



ILMATIETEEN LAITOS  
METEOROLOGISKA INSTITUTET  
FINNISH METEOROLOGICAL INSTITUTE

# Nopeita ja ketteriä meriennusteita surrogaattimalleilla

Antti Westerlund, Benoit Espinola  
ja Aleks Nummelin

Ilmatieteen laitos, Merentutkimus

2023-11-16



# Nopeita ja ketteriä meriennusteita surrogaattimalleilla

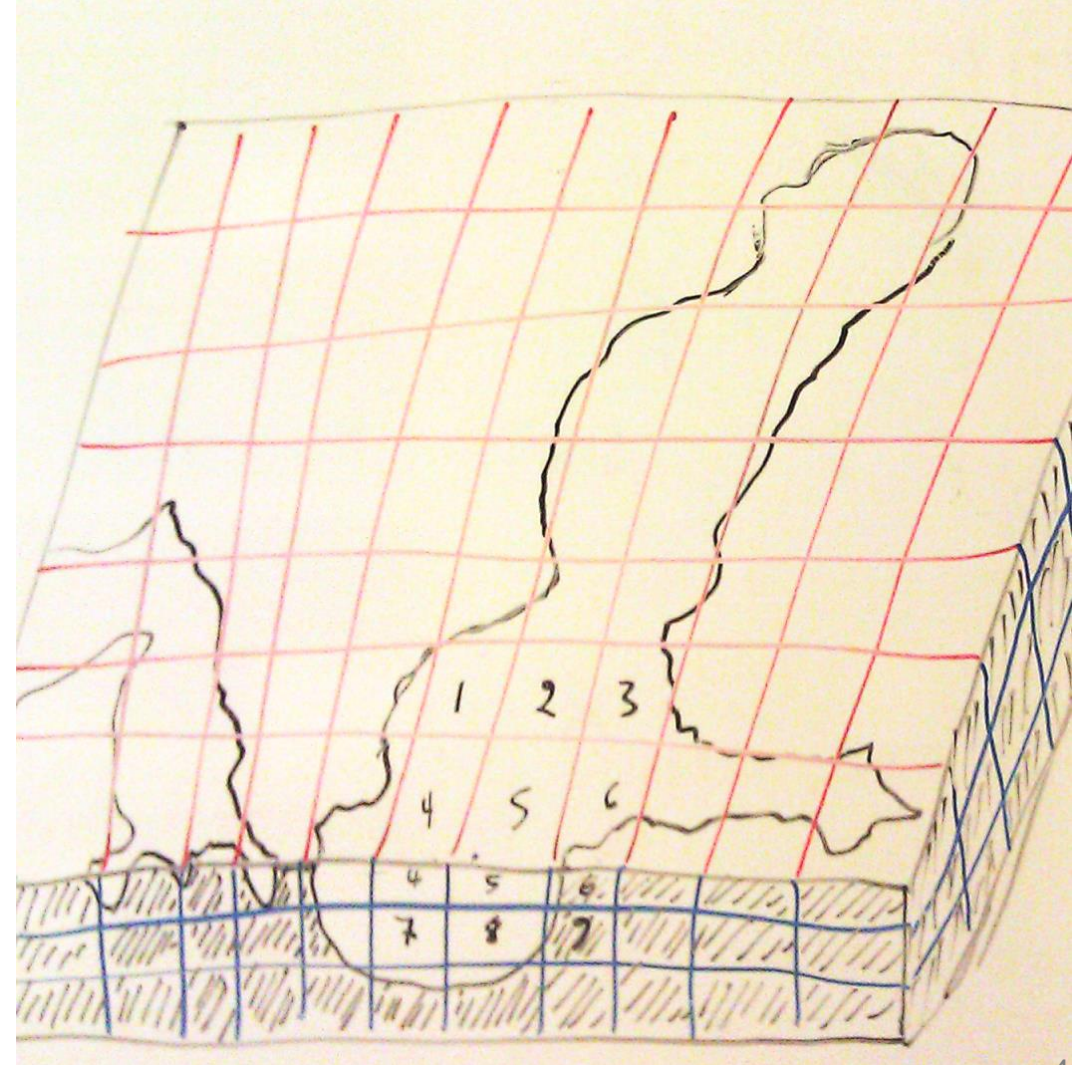
- Hankkeen nimi:
  - "Nopeita ja ketteriä meriennusteita surrogaattimalleilla"
- Toteuttajataho:
  - Ilmatieteen laitos, Merentutkimus
  - Vastuullinen johtaja: Antti Westerlund
- Hankkeelle myönnetty MATINE-rahoitus:
  - 119 618 e
- Hankkeen kesto ja tutkimusvuosi:
  - 1.3.2022-31.12.2023 (toinen vuosi)

# Merensuojelun olosuhdetiedon turvallisuussovellukset

- **Merellisellä fysikaalisella olosuhdetiedolla...**
  - Virtaukset, kulkeutumisen suunnat
  - Lämpötila, suolaisuus
  - Kerrostuneisuus, tiheysjakauma
- **...voidaan varautua riskeihin ja tukea turvallisuus toimijoita**
  - Merivoimien operatiivinen toiminta
  - Merionnettomuudet, öljyonnettomuudet
  - Merivesitulvat, ydinvoimaloiden häiriötilanteet

# Olosuhdetietoa merimalleilla

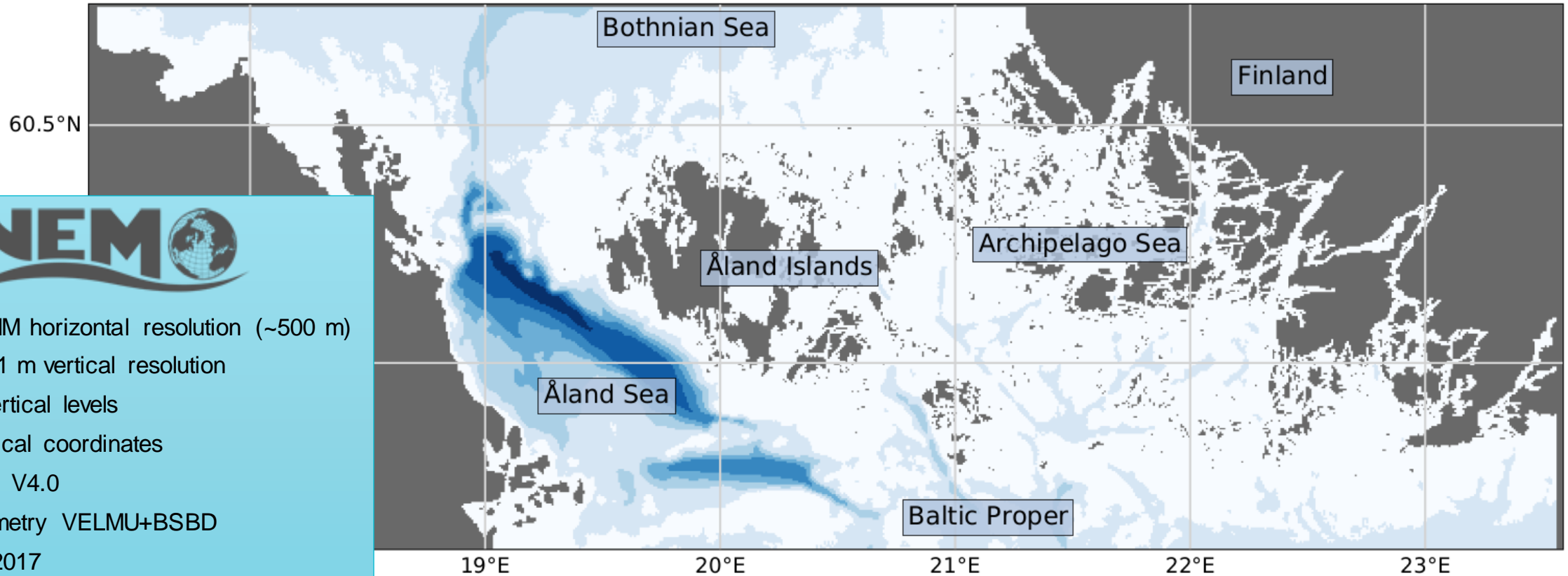
- Havaintojen lisäksi olosuhdetietoa saadaan **laskennallisilla merimalleilla**
  - Analyysyjä ja ennusteita meren tilasta
  - Vuosikymmeniä käytetty ja toimivaksi tiedetty lähestymistapa
- Merialueelle muodostetaan hila. Hilapisteille lasketaan meren tilaa lähtien fysiikan perusyhtälöistä.
  - Vrt. säämallit
- Esimerkiksi Ilmatieteen laitoksella merimallit ovat osa jokapäiväistä operatiivista toimintaa
  - IL:n meripalvelut



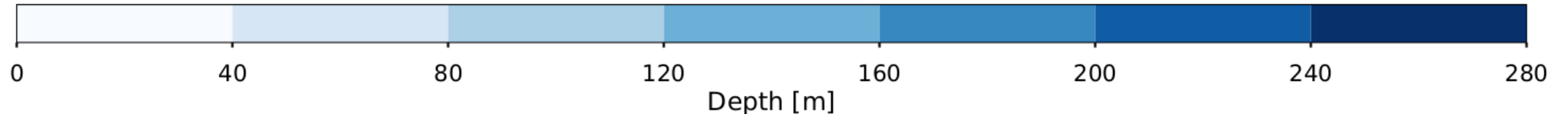
# Merimallintamisen haasteita – yleisesti ja Suomessa

- Vaaditaan suurta laskentakapasiteettia
  - Suurteholaskenta
- Erityisesti Suomen rikkonaisilla rannikoilla tarvetta suurelle erottelukyvylle
  - Resurssivaatimukset yhä vain kasvavat
- Laskenta-ajat voivat olla pitkiä
- Soveltaminen vaatii paljon erityisosaamista
  - Hankaloittaa käyttöä operatiivisessa toiminnassa

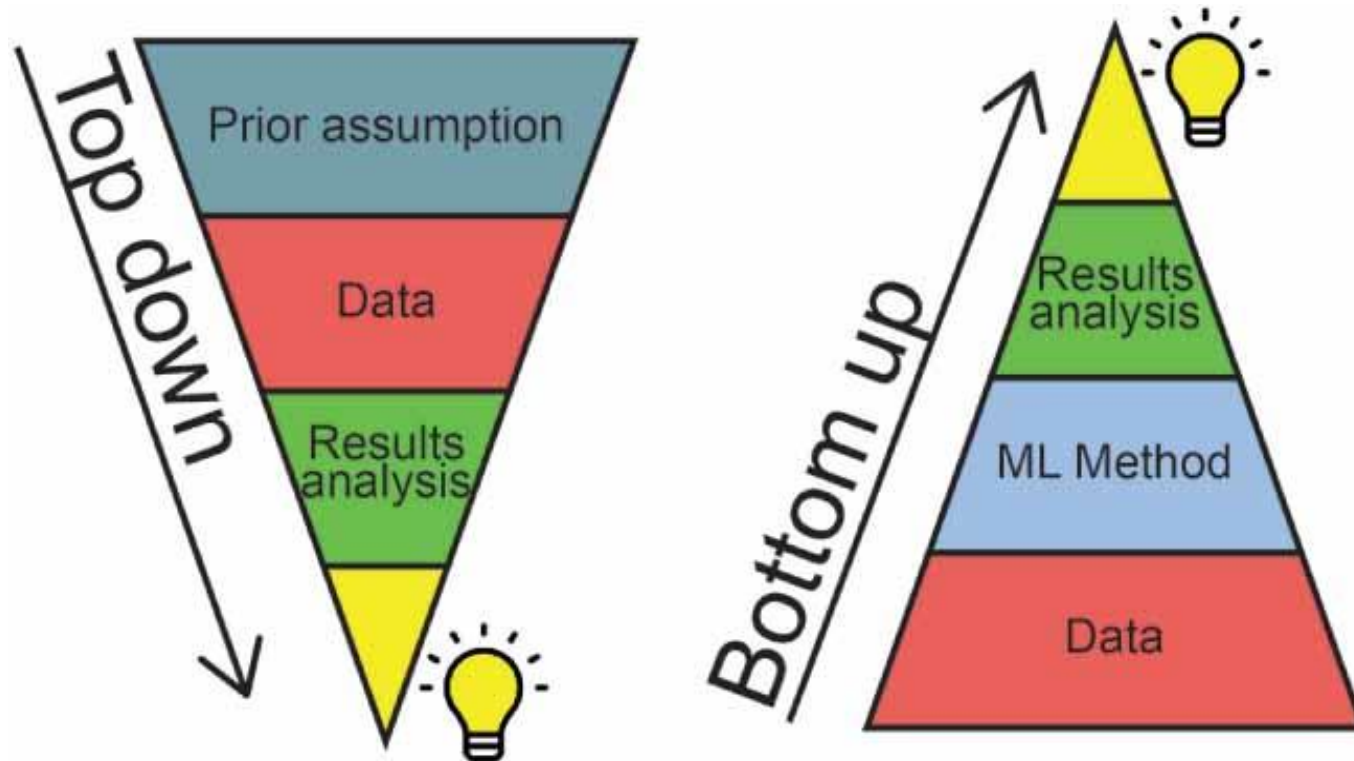
# Lähtökohta: NEMO-merimalli Saaristomerelle ja Ahvenanmerelle



- 0.25 NM horizontal resolution (~500 m)
- Up to 1 m vertical resolution
- 200 vertical levels
- $z^*$  vertical coordinates
- NEMO V4.0
- Bathymetry VELMU+BSBD
- 2013-2017
- ERA5 fcn, CMEMS bc



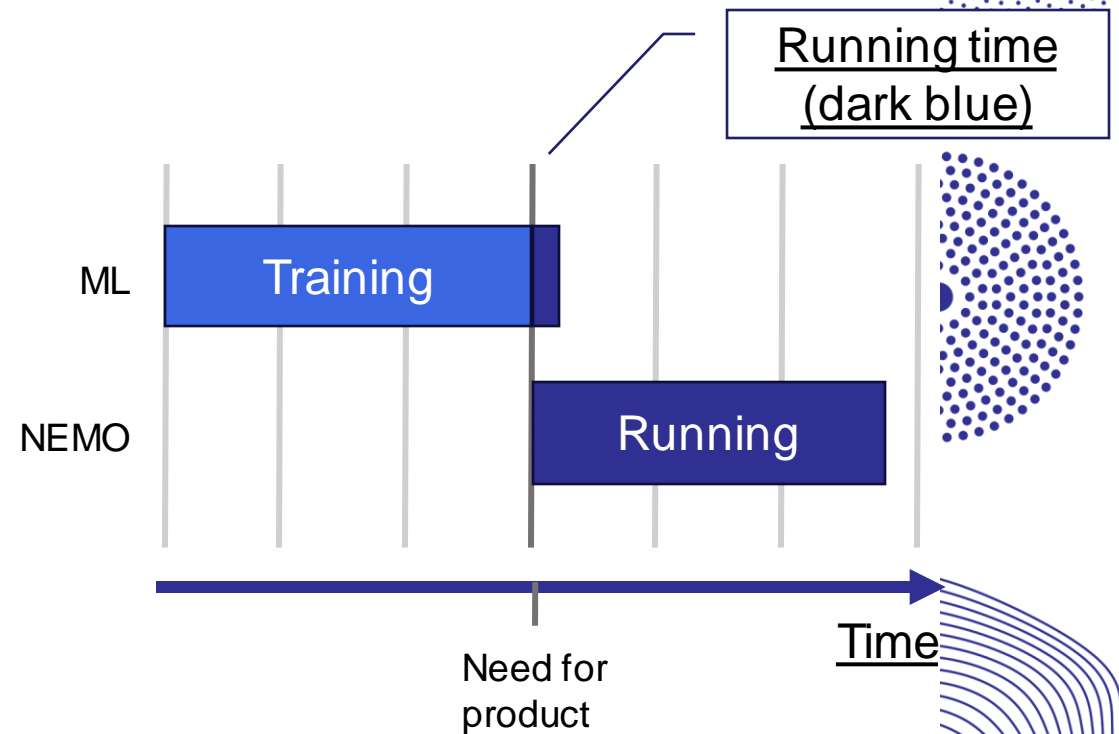
# Datalähtöiset menetelmät



Sonnewald, Maïke, et al. "Bridging observations, theory and numerical simulation of the ocean using machine learning." *Environmental Research Letters* (2021). <https://doi.org/10.1088/1748-9326/ac0eb0>

# Ratkaisu: uudenlainen merimallijärjestelmä

- **Surrogaattimalli** on toisena mallina "esiintyvä" datalähtöinen malli
- Perustuu usein koneoppimismenetelmiin
- Tyypillisesti nopea suorittaa
  - voidaan ajaa läppärillä, datasiirtotarve pienempi
- Koulutetaan esimerkiksi perinteisen simulaatiomallin tuloksilla
- Kokeillaan nyt ensimmäistä kertaa Suomen merialueilla





# Hankkeen tavoitteet

## Surrogaattimallipilotti Merisotakoulun Meritaistelukeskukselle

- Asiakkaalle hyödyllinen tuote
- Laatu riittävän hyvä sovelluskäyttöön
- Tavallista mallia lyhyempi ajoaika ja pienemmät resurssivaatimukset

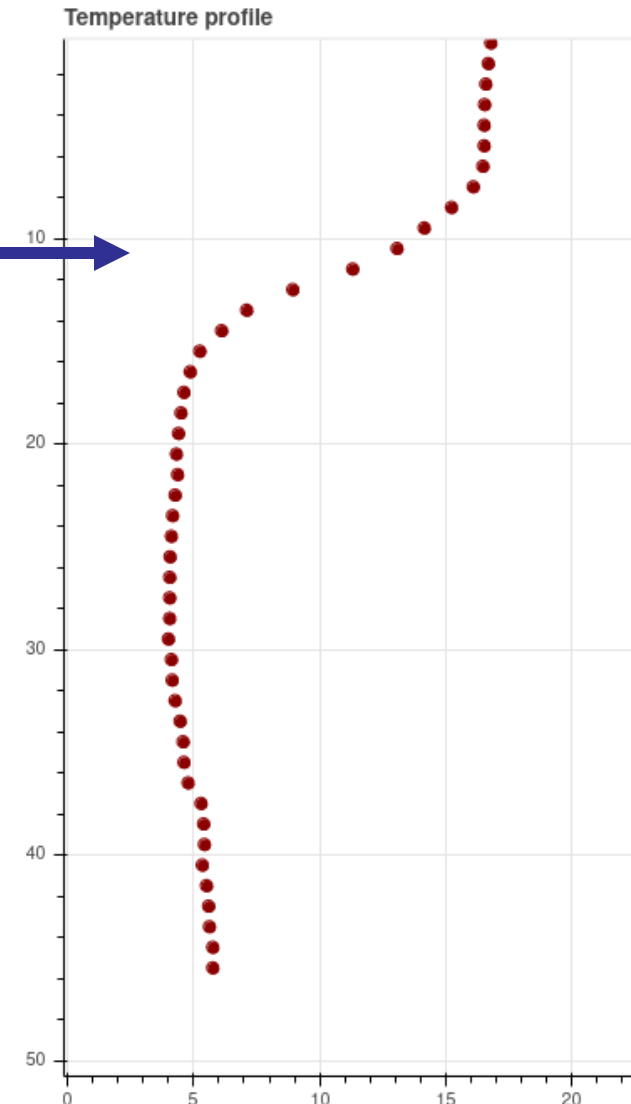
## Ensisijaisesti tarkastellaan veden kerrostuneisuutta

- avoimen veden aikana
- Utön saaren ympäristössä Saaristomerellä
- 1-4 päivän aikajänne



# Lämpötilan harppauskerros eli termokliini

- *Lämpötilan harppauskerros* eli *termokliini* on vyöhyke vesipatsaassa, jossa lämpötila muuttuu nopeasti pystysuunnassa.
- Kuvassa mallinnettu lämpötilaprofiili Utön saaren läheltä kesäkuun 2013 lopulla.
  - Pinnassa sekoittunut kerros
  - Sen alla lämpötila kylmenee nopeasti muutaman metrin matkalla = **termokliini**
  - Termokliini muodostuu vesipatsaaseen alkukesästä ja katoaa syksyllä täyskierron myötä
- Tässä hankkeessa keskityttiin mallintamaan termokliinin paikkaa

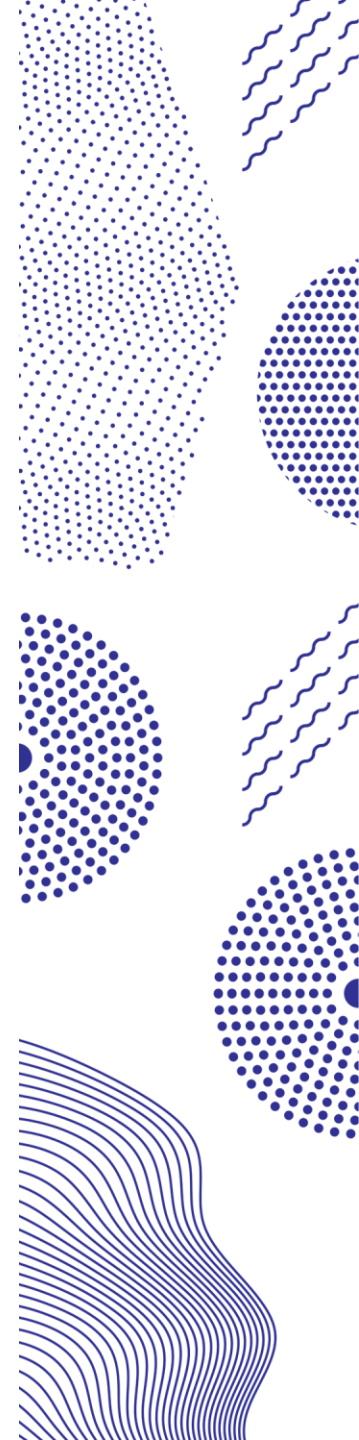


Lämpötila NEMO 0.25NM, Utö 2023-06-30

# Hankkeen eteneminen

- Soveltuvien algoritmien vertailu ✓
- Järjestelmän ensimmäisen version toteutus ✓
- Validointimateriaalin kerääminen ja mittaukset kohdealueella ✓
- Toimivuuden ja tarkkuuden arviointi ✓
- Tuotettavan pilottityökalun kehitystarpeet asiakasnäkökulmasta ✓
- Toisen version toteutus (*loppuvuosi 2023*)
- Artikkelikäsikirjoitus (*valmisteilla*)

# Hankkeen tuloksia



# Tuloksia: pilottityökalu

- Komentorivikäyttöliittymä
- Toteutettu Pythonilla
- Konfiguroidaan tekstitiedostolla
- Tuottaa NetCDF-muotoisen tiedoston
- Ajoaika suuruusluokaltaan noin 5 min, koulutusaika alle tunnin
- Ensimmäinen versio valmis, toinen valmistuu loppuvuonna

```
Open  ~ /Documents/projektit/matine_surr... Save  -  x
cfg.ini
1 # time definition
2 # yyyy-MM-dd
3 # where yyyy is the year
4 # MM is the month (1 to 12)
5 # dd is the day(1 to 31)
6 # e.g.: 01 June 2017
7 # becomes: 2017-06-01
8 [time]
9 start_datetime = 2017-07-
10 end_datetime = 2017-08-0
11
12 # domain definition
13 [domain]
14 # lat in deg N
15 # lon in deg E
16 min_lat = 59.36
17 max_lat = 60.00
18 min_lon = 20.00
19 max_lon = 23.00
20
21 # paths definitions
22 [out_paths]
23 # save results?
24 save_output = yes
25 output_file = ./interfac
26
27 [temp_paths]
28 # save intermediary resu
29 save_temp = no
30 temp_folder_path = ./int
31
32 [path_model]
33 model_path = ./interface
34 model_file_name = LR
35 model_file_extension = p
36
.espinola@puhti-login15:~/Interface
Welcome to Rapid and agile ocean forecasting with surrogate modelling
V.1.9.38r2 (Nov 2023)
Loading configuration file...
#####
Make predictions from 2017-07-31 until 2017-08-01 (YYYY-MM-DD)
#####
Starting program.
[ 2%] |* Loading and preprocessing bathymetry...
[ 4%] |*** Loading bathymetry data...
[ 6%] |***** Processing bathymetry data...
[ 16%] |***** Building model grid...
[ 18%] |***** Loading and preprocessing ERA5 data...
[ 20%] |***** Regridding ERA5 data...
[ 23%] |***** ERA5 Feature engineering...
[ 44%] |***** Making data tabular and splitting domain...
[ 53%] |***** Feature engineering...
[ 79%] |***** Making predictions...
[ 95%] |***** Compiling results...
Saving predictions...
Predictions saved in file: /scratch/project_2001981/espino/interfac/out/thermocline.nc
[100%] |*****
##### Program completed. #####
#####
Summary statistics for the predictions:
          mld_pred
count 252648.000000
mean   12.872058
std    1.826835
min    9.351611
25%   11.636153
50%   12.671737
75%   13.901323
max    17.814205
#####
bash-4.4$
```

# Tuloksia: aika- ja resurssitarpeet

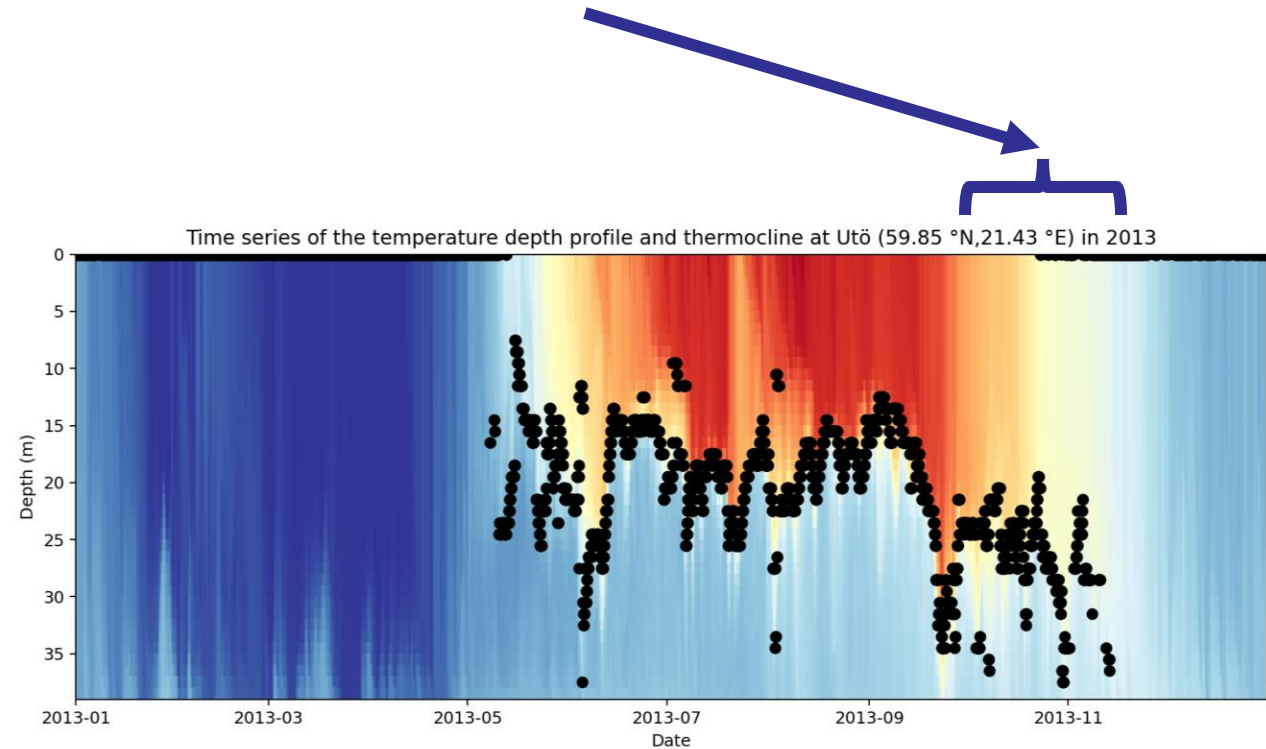
- Ajoalue suuruusluokaltaan n. 30 km X 90 km, resoluutio n. 500 m
- Ajoaika läppärillä noin 5 minuuttia viikon ajolle (4 h välein)
- Koulutusaika superkoneella alle tunnin
  - Riippuu valitusta algoritmista, käytettävissä olevasta kapasiteetista, jne.
  - Joillain algoritmeilla voi olla merkittävästi pidempi (random forest)
  - Koulutusta ei tarvitse tehdä joka ajon yhteydessä
- Johtopäätös: järjestelmä toteutettavissa nykyisellä laskentateholla

# Tuloksia: lähestymistapojen vertailu ja arviointi

- Testattiin harppauskerroksen mallintamiseen sopivia koneoppimisalgoritmeja
  - Lineaariset algoritmit (vakaa vaihtoehto)
  - Random Forest (ei kustannustehokas)
  - Conv3D U-Net (parameterisointi hankalaa, tarvitsee paljon dataa)
  - Vielä testattavia: LSTM, LSTM U-Net, luokittelumenetelmät, ensemblemenetelmät
- Johtopäätöksiä seuraaviin versioihin:
  - Koulutusmateriaalin lisääminen tarpeen
  - Koneoppimisarkkitehtuurin virittämiseen menee huomattavasti aikaa
  - Syötteiden analyysi näytti, että mm. pohjan muodot, tuuli ja säteilypakote merkittäviä mallille
  - Hyvä tieto alkutilasta tärkeä

# Termokliinin syvyys ja jatkuvuus

- Termokliinin syvyyden muutokset ovat epälineaarisia ja riippuvat epälineaarisisista prosesseista (tuulen leikkausjännitys, jne.)
- Pakotteen muutosten ja termokliinin syvyyden muuttumisnopeuden välillä on epälineaarinen viive
- Termokliinin syvetessä pohjaan asti termokliiniä ei enää ole, mikä on myös epälineaarinen muutos
- Rikkonainen mallinnusalue saarineen ja rannikoineen tarkoittaa spatiaalista epäjatkuvuutta



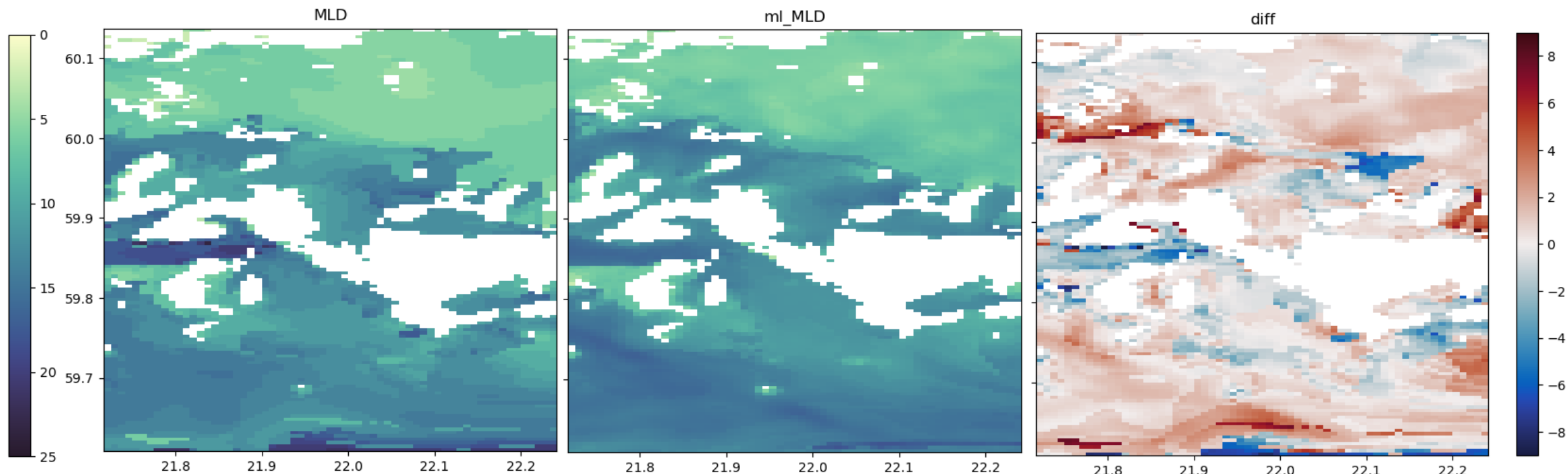
- Epäjatkuvuudet haaste koneoppimisalgoritmeille
- Näihin ongelmiin voidaan osittain vastata käyttämällä mukautettua häviöfunktioita (custom loss function)



# Tuloksia: soveltuvuus käyttöön

- Järjestelmän kehitysversion pohjalta arvioitiin sen käyttökelpoisuutta asiakkaan kanssa
- Keskustelun myötä tunnistettiin seuraavat kehityskohteet
  - Ajoalueen täsmennys
  - Ennustepituuden keskittyminen noin viikon aikajänteelle
  - Reunaehtokäsittelyn uudistaminen
  - Tulosten odotetun tarkkuuden täsmentäminen

# Tuloksia: termokliinin syvyys



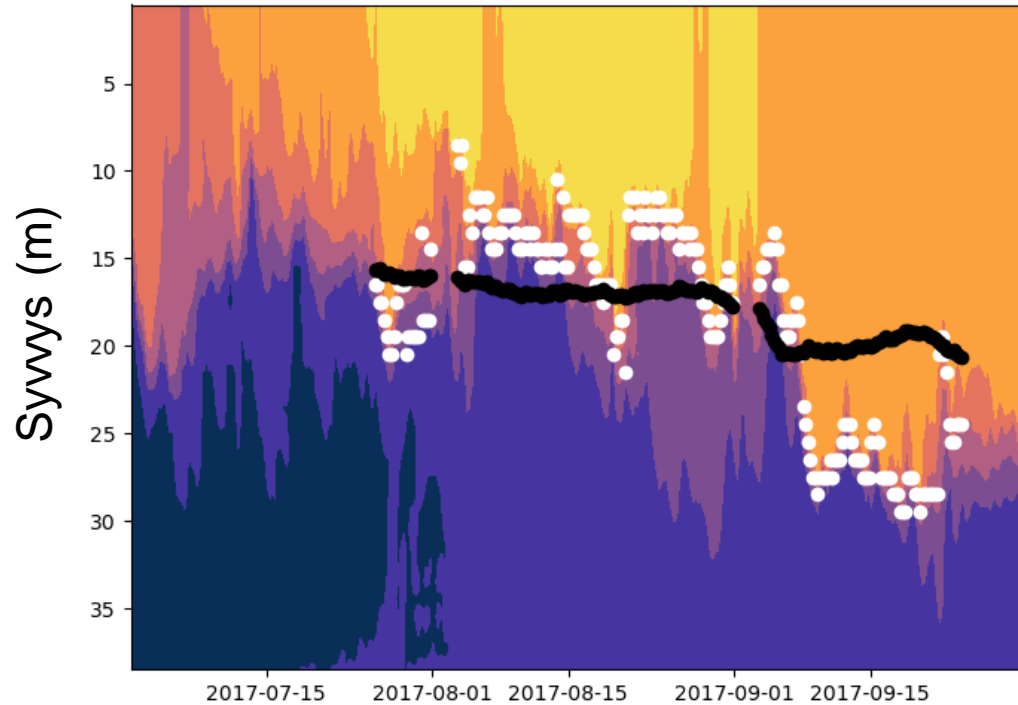
26.07.2017 klo 9:00; 1 päivän termokliiniennuste

NEMO (vas.), koneoppimisjärjestelmä (kesk.), erotus (oik.)

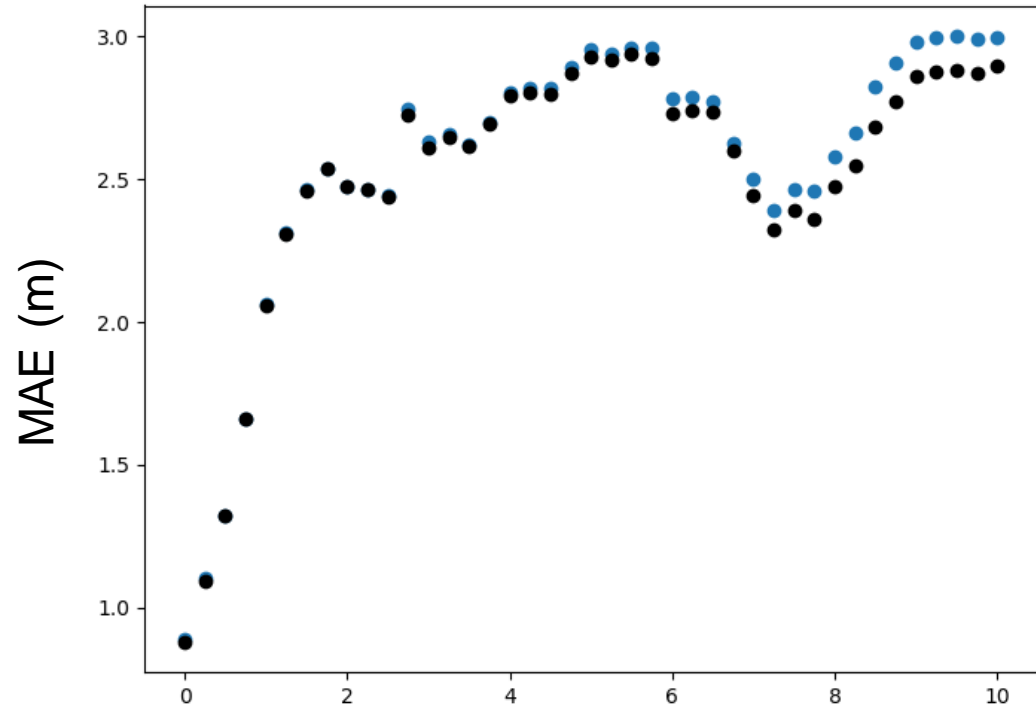
Tulokset näyttävät järkeviltä jopa lineaarisella regressiolla

# Tuloksia: termokliinin syvyys

Termokliinin syvyyden mallinnettu aikasarja,  
Utö 26.7.2017 alkavalle ennusteelle  
valkoinen = NEMOsta analysoitu  
musta = koneoppimisennuste



Ennusteen keski-itseisvirhe (mean absolute error, MAE)  
koko ennustealueella ajan funktiona  
tavallinen (sininen) ja mukautettu häviöfunktio (musta)



# Mahdollisia jatkokehityssuuntia

- Pilotissa tutkittiin lämpötilan harppauskerrosta, mutta jatkokehitystä voidaan suunnata myös esimerkiksi pintalämpötilan, -suolaisuuden ja pinnankorkeuden mallintamiseen
  - Yksivaiheisesta monivaiheiseen prosessiin, jonka tuloksena myös harppauskerroksen syvyys
- Kehitettyjä menetelmiä voitaisiin käyttää myös esimerkiksi ajelehtimislaskelmien kehityksessä (mm. öljyvuodot ja vaarallisten aineiden leviäminen)
- Hyperresoluutiomallit? Puuttuvien arvojen täyttö? Numeeristen mallien biaskorjaus?

# Yhteenveto

- Hankkeessa kehitettiin uudenlaisen merimallijärjestelmän pilotti, joka pohjautuu surrogaattimalleihin.
- Surrogaattimalli on toisena mallina "esiintyvä" datalähtöinen malli
  - Koneoppimispohjainen
  - Ajoaika ja resurssivaatimukset perinteisiä malleja pienemmät
- Hyvä ensimmäinen askel, kehitystyö jatkuu toisissa hankkeissa