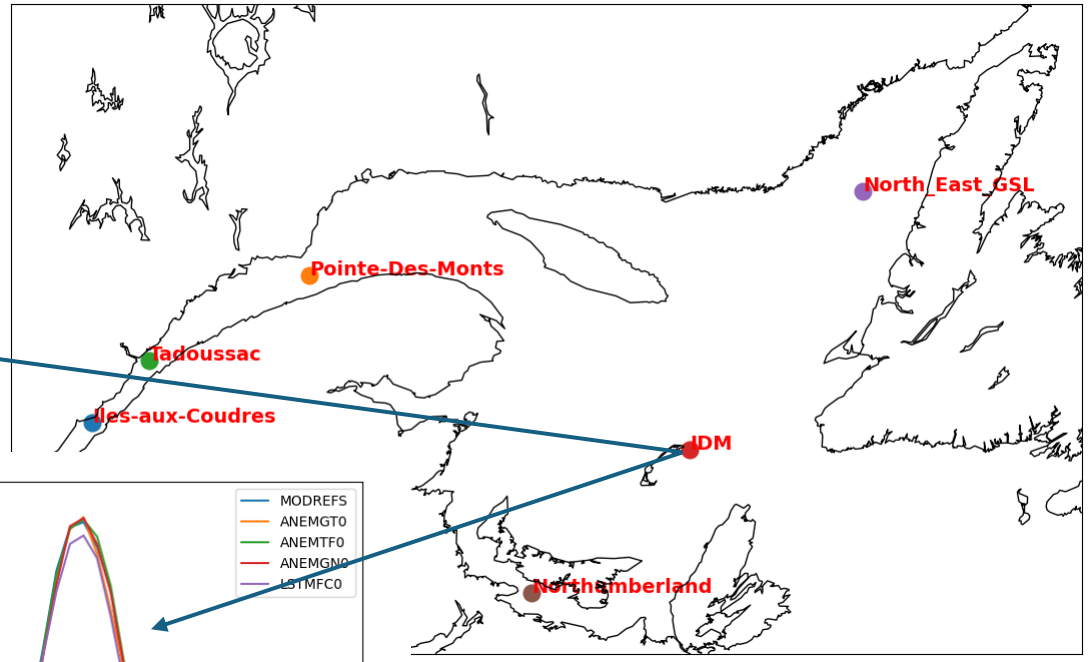


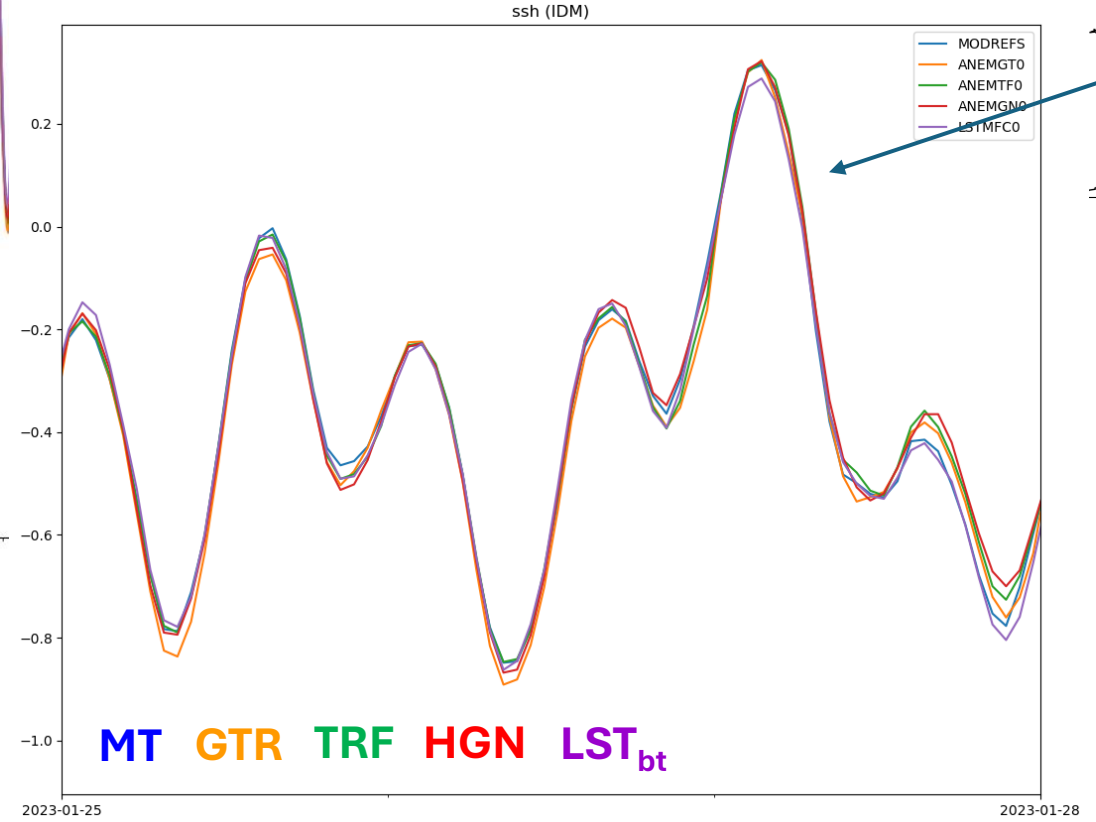


ML results (end-to-end 0-23h forecasts, SSH)

Surge examples







Anemoi, SSH error stats compiled for 2023 (against **MT**) (error growth)

LEAD TIME = **12h**

BIAS

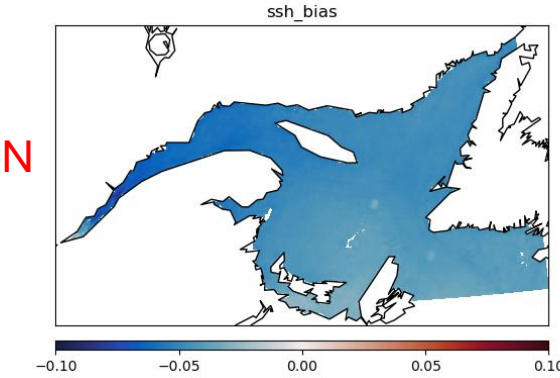
STD error

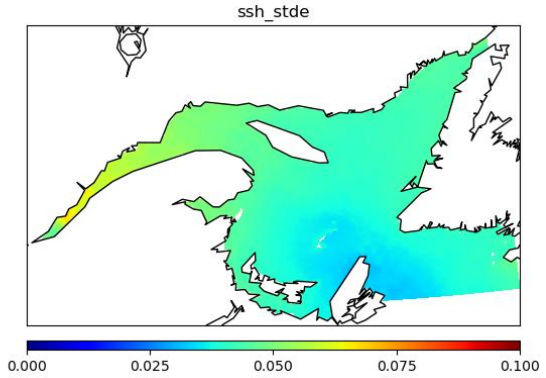
LEAD TIME = **23h**

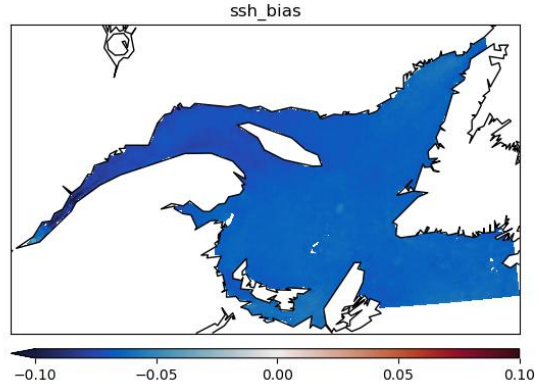
BIAS

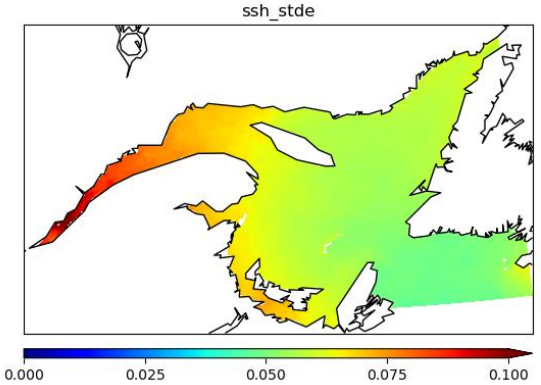
STD error

HGN

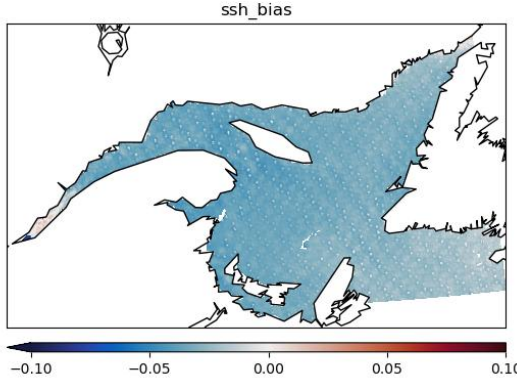


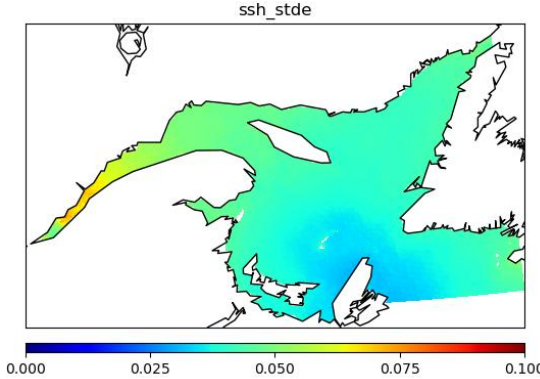


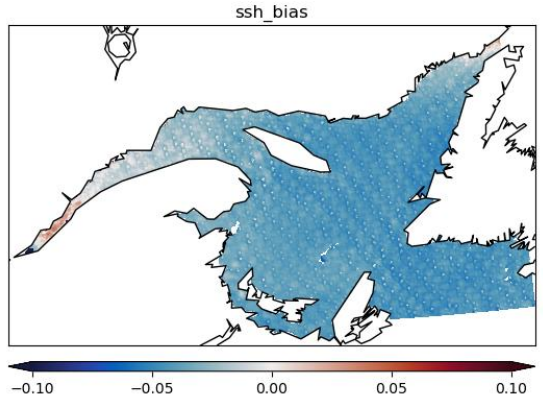


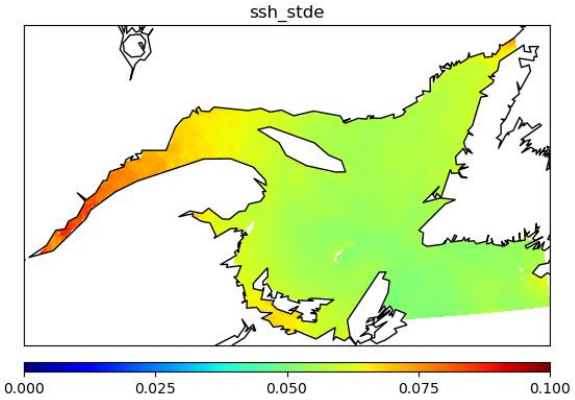


GTR

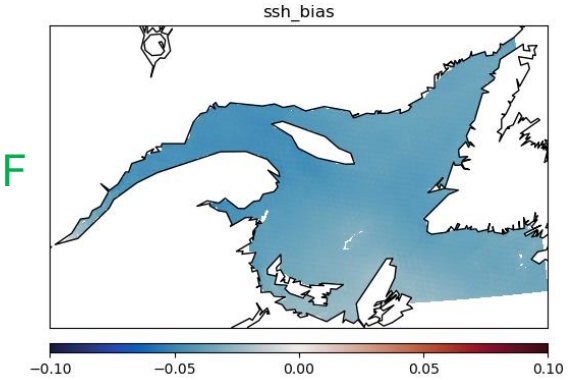


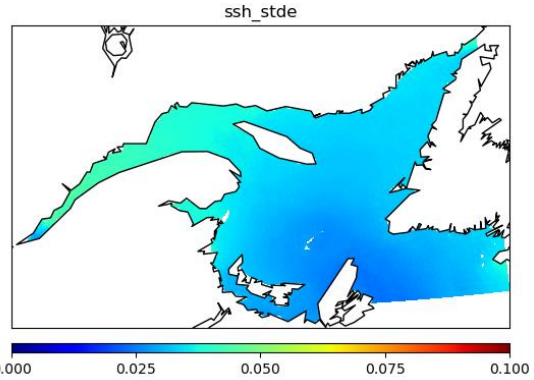


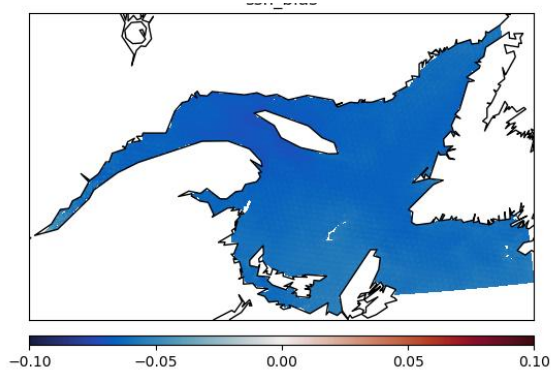


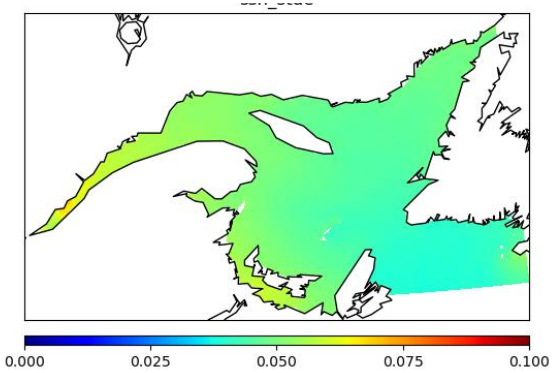


TRF









Anemoi (U,V) norm & direction, error stats (2023)

LEAD TIME = **12h**

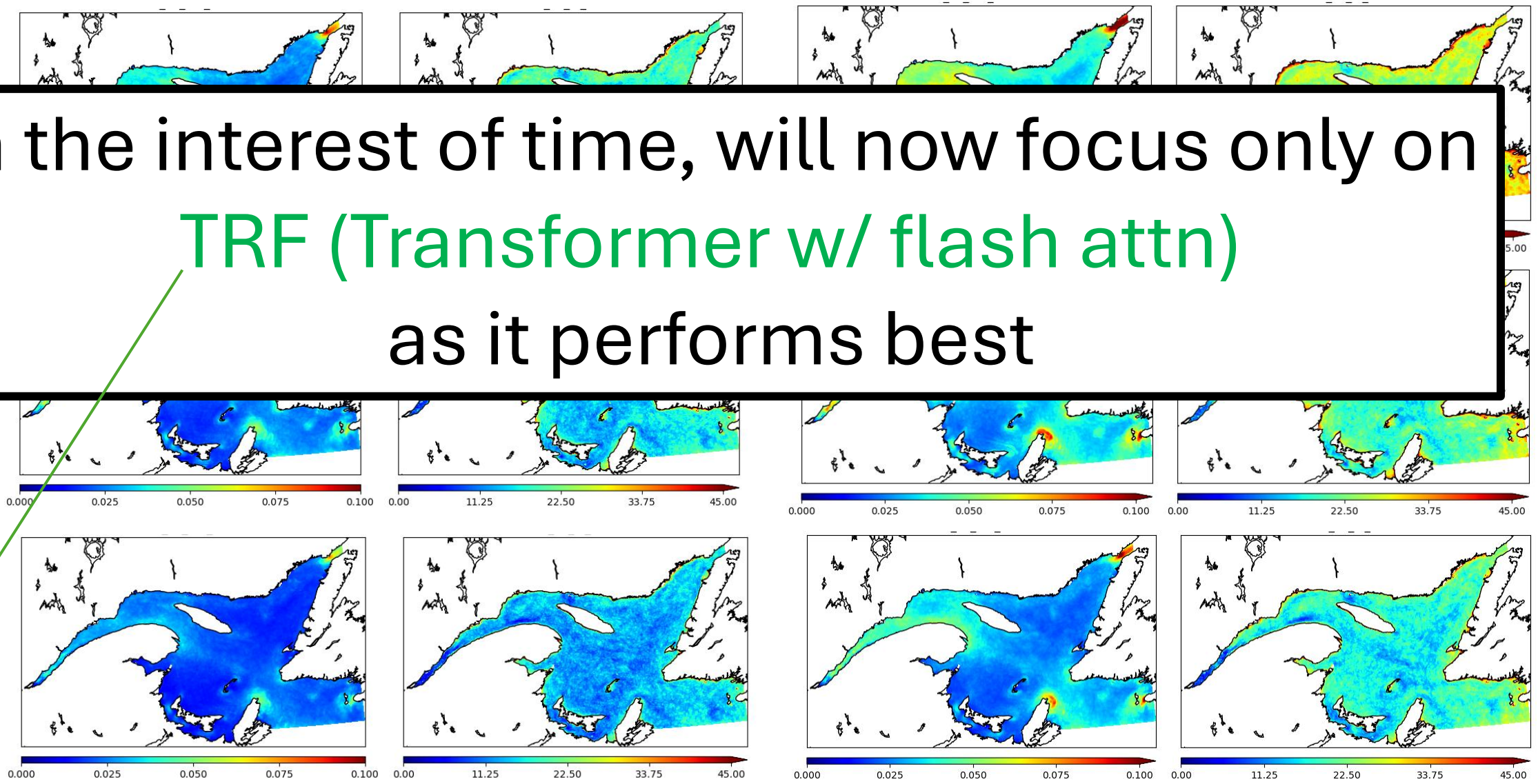
(U,V) NORM STD E

(U,V) direction STD E (deg.)

LEAD TIME = **23h**

(U,V) NORM STD E

(U,V) direction STD E (deg.)



In the interest of time, will now focus only on

TRF (Transformer w/ flash attn)

as it performs best

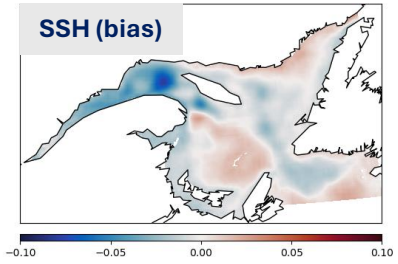
TRF

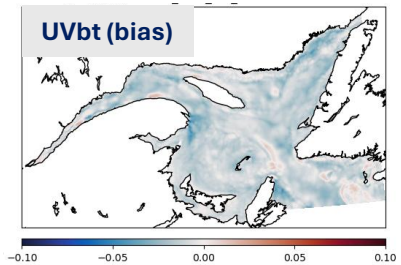
Error stats lead time 12h vs 23h, 2023 (flash attention game!)

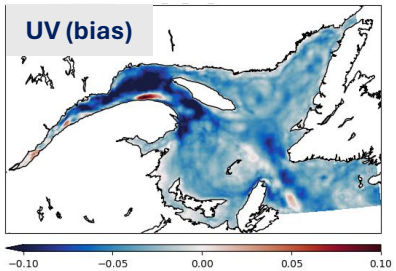
LST_{bt}

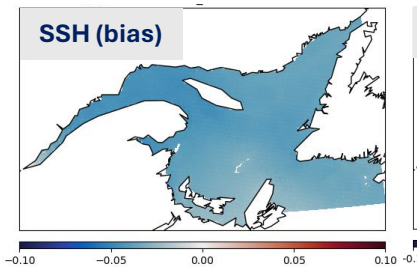
Anemoi TRF

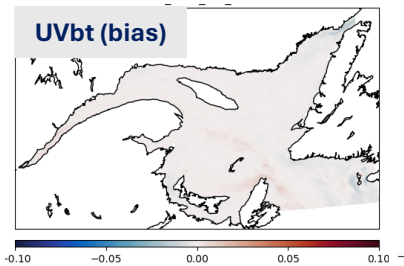
Lead time **12**

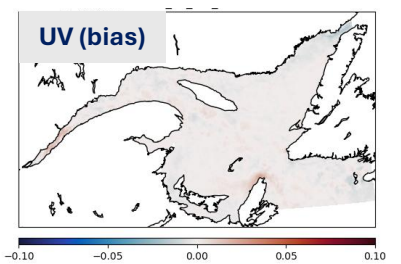


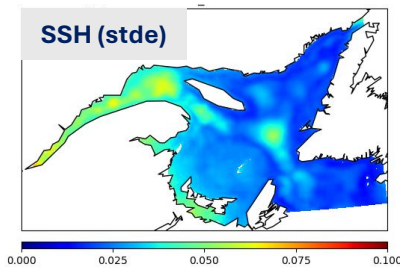


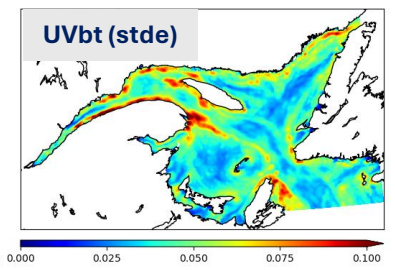


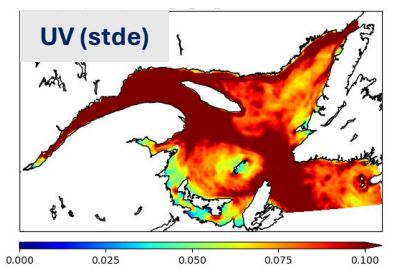


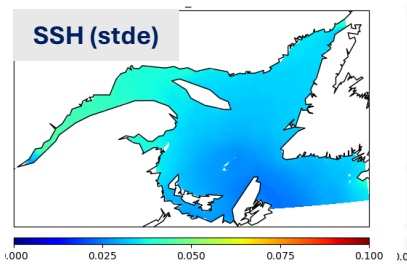


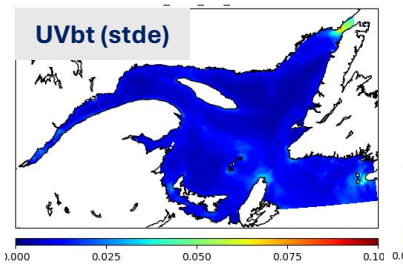


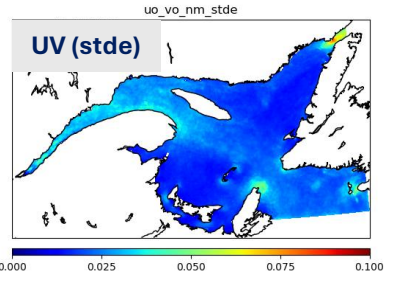




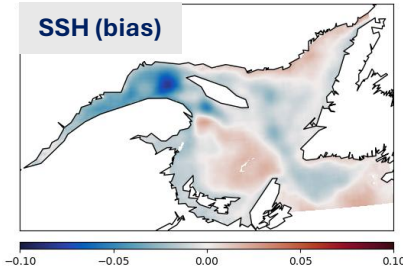


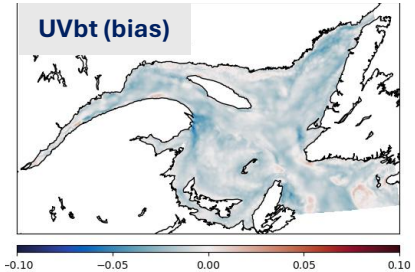


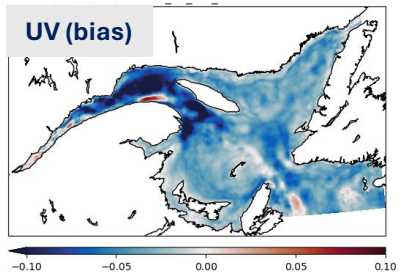


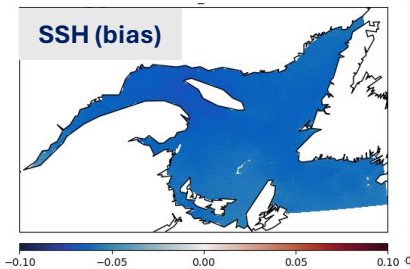


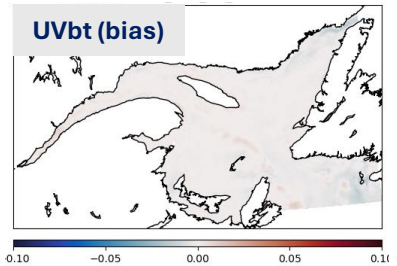
Lead time **23**

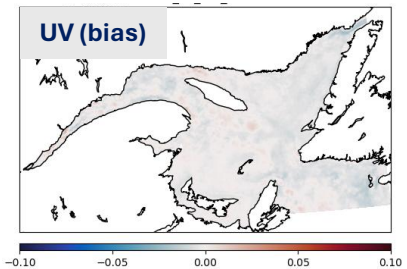


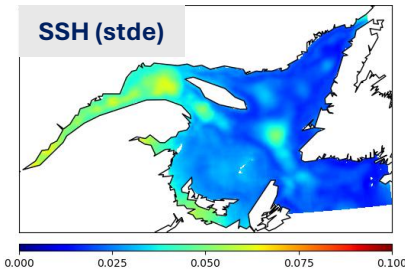


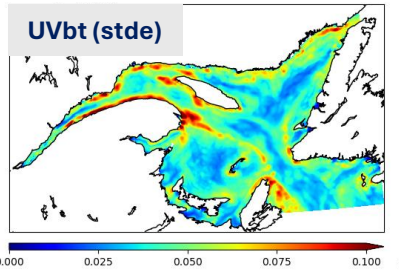


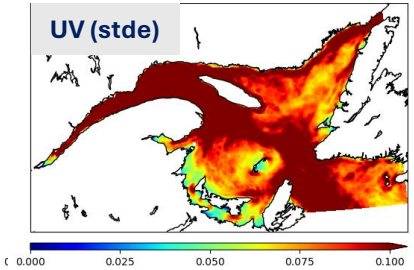


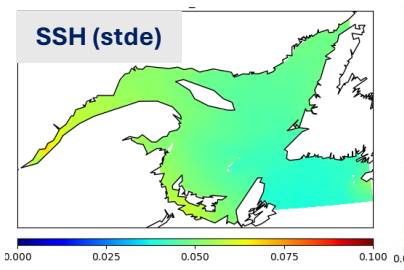


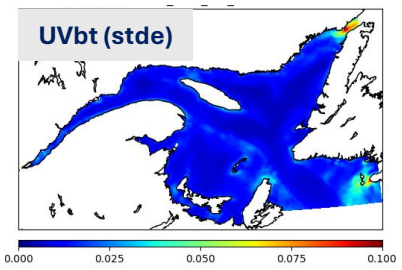


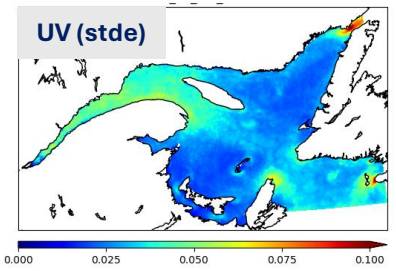






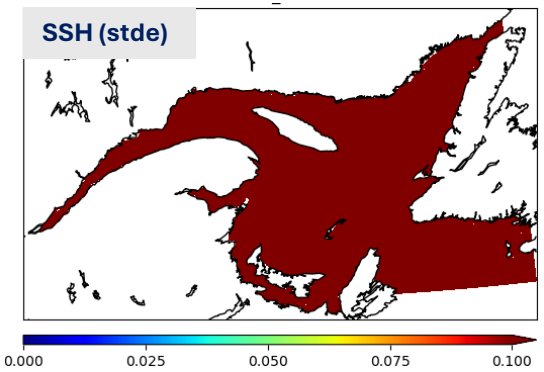


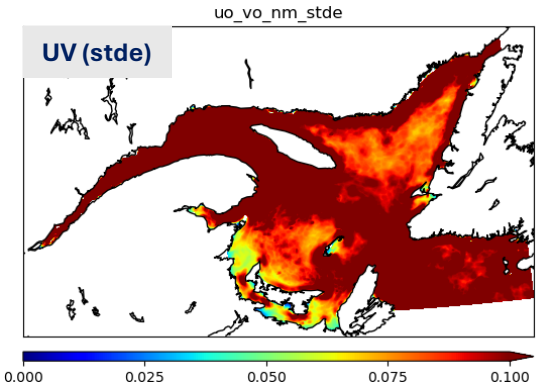


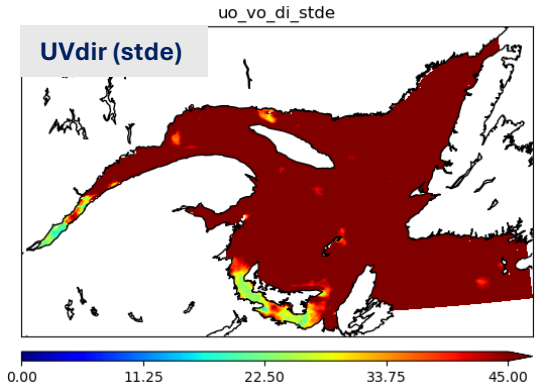


10 day *pseudo-forecast experiment with Anemoi

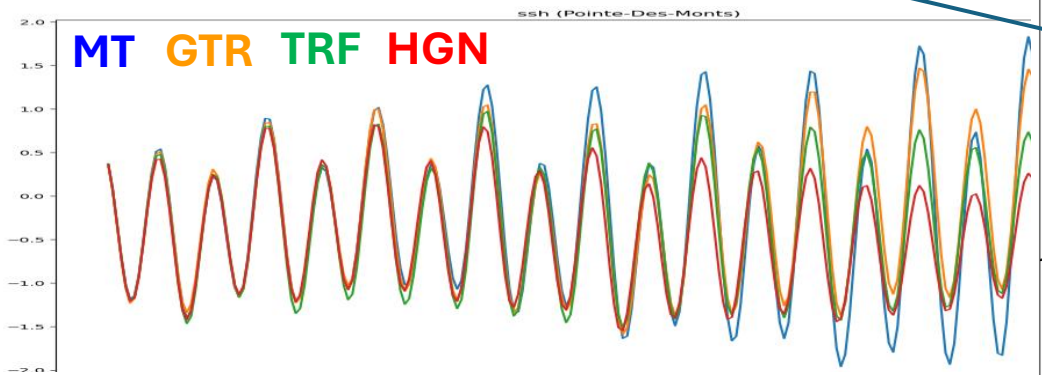
STD E (lead time = 3 days)

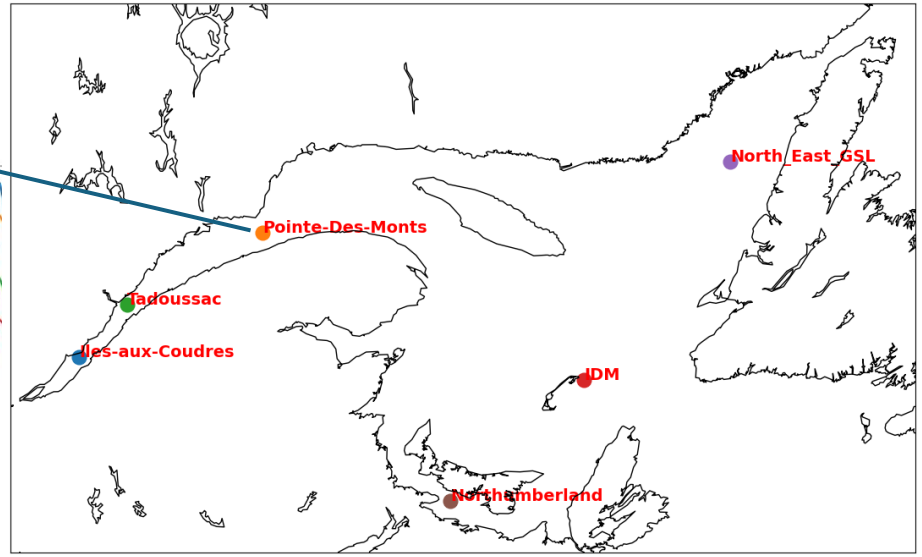


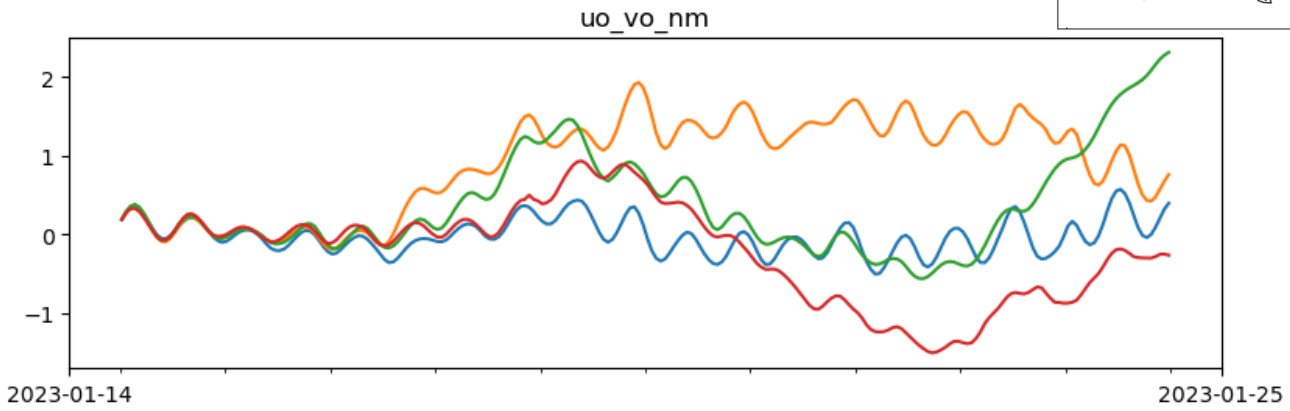


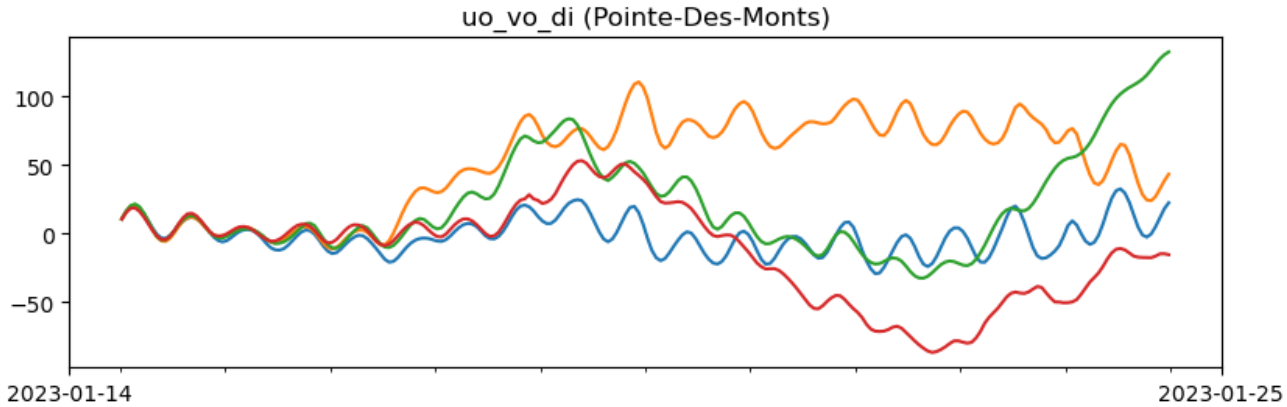


Pointe-Des-Monts









*atm forcing as in **MT** (short forecast, daily)

Conclusions

Both Anemoi and LSTM capture SSH variability with tides

- better spatial resolution in Anemoi but larger error growth as a function of lead time compared to LST_{bt}

For current emulation, Anemoi outperforms LST_{bt}

- larger error in chaotic flow regions (e.g. Gaspé current)
- STD error (total current) is dominated by the baroclinic component (not shown)

Error growth in Anemoi relatively uniform spatially, might be reduced with a boundary forcing constraint (w/ constrained mean sea level)

10 day forecast with Anemoi

- serious degradation after 3 days lead time

Next steps

1- ML results improvement

coastal ocean boundary forcing in Anemoi

hybrid approach? SSH (w/ LST_{bt}), currents (w/ Anemoi)

2- Forecast emulation applications

lead time > 48, ensembles, full CIOPS-E domain

3D currents

pre-training (30 years), U, V = F(UVbt, UVbc, UVbc_slow, …), with **UVbc=UVbc_slow**

fine-tuning, a few years of 3D hourly outputs

3- More detailed evaluation

harmonic analysis, storm surge, observations, ML error versus ocean model uncertainty

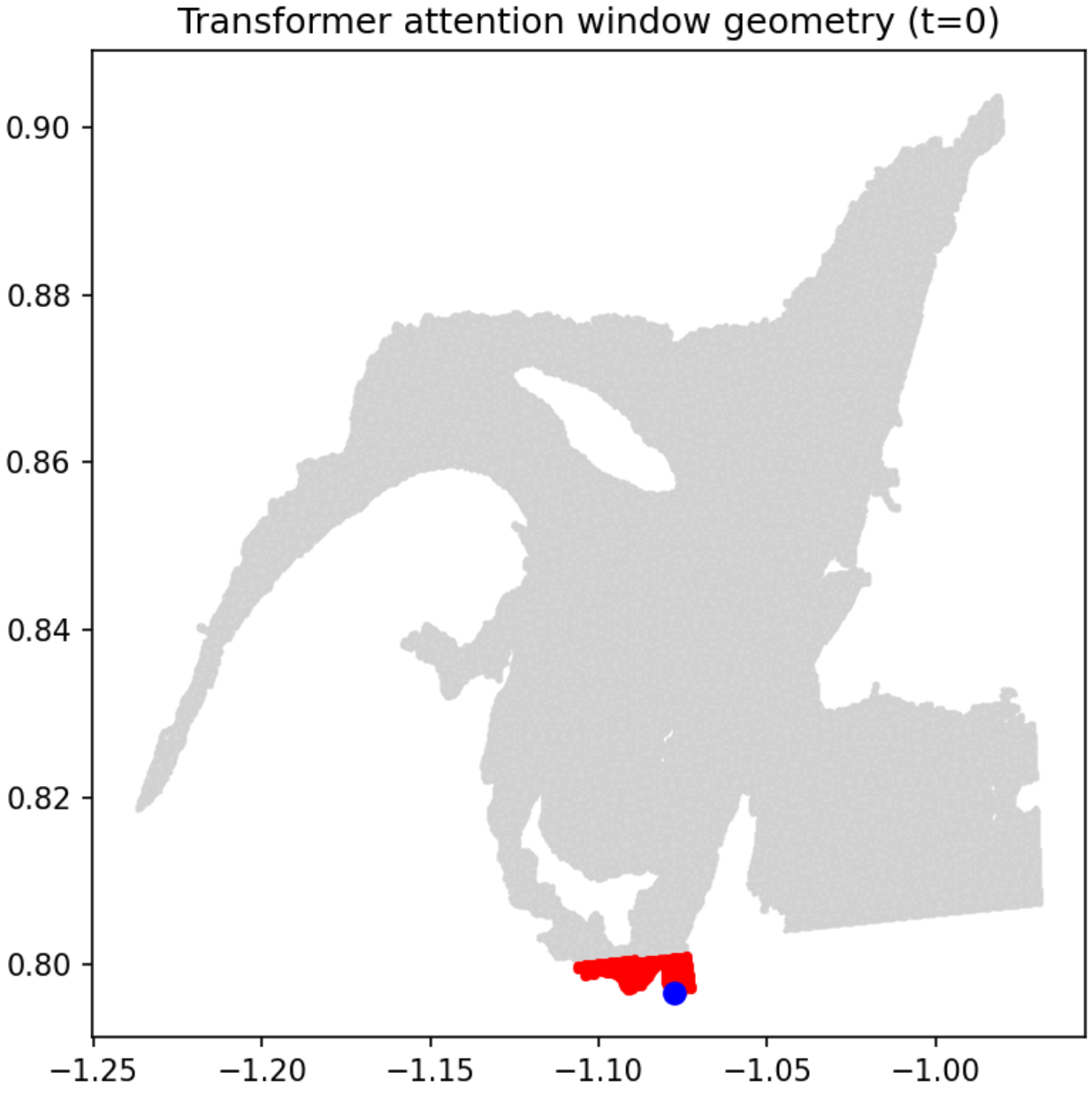
3- Downscaling

Twin 30 years simulations ongoing

GSL144 (1/144, ~500m)

STLE200 (200m)

Thank you!



Extras...

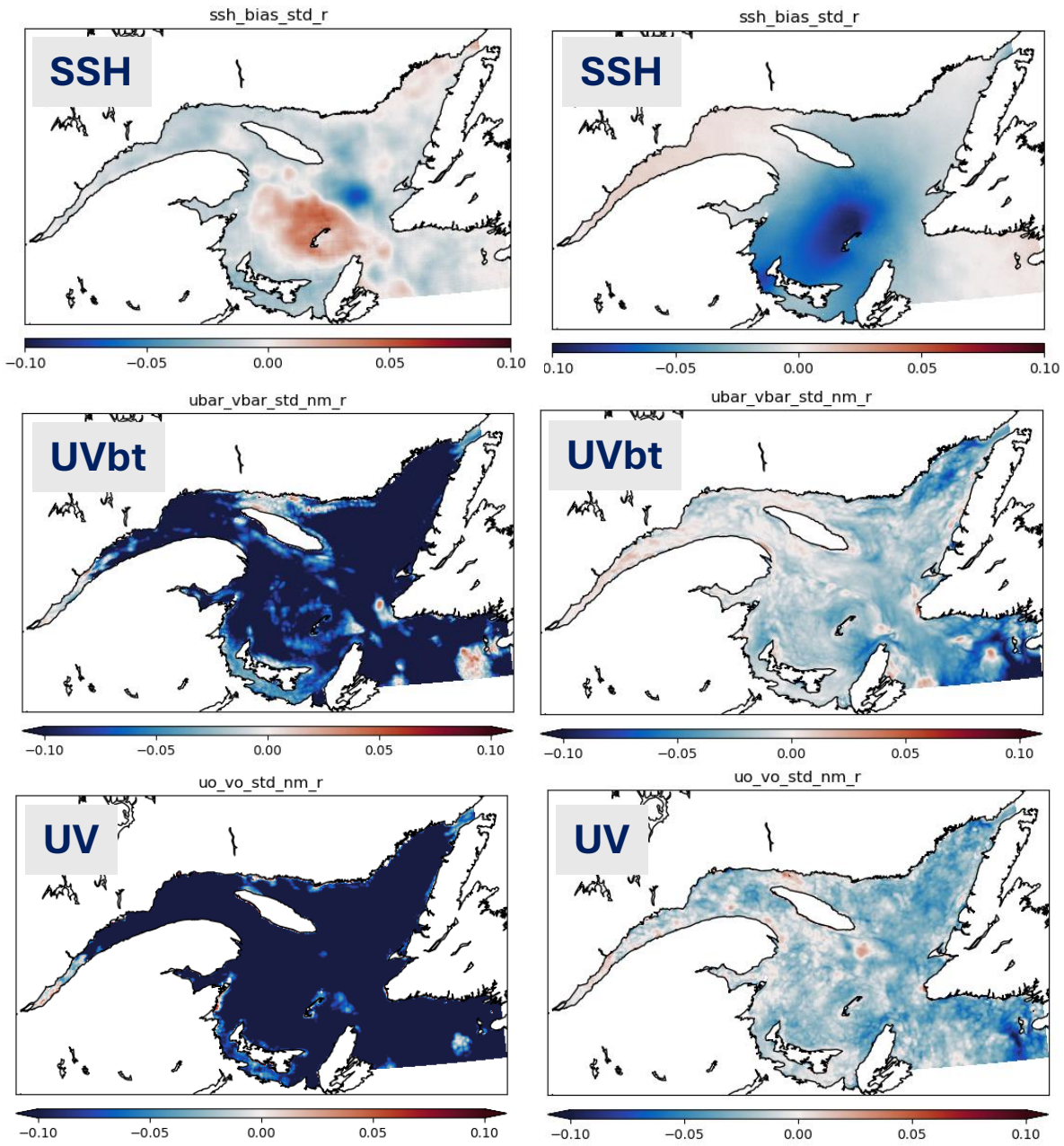
Variability metrics

(lead time = 23h)

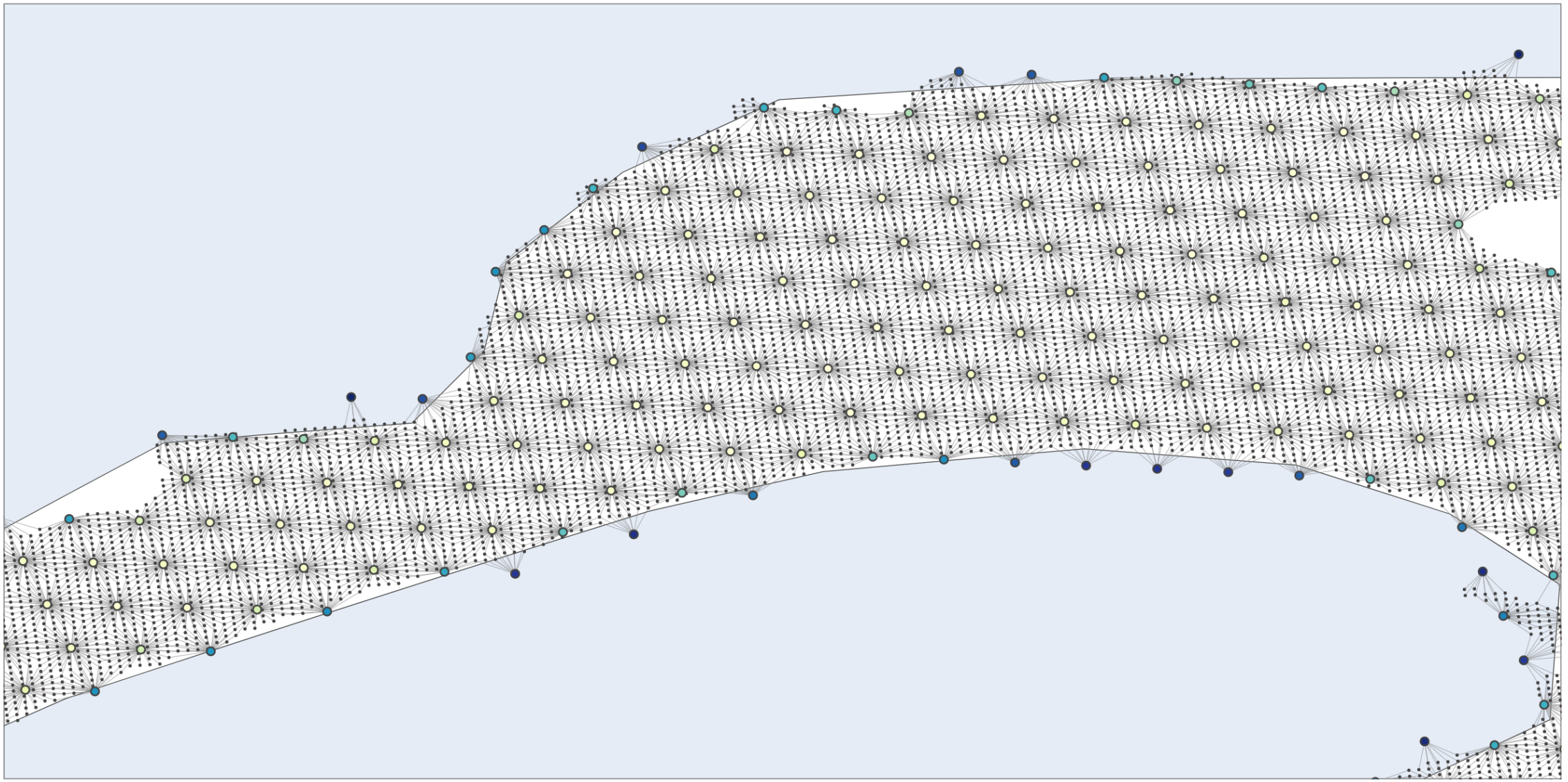
σ(LST$_{bt}$ or TRF)/ σ(MT) - 1

Red: too much variability

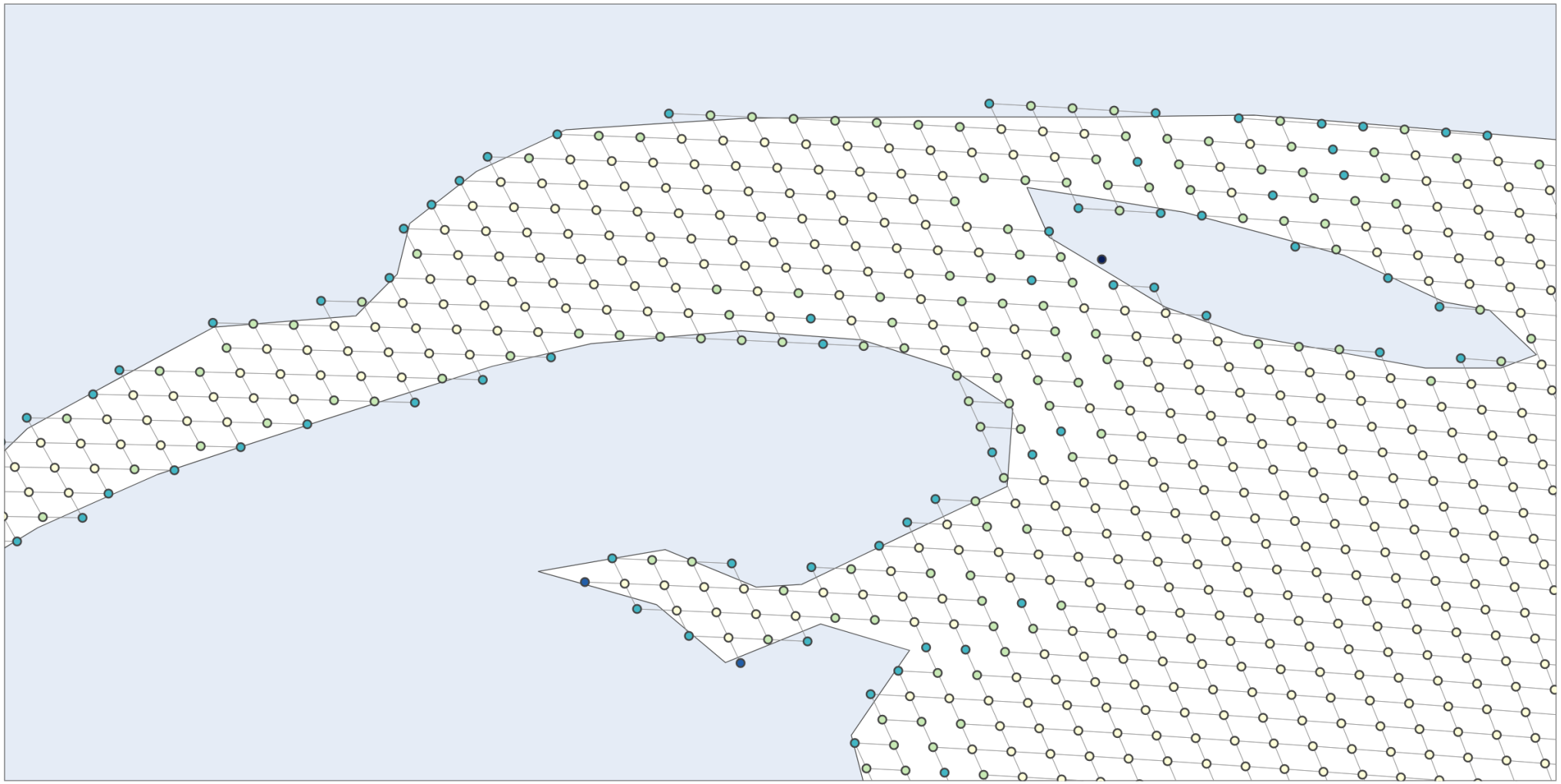
Blue: missing variability



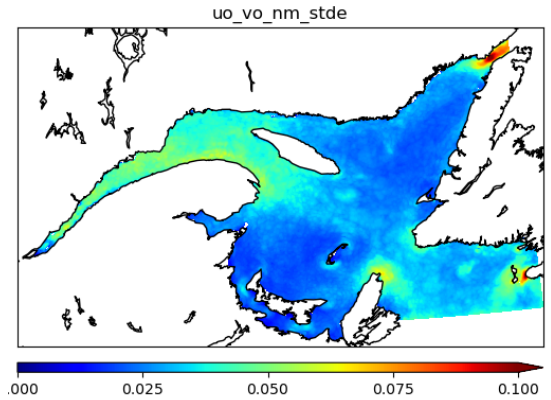
Portion of graph encoder grid (data to hidden)

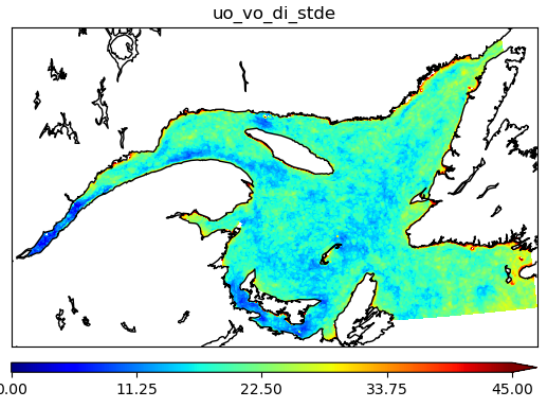


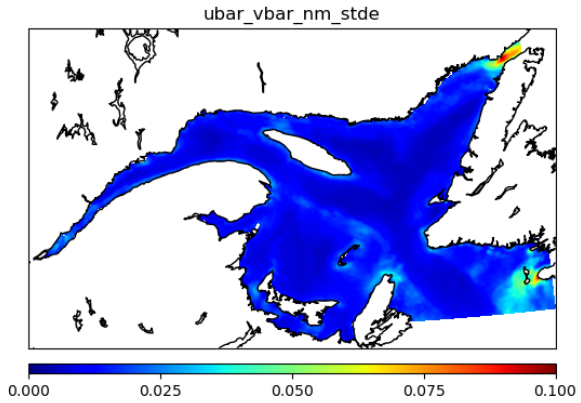
Portion of graph encoder grid (hidden to hidden)

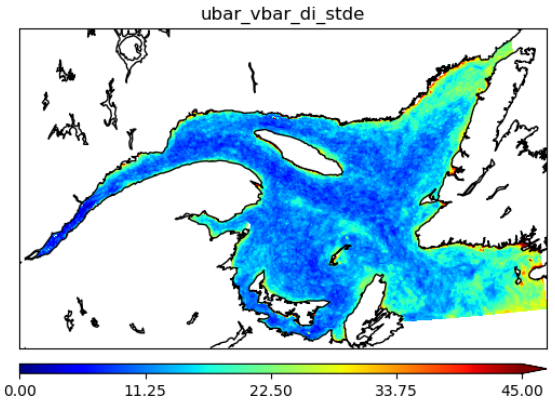


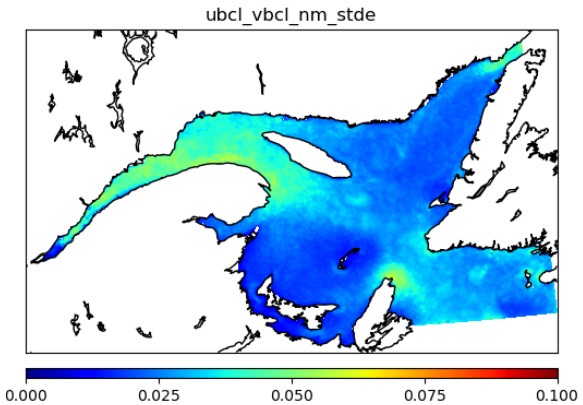
Barotropic/Baroclinic relative error (TRF)

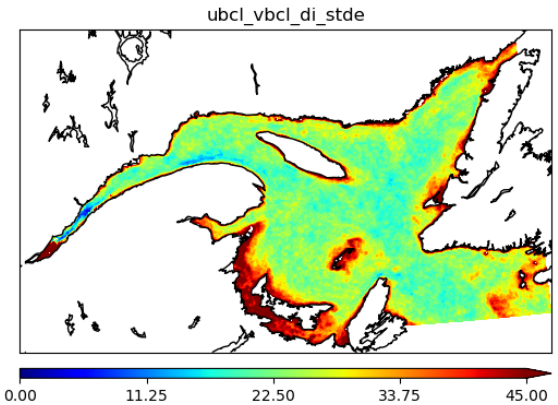


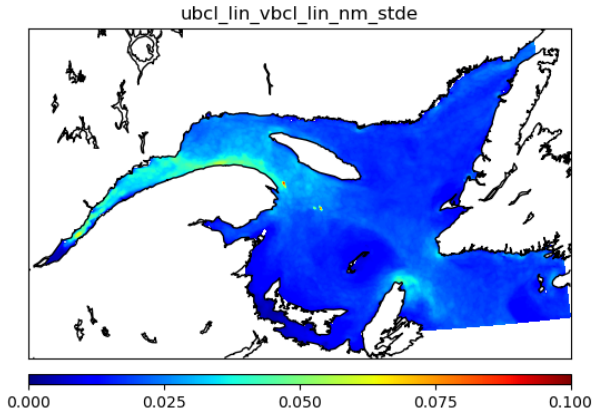


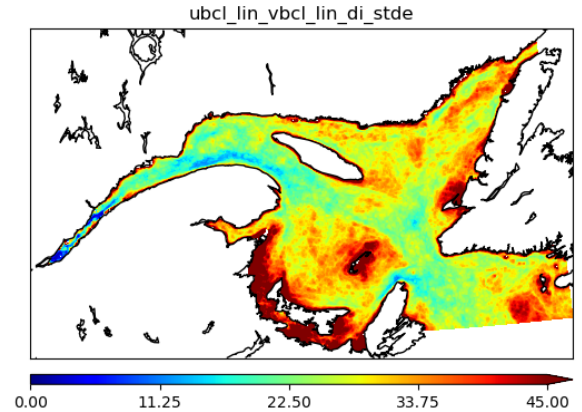




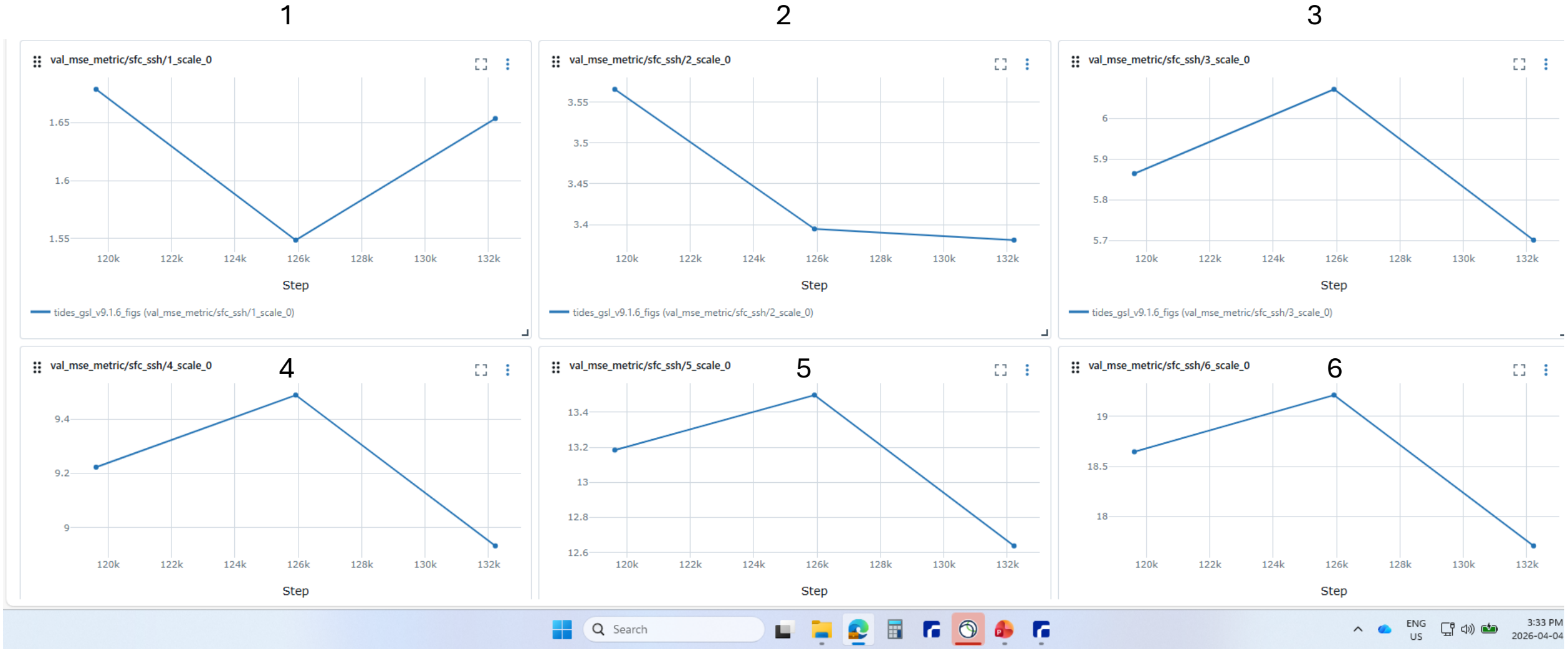








Training error growth example (rollout=6, SSH)



x 16 GPU x 2

YAML's model override section

GTR :

```
model:
 num_channels: 512
 processor:
  _target_:
anemoi.models.layers.processor.GraphTransformerProcessor
   num_layers: 8
   num_heads: 8 # GraphTransformer or Transformer only
```

TRF:

```
model:
 num_channels: 1024
 processor:
  _target_: anemoi.models.layers.processor.TransformerProcessor
  num_layers: 16
  num_heads: 16 # GraphTransformer or Transformer only
  window_size: 512
```

HGN:

```
model:
 model:
  _target_:
anemoi.models.models.AnemoiModelEncProcDecHierarchical
 num_channels: 512
 enable_hierarchical_level_processing: True
 level_process_num_layers: 8
 processor:
  num_layers: 16
  num_chunks: 2
            (processor is GNN here..., in defaults)
```

ALL configs:

```
training:
   gradient_clip:
      val: 1.
      algorithm: norm
```