# Synergy of oceanographic and geodetic data using machine learning

Artu Ellmann, TalTech geodesy group

**BSHC CHART DATUM, WATER LEVEL AND CURRENTS WORKING GROUP (CDWCWG)**

Tallinn, March 25-26, 2025

---

## Past, Present and Future Research

By synergy various sea level data sets

- Improve on hydrodynamic model
- Improve on satellite altimetry ← **Phase 1 (accurate data)**
- Improvements in geoid

By utilizing machine learning with Dynamic Topography
- Sea level forecasting (1-5 days ahead)
- Forecasting of extremes
- Water Balance in Baltic Sea ← **Phase 2 (understanding and applicability)**
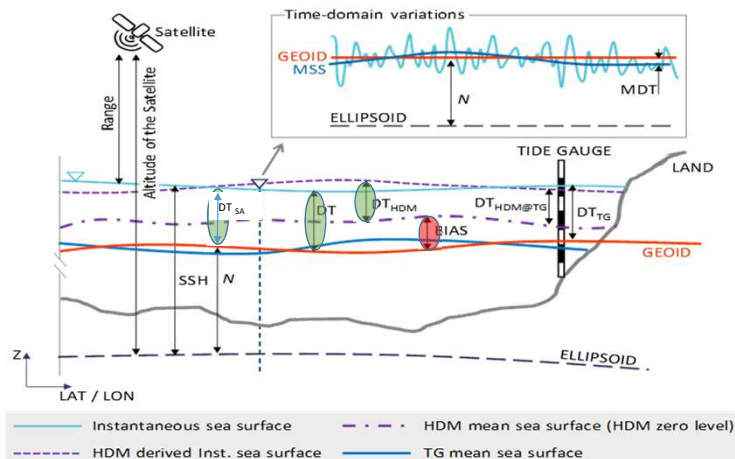
Future Research
- Determination of ocean dynamic patterns and processes
- Determination of Realistic Under Keel Clearance ← **Phase 3 (understanding and applicability**

**Research funded by Estonian Research Council (2019-2027)**
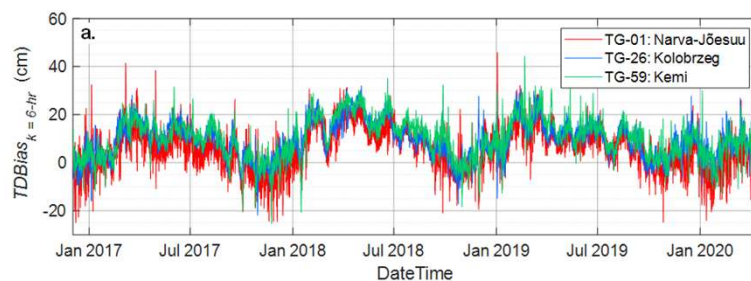
## Sources of sea level data: vertical reference
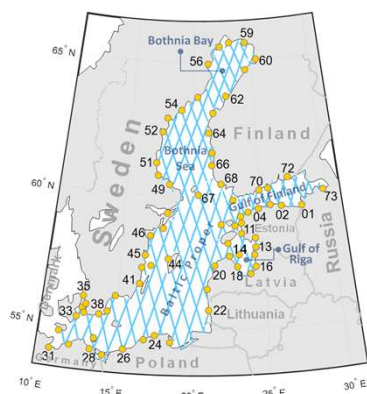


# Improve on hydrodynamic models

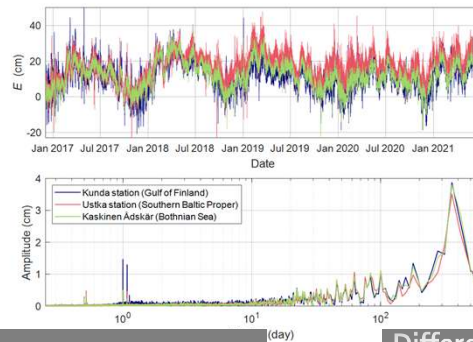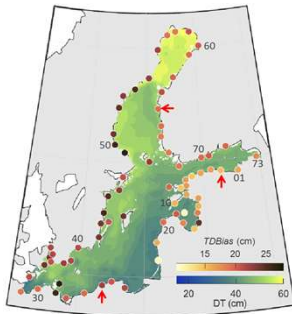Limitation: Hydrodynamic models may have a spatial and temporal bias that exist

Advantage: Tide gauges can be referred to the geoid (i.e Baltic Sea Chart Datum) by implementing corrections



The temporal domain Bias time-series of Nemo-Nordic model, for example, at three TG stations with k = 6-hour backward moving average method

# Vertical Reference Differences: HDM vs TG

$$E(\varphi_{TG}, \lambda_{TG}, t) = DT_{HDM}(\varphi_{TG}, \lambda_{TG}, t) - DT_{TG}(\varphi_{TG}, \lambda_{TG}, t)$$



**Question/Challenge:**
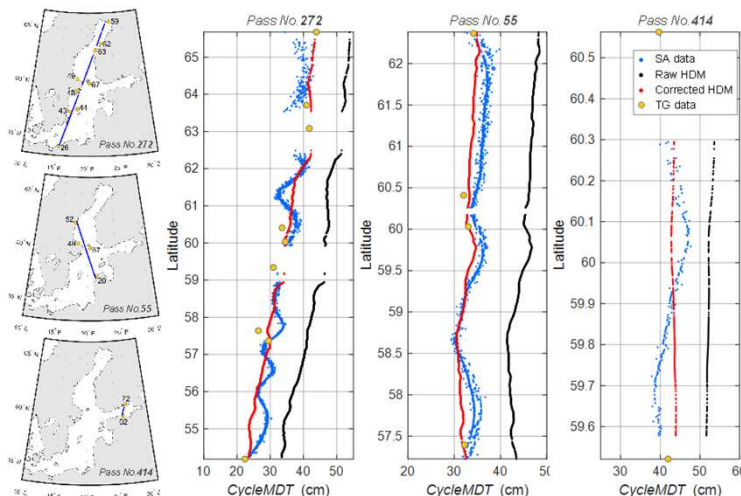- Coastal areas can be corrected by TG but what is the procedure in the offshore areas

**Observations:**
- Difference can be as much as -20 to 40cm
- Stations follows similar pattern and frequency of error

**Differences:**
- Spatial and temporal resolution differs
- Vertical datum differs
- Different mode of measurement

# Validation with Satellite Altimetry



Comparison of raw/corrected Nemo-Nordic model with the SA data (Sentinel-3A) and TG records via CycleMDT (i.e., mean of 33 cycles) in three example

TAL
TECH

## Machine Learning Background

**Deep Learning/Machine Learning:** computer learns to perform tasks based on experience it gains during training.
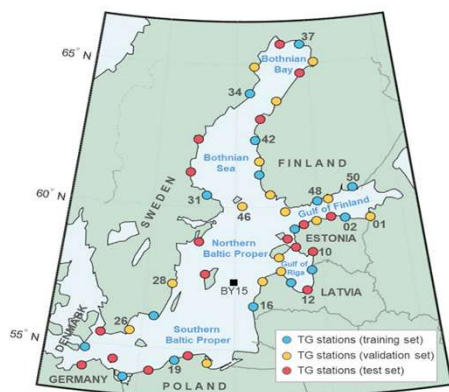
Basic components:

- **Data (as input)**:Good quality (so you receive good results)
- **A model (i.e a hypothesis):** to predict quantities of interest (model chosen by user)
- **Loss function**: the discrepancy (difference between prediction and observed)

An iterative approach is used until the loss function is minimum

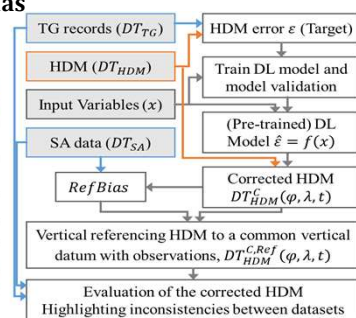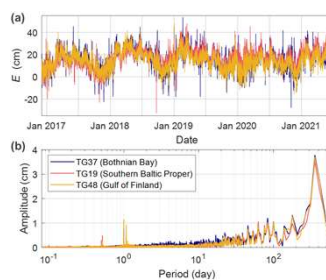- Validation: internally and/or externally (independent source)

**TAL TECH**

---

## Results: Method 2, Deep Learning (WaveNet Approach)

$$E(\varphi, \lambda, t) = \varepsilon(\varphi, \lambda, t) + \textbf{RefBias}$$



- 4.5 years examined
- Train: 16 TG stations (blue)
- Test: 18 TG stations (red)
- Validation : 16 stations (yellow)
- Evaluated: 52 stations

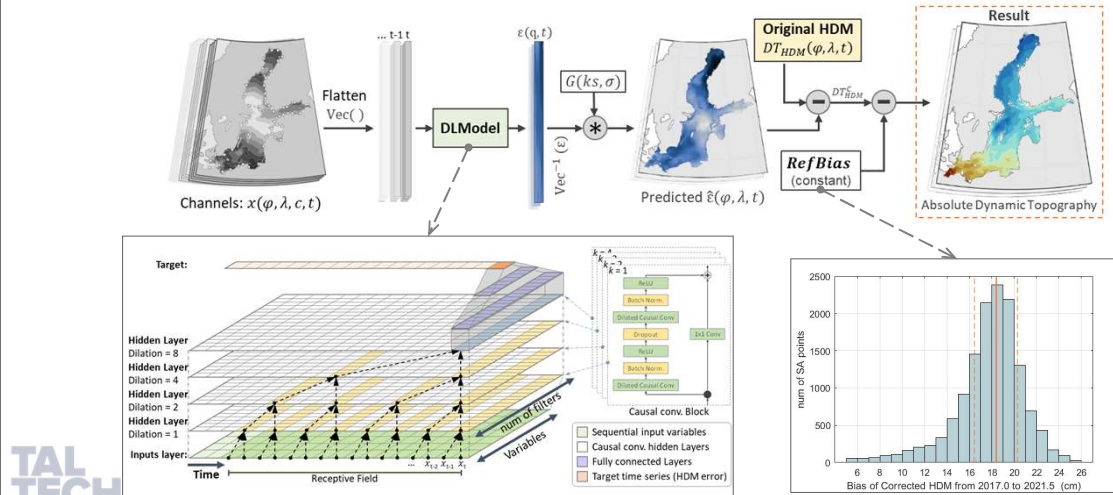- The HDM error $\varepsilon$ expected to consist of different components that are most likely to be predictable both in time and space.
- RefBias is expected to be constant both in space and time
- DL model with temporal dilated causal convolution layers inspired by **WaveNet (Oord et al., 2016)**…(spectrum analysis)
- Causal convolution is unidirectional (1D), and the learnable parameters (i.e., weights and biases) are trained to predict

# Method

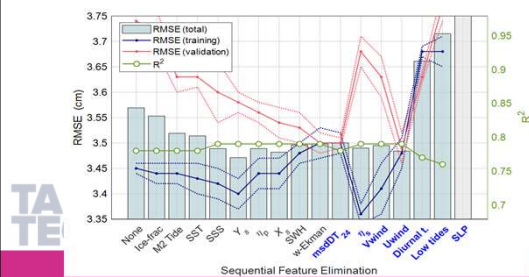$$E(\varphi, \lambda, t) = \varepsilon(\varphi, \lambda, t) + RefBias$$



Page | 9

---

# Method II: Determine relevant variables
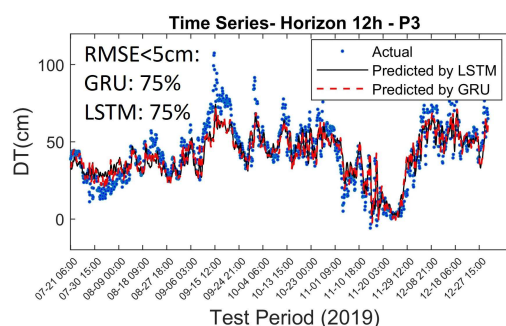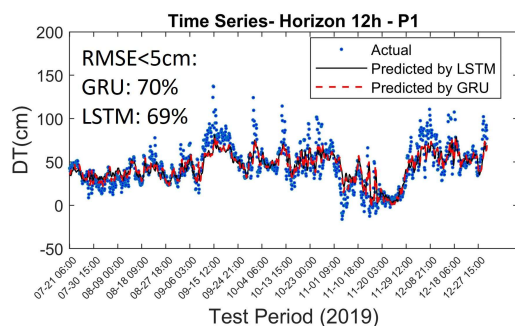
$$E(\varphi, \lambda, t) = \varepsilon(\varphi, \lambda, t) + RefBias$$

- A wrapper-type sequential feature elimination algorithm was utilized
- The algorithm states training with a subset of variables and then removes a variable based on an elimination criterion. This criterion is a combination of the RMSEs from both the training and validation sets,
- **DL model was generalized** over the spatial dimension using input variables: '$msdDT_{24}$', '$\eta_s$', 'Uwind', 'Vwind', 'Diurnal tides', 'Low tides', and 'SLP'.



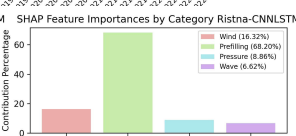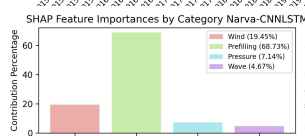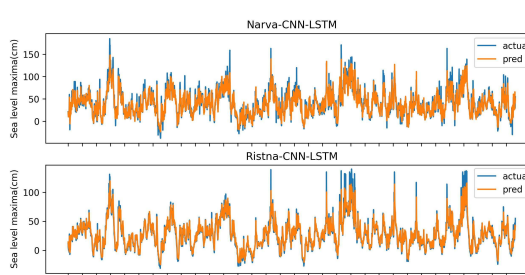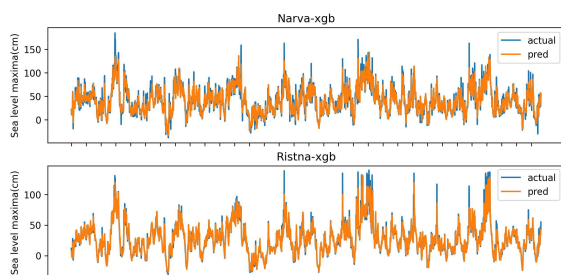|  | Variable | units | Sourced resolution | | Data source |
|---|---|---|---|---|---|
|  |  |  | Temporal | Spatial |  |
| 1 | Zonal wind (Uwind) | m/s | Hourly | 1 NM | Sourced from Nemo-Nordic dataset |
| 2 | Meridional wind (Vwind) | m/s | Hourly | 1 NM |  |
| 3 | Sea surface temperature (SST) | °C | Hourly | 1 NM |  |
| 4 | Sea surface salinity (SSS) | psu | Hourly | 1 NM |  |
| 5 | Ice fraction (Ice-frac) | % | Hourly | 1 NM |  |
| 6 | Zonal wind stress ($X_s$) | Pa |  |  | Computed at the HDM grid points with an hourly temporal resolution using U and Vwind |
| 7 | Meridional wind stress ($Y_s$) | Pa |  |  |  |
| 8 | Ekman pumping (w-Ekman) | m/s |  |  |  |
| 9 | Sea surface pressure (SLP) | Pa | 3-hourly | 5.5 km | Copernicus: https://doi.org/10.24381/cds.622a565a |
| 10 | Precipitation | cm | Hourly | 0.25 | MTPR was sourced |

5

## Results: Instantaneous sea level forecasting 12 hours



> Extremes are problematic in both methods for forecasting
> P1 located in eastern Gulf showed lower performance perhaps due to not including river discharge and sea ice

Rajabi-Kiasari, Saeed; Delpeche-Ellmann, Nicole; Ellmann, Artu (2024). 'Dynamic Topography Forecasting using Deep Recurrent Neural Networks with high resolution hydrodynamic model'. Published

## Forecasting sea level extremes using DL (1971-2022)



- 70% of data (1971-01-08-2007-05-28) was used for training,
- 15% (2007-05-29-2015-03-15) for validation,
- 15% for test (2015-03-16-2022-12-31).

Narva
XGB - R-squared (Test): 0.83, RMSE: 10,36
RF - R-squared (Test): 0.83, RMSE: 10,42
**CNN-LSTM - R-squared (Test): 0.84**, RMSE: 10,32

Ristna
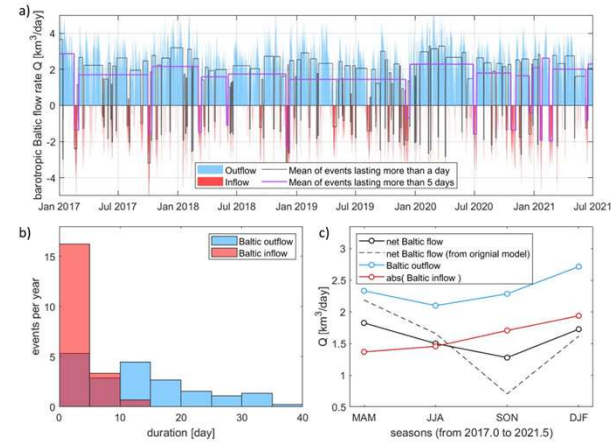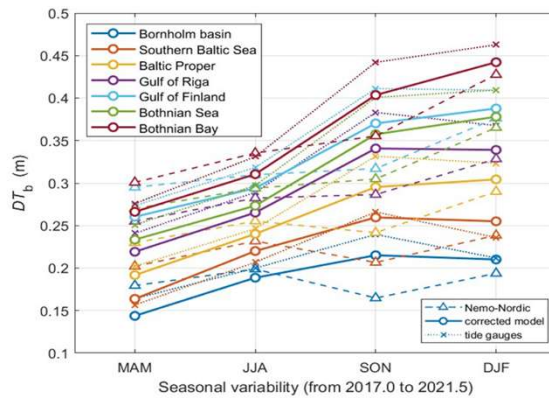XGB - R-squared (Test): 0.78, RMSE: 13,6
RF - R-squared (Test): 0.77, RMSE: 14
**CNN-LSTM - R-squared (Test): 0.81**, RMSE: 12,90

Rajabi-Kiasari, Saeed; Delpeche-Ellmann, Nicole; Ellmann, Artu (2024). Deep Learning to forecast extremes in Baltic Sea. In review

6

## Water balance in Baltic Sea using DT

Published (almost)



Jahanmard, V.; Delpeche-Ellmann, N.; Ellmann, A. (2024). published

---

## Summarizing forecasting

- Recurrent neural networks are deep learning models suitable for forecasting
- The GRU model slightly outperformed LSTM model. GRU model simpler
- Forecasting of extremes can be problematic. The DL method, time span and input variables choosen is vital for good forecasting

### General summary

- New demands are technology required for oceanograhy applications. It is important to get it right from the beginning
- For seal level data; common and accurate vertical reference is essential for linking with ohter data sets
- DL/ML methods can assist improving the models and understanding underlying processes
- Forecasting of extremes can be problematic. The DL method, time span and input variables choosen is vital for good forecasting

## References

Jahanmard, Vahidreza; Hordoir, Robinson; Delpeche-Ellmann, Nicole; Ellmann, Artu (2023). Quantification of Hydrodynamic Model Sea Level Bias Utilizing Deep Learning and Synergistic Integration of Data Sources. Ocean Modelling, 186, #102286. DOI: 10.1016/j.ocemod.2023.102286

Jahanmard, V.; Delpeche-Ellmann, N.; Ellmann, A. (2022). Towards realistic dynamic topography from coast to offshore by incorporating hydrodynamic and geoid models. Ocean Modelling, #102124. DOI: 10.1016/j.ocemod.2022.102124.

Jahanmard, Vahidreza; Delpeche-Ellmann, Nicole; Ellmann, Artu (2021). Realistic dynamic topography through coupling geoid and hydrodynamic models of the Baltic Sea. Continental Shelf Research, 222, 104421. DOI: 10.1016/j.csr.2021.104421

Rajabi-Kiasari, Saeed; Delpeche-Ellmann, Nicole; Ellmann, Artu (2023). Forecasting of absolute dynamic topography using deep learning algorithm with application to the Baltic Sea. Computers & Geosciences, 178, #105406. DOI: 10.1016/j.cageo.2023.105406.

Rajabi-Kiasari, Saeed; Delpeche-Ellmann, Nicole; Ellmann, Artu (2024).Submitted.

Rajabi-Kiasari, Saeed; Delpeche-Ellmann, Nicole; Ellmann, Artu (2024). Deep Learning to forecast extremes in Baltic Sea. In progress

Mostafavi, Majid; Delpeche-Ellmann, Nicole; Ellmann, Artu; Jahanmard, Vahidreza (2023). Determination of Accurate Dynamic Topography for the Baltic Sea Using Satellite Altimetry and a Marine Geoid Model. Remote Sensing, 15 (8), #2189. DOI: 10.3390/rs15082189

**TAL TECH**

---

## Validation options using SWOT satellite altimetry over the open seas

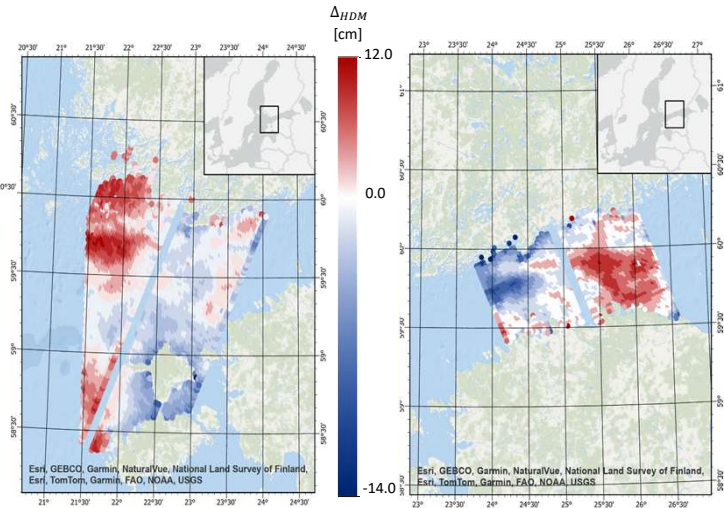### OCEANOGRAPHIC APPROACH

STUDY AREA

- 6 Estonian and 4 Finnish tide gauges and **Baltic Sea Physics and Forecast HDM**.
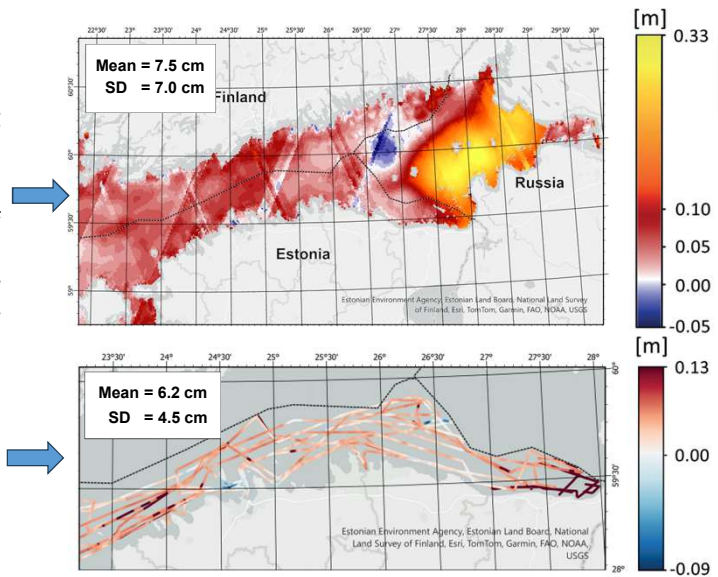
## OCEANOGRAPHIC APPROACH

RESULTS 1/3

- **RMSE average is 15.38 cm** with maximum of 53.52 cm and minimum of 3.98 cm.

- **SD average is around 8 cm**, maximum 15.07 cm and minimum 2.19 cm



## OCEANOGRAPHIC APPROACH

RESULTS 2/2

- Area of **possible marine geoid modelling error was investigated** encompassing sea level data (Varbla et al., 2021)
- Due to high coverage and resolution of KaRIn data **we managed to quantify BSCD2000 modelling errors in previously inaccessible area** (due to country boundaries).

- **Shipborne GNSS and ALS** data used for geoid modelling was also used to assess **KaRIn-derived geoidal heights**.

**Thank you for your attention**

TAL
TECH