# New assessment of MyOcean forecasting systems in European Seas using process oriented metrics

Jan Maksymczuk, UK Met Office

- As part of the MyOcean project, 3 institutions worked to adapt and trial NWP validation techniques to ocean products from the MyOcean catalogue:
  - Biogeochemistry-specific metrics (OD Nature) - Katrijn Baetens
  - Site-specific metrics for ocean currents (UK Met Office) - Ray Mahdon
  - Triple-collocation methods (Mercator-Ocean) - Marie Drevillon

- This presentation gives a brief overview of results from these investigations

- Further results are available from the CMEMS validation website:
  - http://marine.copernicus.eu/web/103-validation-statistics.php


- Simple metrics such as bias and RMS errors are routinely applied to products from the MyOcean catalogue
- Whilst useful, these metrics can sometimes be of limited value in
  - identifying timing or spatial errors
  - capturing intense or high impact events
- Atmospheric science has a long history of developing more informative statistics

# Categorical metrics and neighbourhood methods

• Categorical metrics transform a continuous dataset to a binary data set by determining if the value of the continuous data set exceeds a certain user specified threshold (yes = 1), or not (no = 0), which allows a *contingency table* to be constructed

| Event Forecast | Event Observed | | |
|---|---|---|---|
| | YES | NO | Marginal Total |
| YES | a | b | a + b |
| NO | c | d | c + d |
| Marginal Total | a + c | b + d | a + b + c + d |

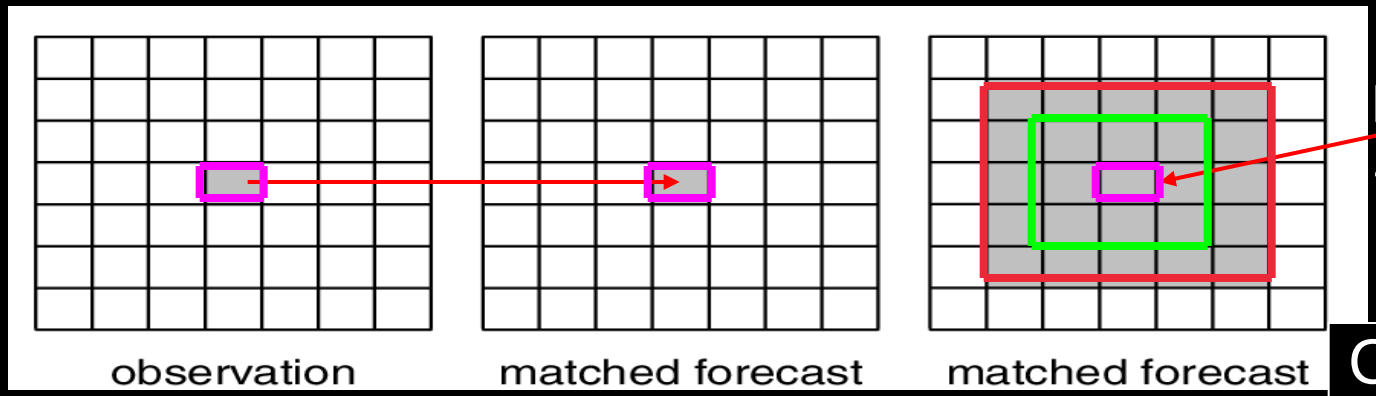
• From these contingency tables a number of categorical metrics can be derived
• In a perfect system *a* and *d* would be maximised, whilst *b* and *c* would be zero
    • There may be cost implications for users if a False Alarm (b) is raised, for example, unnecessarily implementing contingency plans
    • A miss (c) can lead to damage to reputation and lack of user confidence, especially if it occurs too many times

- Closeness is not rewarded using traditional validation techniques:
    - usually rely on extracting the nearest grid point or using bilinear interpolation in order to produce a matched forecast/observation pair
    - forecasts that are slightly offset with regards to the observation are penalised as severely as forecasts that are significantly offset

- Neighbourhood techniques are a departure from traditional validation approaches:
    - compare single observations to a forecast region around the observation location, either temporally or spatially

- Neighbourhood techniques are particularly suited to dealing with the ‘double penalty problem’:
    - often occurs when validating higher resolution models
    - penalised for both timing and location errors
    - for example, if the shape or intensity of an eddy is correctly simulated, but forecast with some time lag, or not at the exact location

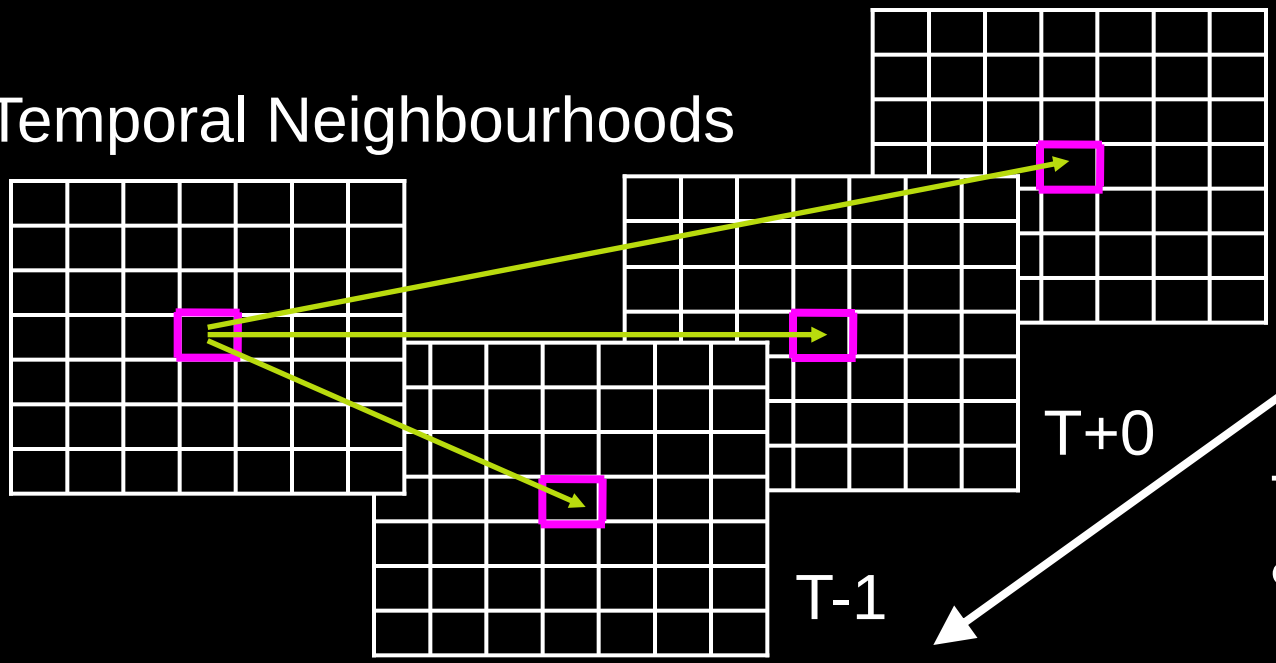
# Spatial Neighbourhoods




Neighbourhoods:
1x1, 3x3, 5x5,..,NxN

Combinations spatial & temporal neighbourhoods trialled

# Temporal Neighbourhoods




Time averaging & shifting

# Neighbourhood examples from an atmospheric 1.5km model

3 x 3 7 x 7 17 x 17

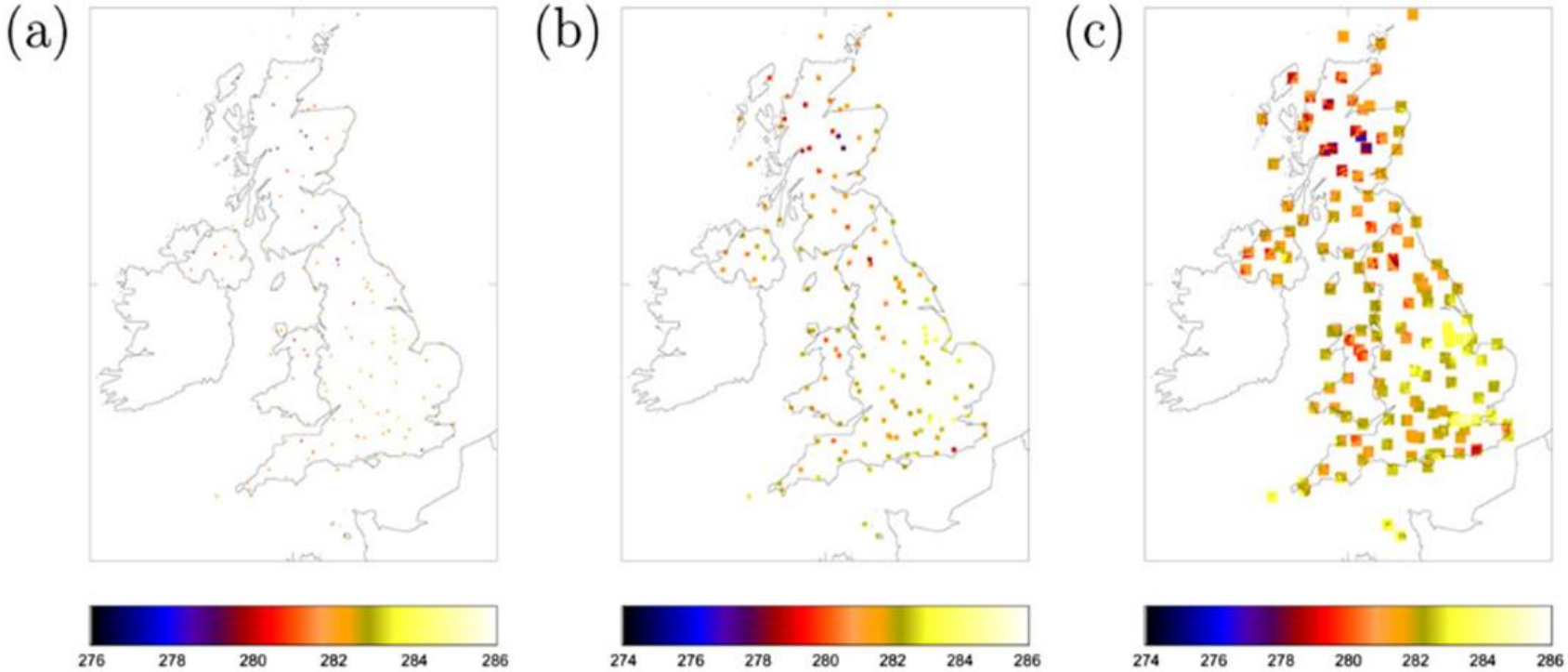



FIG. 3. An illustration of neighborhood sizes around each UK observing site based on 1.5-m temperature (K)

Met Office

# Biogeochemistry-specific metrics:

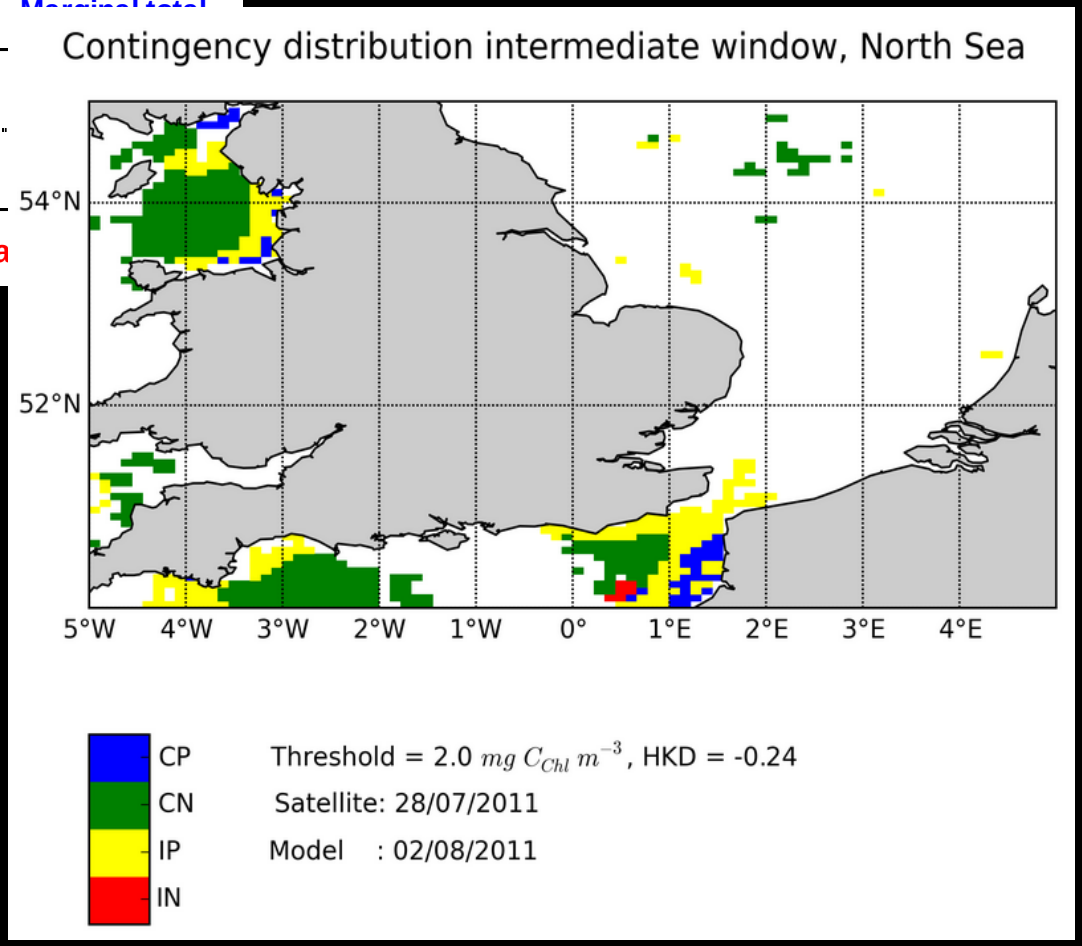
Katrijn Baetens, Sebastien Legrand

# Biogeochemistry metrics

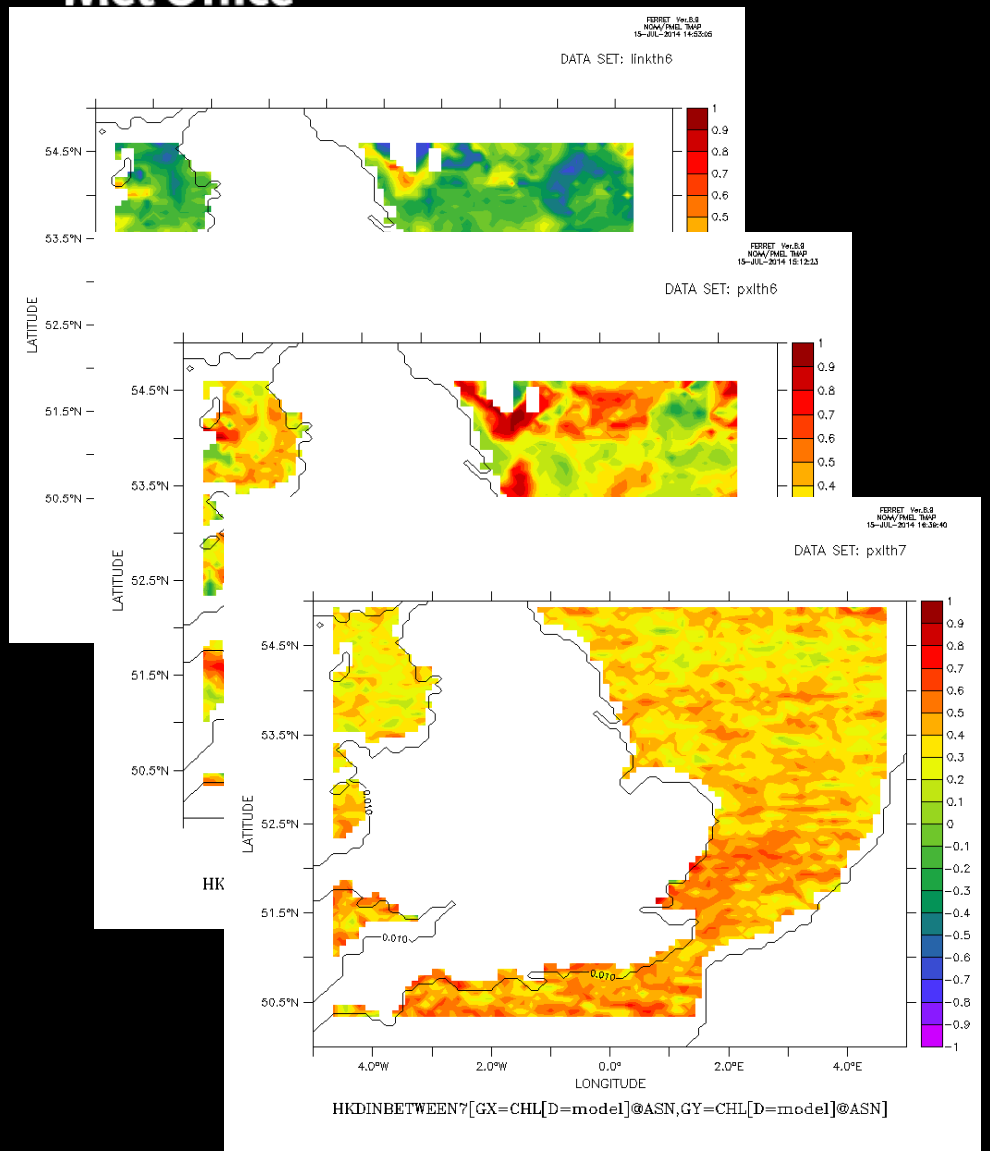
- Forecasts of chlorophyll-a compared against satellite products
- North Sea, Mediterranean Sea and Baltic region assessed
- Categorical metric Hanssen-Kuippers Discriminant (HKD) assessed
    - Range: -1 to 1, 0 indicates no skill
- Neighbourhood techniques trialled
    - up to +/- 6 units of model resolution temporally
    - up to 3 grid cells spatially
- Various combinations of temporal and spatial averaging tested

# Biogeochemistry metrics

| Event forecast | Event observed | | |
|---|---|---|---|
| | Yes | No | Marginal total |
| Yes | a | b | |
| No | c | d | |
| Marginal total | a + c | b + d | a |



# Biogeochemistry metrics




No neighbouring distance applied – poor agreement, low/-ve HKD scores

Temporal neighbouring distance of +/- 6 days applied – good agreement, high/+ve HKD scores

Spatial permutation with 3 grid cells applied – model skill smoothed, features lost

# Biogeochemistry metrics

## VALIDATION STATISTICS

MFCs: global arctic balticsea nws ibi medsea TACs: sealevel sst wind oceancolour insitu novelmetrics

biogeochemistry ocean-currents triple-collocation

continuous-metrics categorical-metrics p90

**Area Timeseries**

baltic-sea-flexible-translation

med-sea-flexible-translation

north-sea-flexible-translation

**Depth**

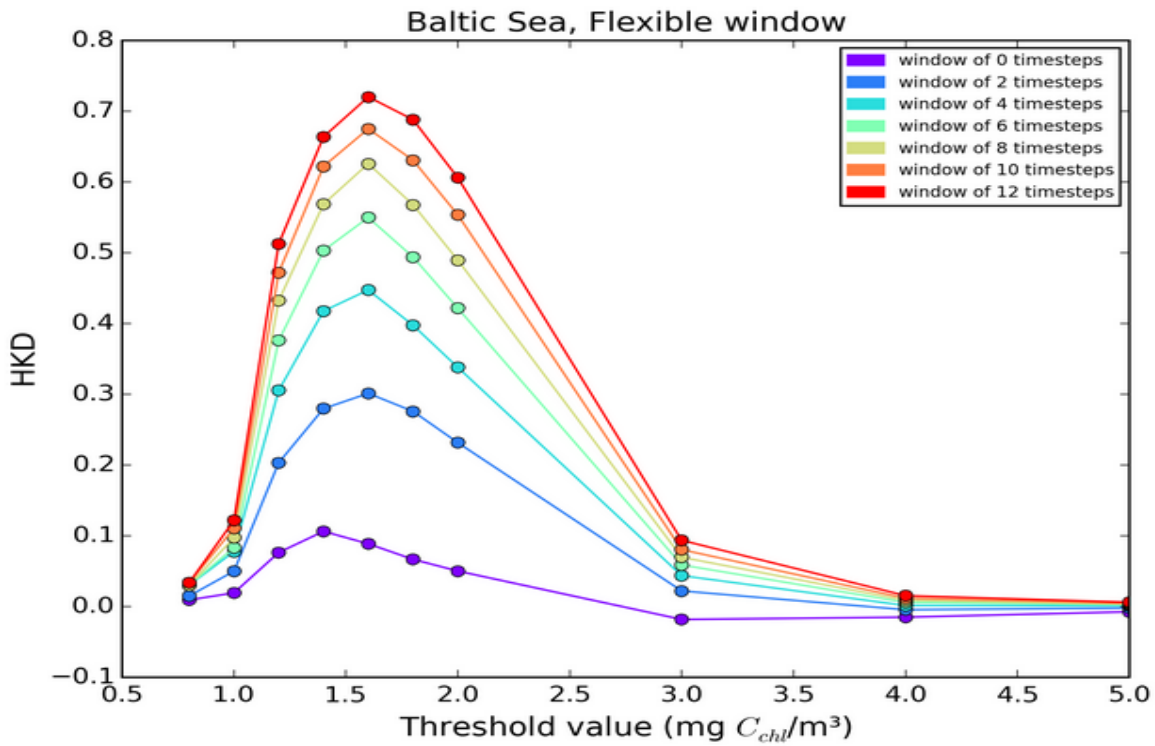
surface

**About**

These pages contain summary statistics for MyOcean products. More detailed validation results can be found in a Quality Information Document (QuID) for each product, which can be accessed via the MyOcean catalogue

For further information please contact: MyOcean Service Desk

Last Updated: 18 May 2015




Threshold value (mg $C_{chl}$/m³)

Threshold value (mg $C_{chl}$/m³)

Met Office

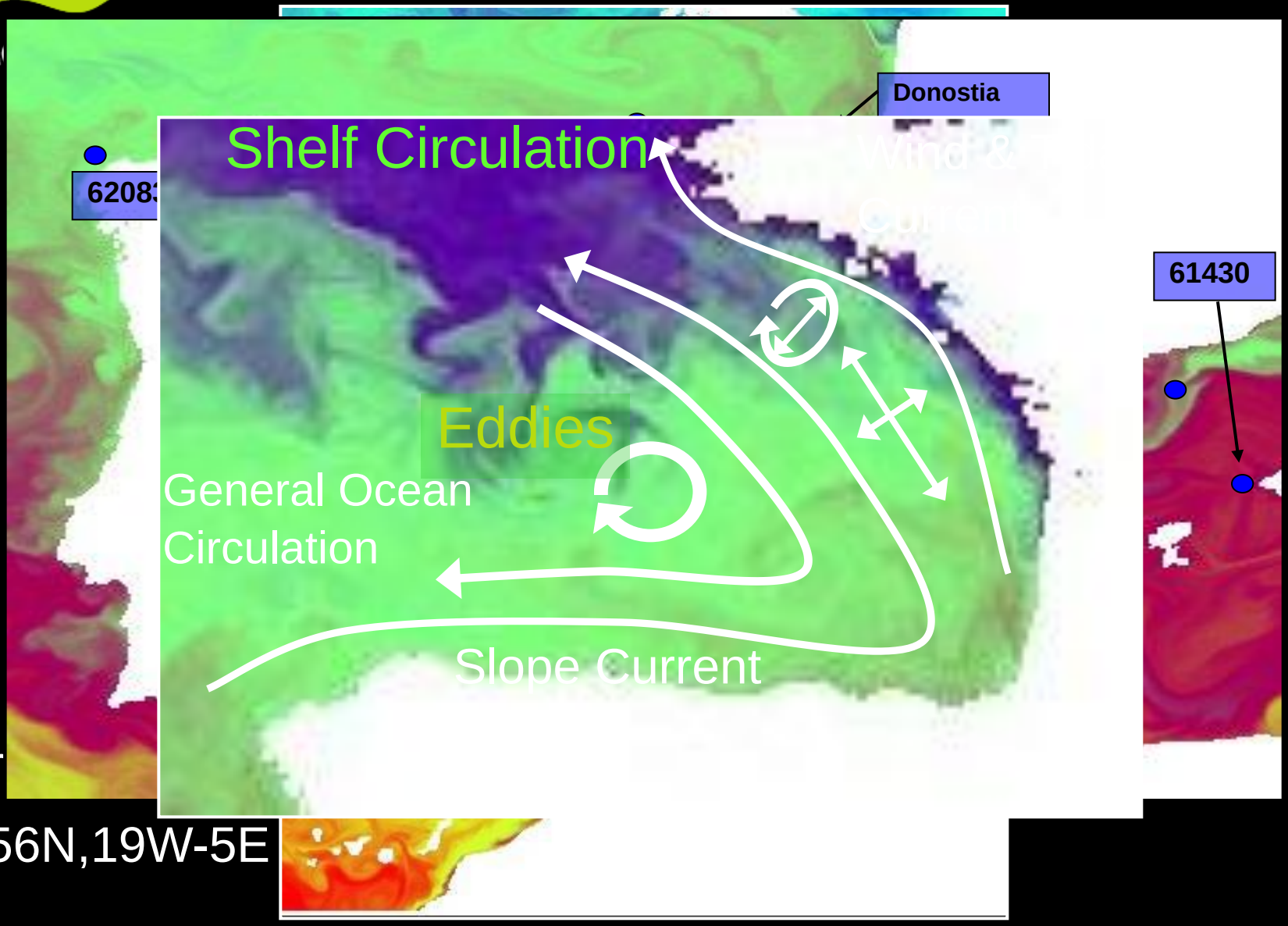
# Site-specific metrics for ocean currents:

Ray Mahdon

# Ocean currents metrics

- Surface currents forecasts are important for commercial or defence weather-windows
  - e.g. Current speed below 1kt for 12 hours.
  - e.g. Does not exceed 1kt more than x times
- Does the model capture extreme events?
- In which locations or time of year do the models have the best performance?

# Ocean currents metrics



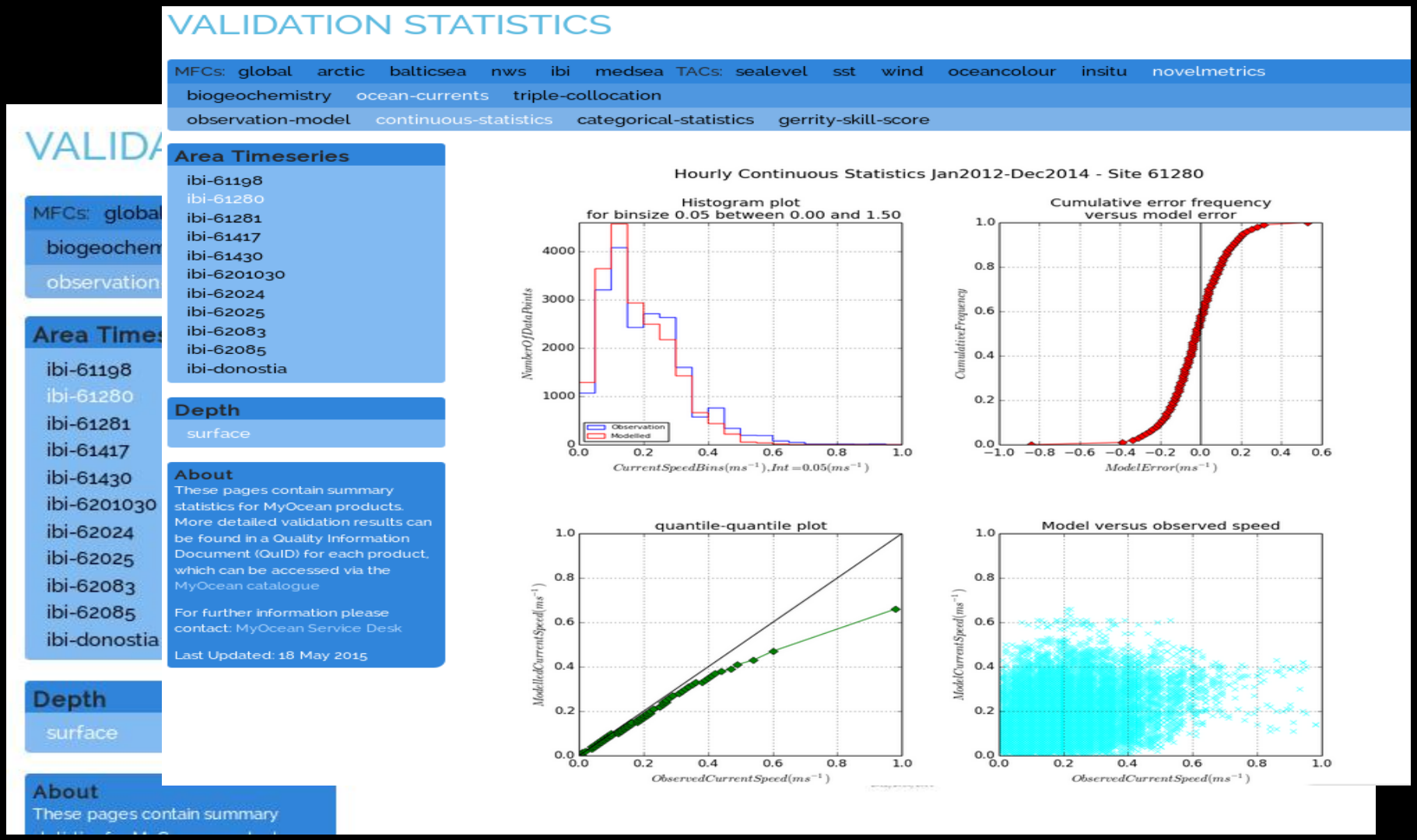

IBI r

26-56N,19W-5E

# Ocean currents metrics

- Continuous statistics are often helpful to describe overall behaviour
  - e.g. q-q & histogram plots describe climatology
  - time series can show seasonal patterns or significant events
- Continuous statistics do not quantify the performance of a system when exceeding thresholds or high impact events are of interest

# Ocean currents metrics



# Ocean currents metrics

- Categorical assessment performed on hourly data from Jan 2012–Dec 2014
  - simple metrics assessed initially (ETS, hit rate, CSI,...)
- Combinations of spatial and temporal neighbourhoods trialled
- Multi-event contingency tables analysed using Gerrity Skill Score
  - good observation (sample) climatology required
  - rewards rare/extreme events
  - does not reward (penalises) conservative forecasting
- Combinations of spatial and temporal neighbourhoods trialled
- Multi-category methods on surface ocean current speed are relatively new, so expectation of skill level is unknown

# Ocean currents metrics

## VALIDATION STATISTICS

MFCs: global arctic balticsea nws ibi medsea TACs: sealevel sst wind oceancolour insitu novelmetrics
biogeochemistry ocean-currents triple-collocation
observation-model continuous-statistics categorical-statistics gerrity-skill-score

**Area Timeseries**

- ibi-61198
- ibi-61280
- ibi-61281
- ibi-61417
- ibi-61430
- ibi-6201030
- ibi-62024
- ibi-62025
- ibi-62083
- ibi-62085
- ibi-donostia

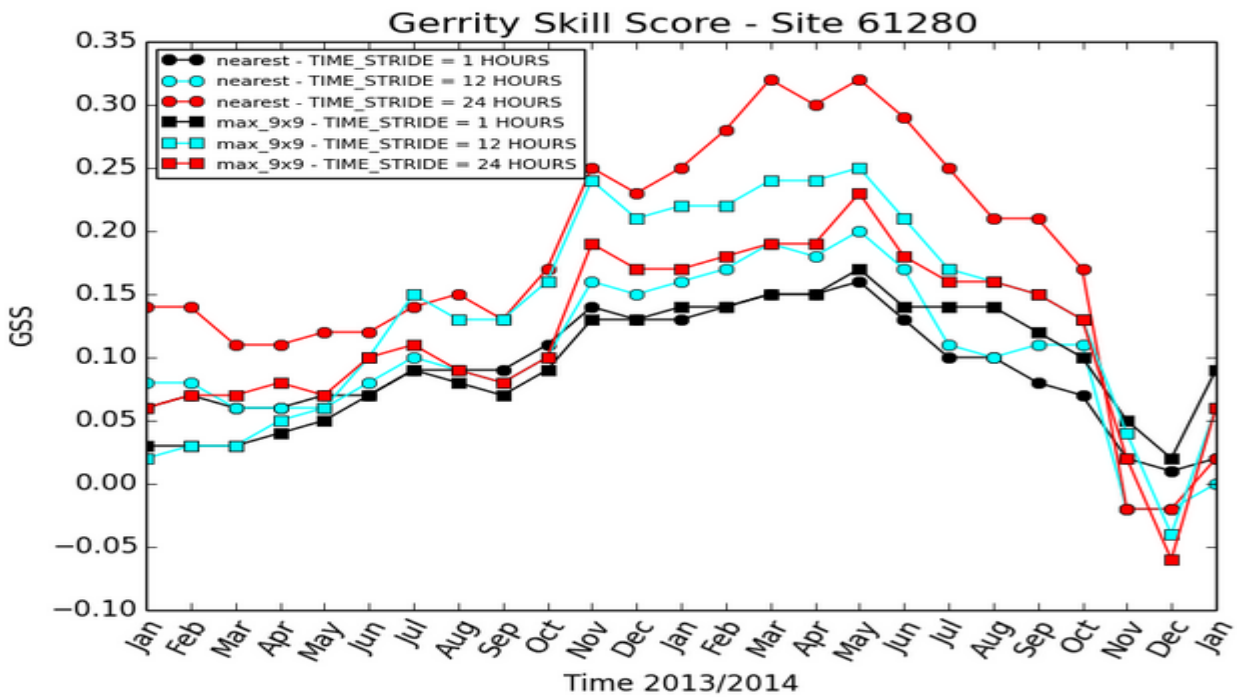
**Depth**

surface

**About**

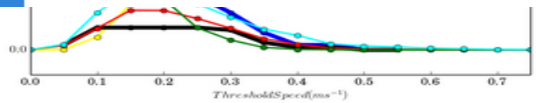
These pages contain summary statistics for MyOcean products. More detailed validation results can be found in a Quality Information Document (QuID) for each product, which can be accessed via the MyOcean catalogue

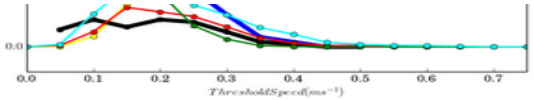
For further information please contact: MyOcean Service Desk




For further information please contact: MyOcean Service Desk

Last Updated: 18 May 2015





Met Office

# Triple-collocation methods for SST:

Marie Drevillon, Bruno Levier

# Triple collocation for SST

- Triple collocation is a method used to compare different datasets of the same geophysical field
- It provides an estimate of the error of each dataset, and can be used to calibrate the datasets
- Long time series of data required for meaningful conclusions
- In-situ drifting buoys, satellite (AVHRR METOP-A) and forecast systems (GLO, NWS, IBI) were collocated for SST
- Due to data assimilation the datasets are not truly independent

# Triple collocation for SST

## VALIDATION STATISTICS

MFCs: global arctic balticsea nws ibi medsea TACs: sealevel sst wind oceancolour insitu novelmetrics

biogeochemistry ocean-currents triple-collocation

sst

**Area Timeseries**

insitu-locations
measurements-per-month
error-variance-triplets

**Depth**

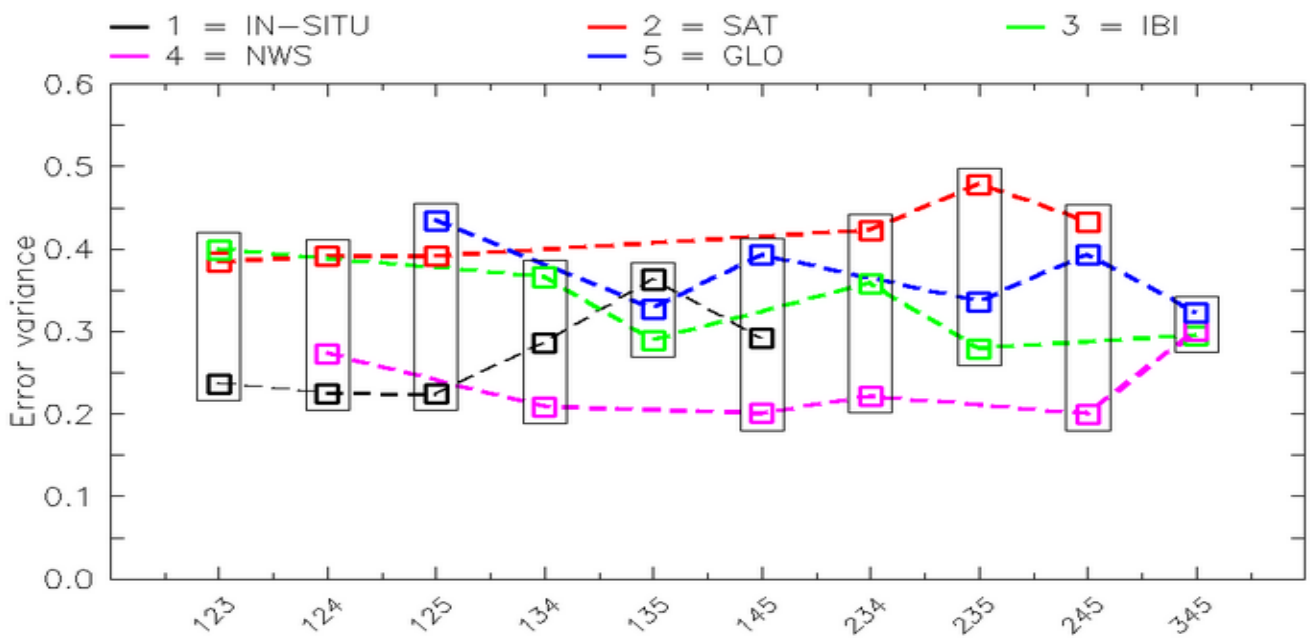
surface

**About**

These pages contain summary statistics for MyOcean products. More detailed validation results can be found in a Quality Information Document (QuID) for each product, which can be accessed via the MyOcean catalogue

For further information please contact: MyOcean Service Desk

Last Updated: 18 May 2015



# Conclusions from metrics investigations

- Biogeochemistry:
  - Application of the Hanssen Kuippers Discriminant and neighbourhood techniques has shown that it is able to detect systemic time lags of the model forecasts
  - Expanding the time frame of the model does not lead to a huge improvement in model skill
  - The neighbourhood techniques are not a means to improve the model performance in an unjustified way

- Ocean currents:
  - The Gerrity Skill Score (GSS) reflects what can be deduced from the continuous statistics and the simpler categorical metrics
  - Modelled currents show some skill when both the modelled and observed currents are coarsened temporally
  - Where the model has a good representation of the current distribution, as well as an accurate representation of the phase of the current cycle, the GSS is higher relative to those that are out of phase or do not have a good current distribution or climatology
  - The GSS is a supplemental score to the other statistical metrics examined - could be a useful summary metric, but may be too complex for basic user requirements

- Triple collocation:
  - Several applications of the method were tested with Sea Surface Temperature
  - A long time period is required in order to gather enough data, so it is more suited to delayed time studies
  - The data sets compared are not independent due to data assimilation in operational systems
  - The data sets do not necessarily represent the same space and time scales
  - The error variances obtained are not obvious to interpret - the method should be seen as a complement to other statistical methods

Met Office

Thank you - any questions?