COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

European Union Programme

Copernicus Europe's eyes on Earth

MERCATOR OCEAN
OCEAN FORECASTERS
Entrusted entity by mandate of the European Commission

Met Office

EGU session OS4.7
High-resolution model Verification Evaluation (HiVE). Part 2: Using object-based methods for the evaluation of chlorophyll blooms

Marion Mittermaier, Rachel North,
Christine Pequignet and Jan Maksymczuk

Objectives of the HiVE project

The **High-resolution model Verification Evaluation (HiVE) project** considered CMEMS forecast products applicable to regional domains and aimed to demonstrate, for the first time, the utility of spatial verification methods (originally developed to evaluate km-scale forecasts from atmospheric models), for verifying km-scale ocean model forecasts. It was undertaken to address the need for new metrics adapted to the increased resolution in both observations and models, as identified in the CMEMS Service Evolution Strategy.

The project had two key objectives relating to the ongoing assessment protocols for ocean forecast models, and how they could be evolved to cope with future modelling systems.

1. *To understand the accuracy of CMEMS products at specific observing locations using neighbourhood methods and ensemble techniques – see* **High-resolution model Verification Evaluation (HiVE). Part 1: Using neighbourhood techniques for the assessment of ocean model forecast skill**

2. *To understand the skill of CMEMS products in forecasting events or features of interest in space and time*

Context

Understand the skill of CMEMS products in forecasting events or features of interest in space and time.

- Trial the use of an object-based spatial method called MODE (Method for Object-based Diagnostic Evaluation) to evaluate the evolution of events in both forecast and observation fields.
- MODE aims to evaluate forecast quality in a manner similar to that of a user making a subjective assessment. Object-based verification methods were developed to provide an objective link to the way forecasts are used subjectively, i.e. focusing on features or events of interest.
- Outcomes will have applications when monitoring feature evolution (for example, eddies, chlorophyll blooms), and could be extendable and applicable to global model assessments.

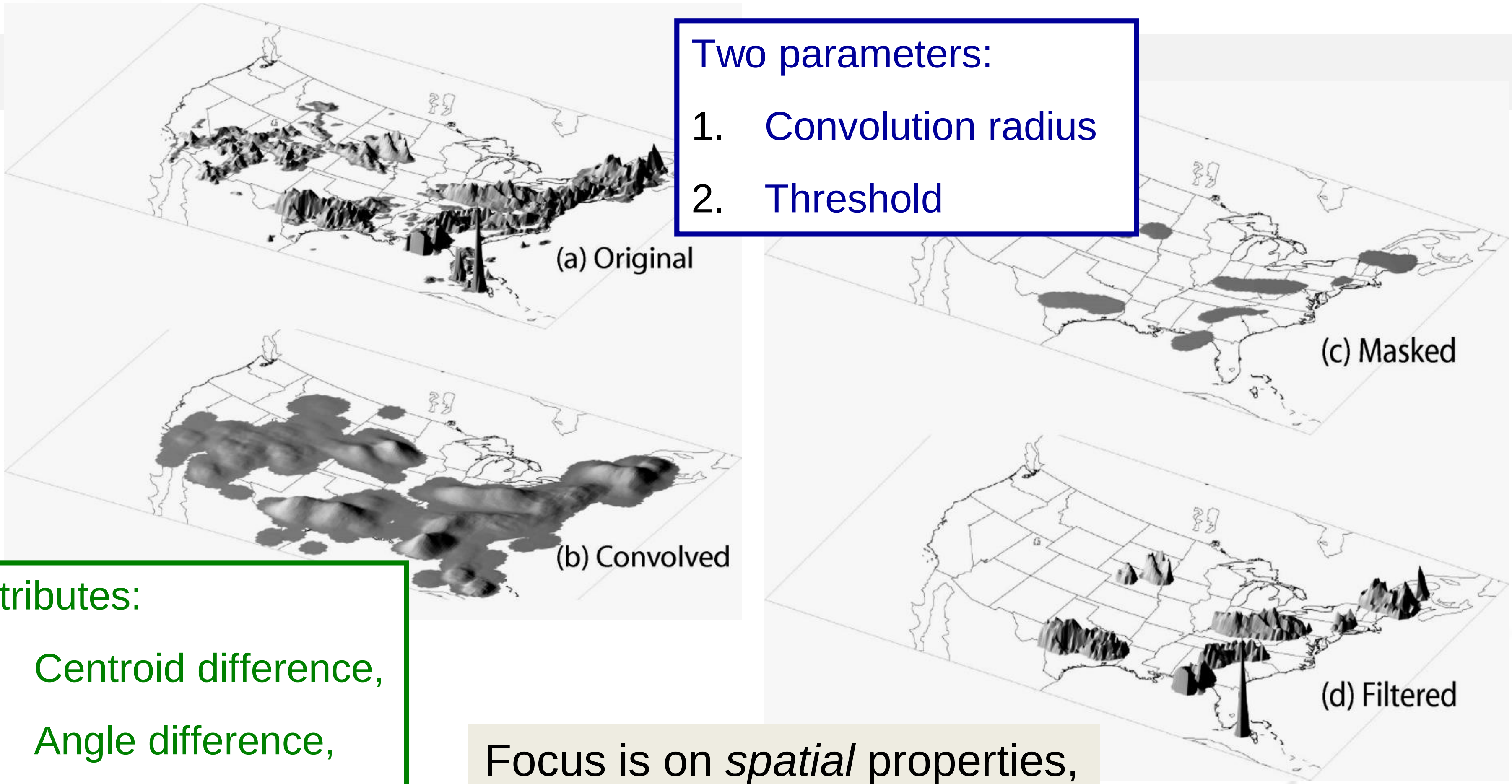
COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

MODE – Method for Object-based Diagnostic Evaluation

- The **Method for Object-Based Diagnostic Evaluation (MODE)** tool was developed in response to a need for verification methods that can *provide diagnostic information which is more intuitively useful and visual* than the information that can be obtained from traditional verification metrics, especially for high-resolution NWP output.
- MODE was originally demonstrated for precipitation forecasts, but it can also be applied to other fields with coherent spatial structure and can in theory be applied to any fields with coherent spatial structures, provided a gridded analysis exists.
- MODE can be used in a very generalised way, comparing two fields: in this context one is a forecast, the other an observation-based or model-based analysis.
- MODE identifies objects in both the forecast and observed fields. These objects mimic what humans would do to define features of interest.
- Characteristics or attributes of objects can be analysed to understand how the feature is being forecast. Objects can also be matched or paired up between the forecast and observations to analyse the attributes of the matched pairs, e.g. overlap, mismatch, distance between centre of objects.
- Summary statistics describing the objects and object pairs are produced. These statistics can be used to identify similarities and differences between forecast and observed objects, which can provide diagnostic insights of forecast strengths and weaknesses.
- The object time evolution can be analysed with MODE Time Domain (MTD).

MODE – Method for Object-based Diagnostic Evaluation
Davis et al., *MWR*, 2006

Highly configurable



Two parameters:

1. Convolution radius
2. Threshold

Attributes:

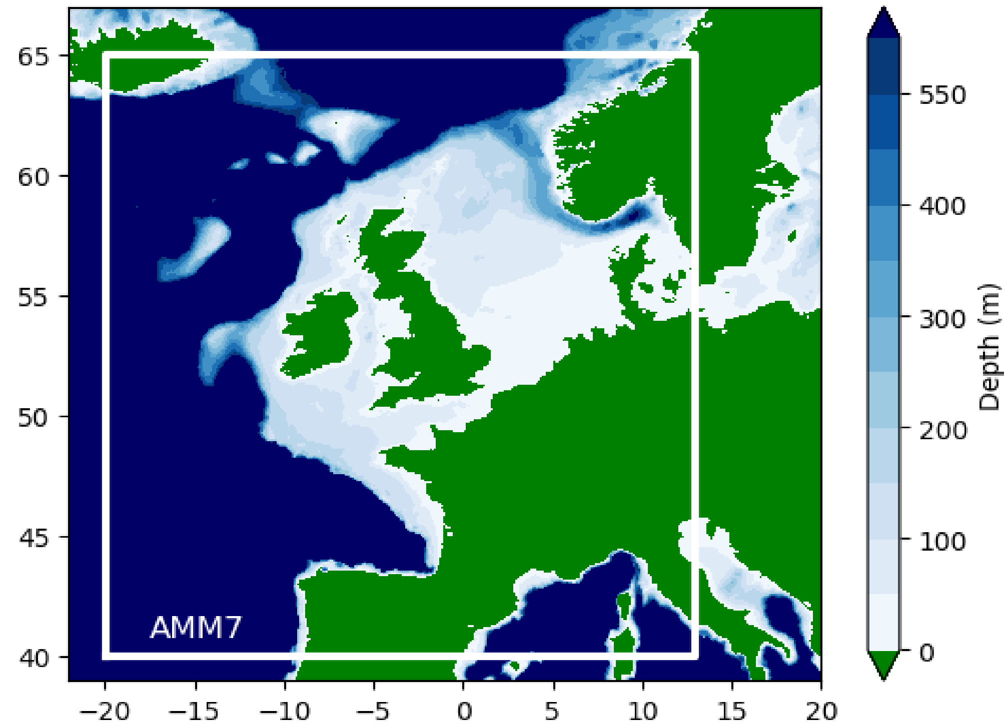
- Centroid difference,
- Angle difference,
- Area ratio etc

Focus is on *spatial* properties, especially the *spatial* biases

Model Domains

Assessments of **chlorophyll concentration** forecast were made using satellite observations and model products from the CMEMS catalogue on the European North West Shelf domain during the 2019 bloom season:

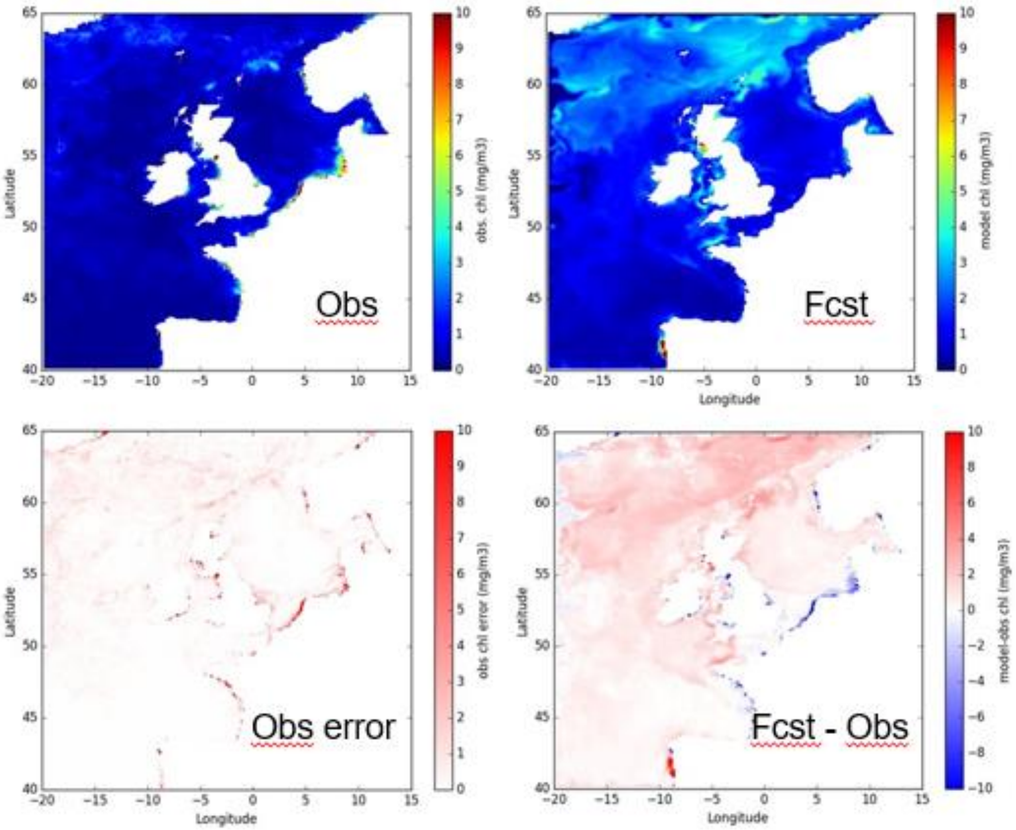
- The L4 NRT multi-sensor ocean colour Chlorophyll (~1km) (hereafter labelled L4) OCEANCOLOUR_ATL_CHL_L4_NRT_OBSERVATIONS_009_037
- The Met Office chlorophyll forecast from the FOAM-ERSEM on the AMM7 (1/10°) domain (labelled AMM7v8) NORTHWESTSHELF_ANALYSIS_FORECAST_BIO_004_002_b



- In addition, a new research chlorophyll analysis product based on FOAM-ERSEM and using ocean colour data assimilation was also compared (hereafter labelled AMM7v11 BIO DA)

Understanding biases

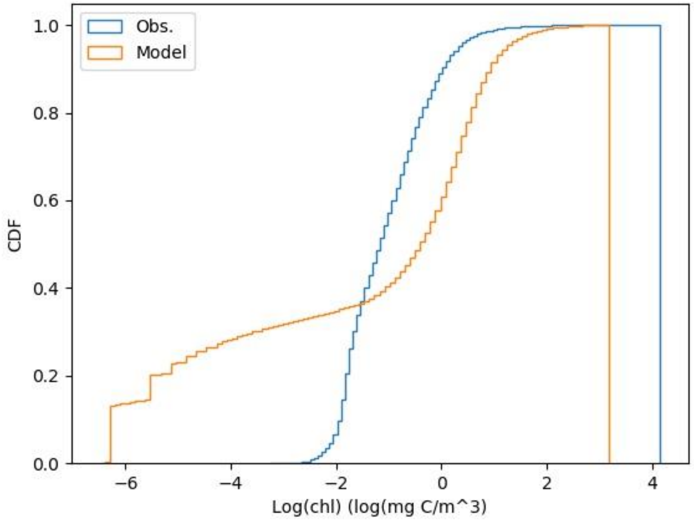
- Any threshold-based method can be sensitive to bias.
- A visual/subjective inspection of the AMM7v8 analyses and L4 product shows that some biases exist which must be considered during the results analysis. The biases appear to be largest near the coast.



Daily mean L4 multi-sensor regridded observations (left) and AMM7v8 output (right) chlorophyll for 10 July 2018. Bottom: Error estimates on the multi-sensor chlorophyll (left) and difference between model and observations (right).

MODE – Method for Object-based Diagnostic Evaluation

- There was a significant bias between the AMM7v8 forecasts and the L4 satellite product
- The model produces many very low concentrations (at the numerical noise level) which are not observed
- Whilst the shape of the upper half of the forecast CDF shows the same rate of increase, by this stage there are too many forecast concentrations compared to those observed, though with a fairly constant offset
- This makes the bias correction using quantile mapping possible
- The observations have a sharper curve, more like that from a normal distribution
- The two curves cross over each other, at a threshold of ~0.5 mg/m³. The observed distribution is very skewed at the upper end; approximately 95% of the values are below concentrations of 1 mg/m³, but have higher concentrations than the forecasts, larger by a factor of about 3.
- This implies that in the tail, forecast concentrations appear to be underdone compared to those observed
- Due to this bias and the influence it has on identifying objects to be compared across forecasts and observations a decision was taken to minimize the effect of the bias where possible.

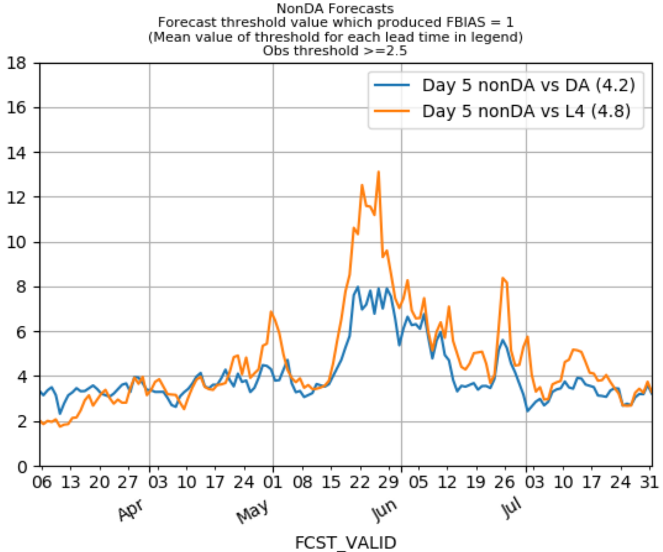


Cumulative distribution of observation and +12h hour forecast log Chlorophyll concentration for the bloom season 2019.

MODE – Method for Object-based Diagnostic Evaluation

- A quantile mapping approach is applied to set the forecast threshold to a value which will vary in time to ensure that the frequency bias of the paired fields is equal to one
- This means that the threshold-exceedance seen in the forecasts occurs at the same proportion as that seen in the observations
- This frequency equivalence, applied across the whole field, behaves as a bias removal tool and subsequent identified objects can then be analysed in the usual way.

- There was very little variation with lead time, so only the day 5 forecast data are shown here.
- These show that the values are generally above the threshold used for the observations (2.5 mg/m^3) and there is a spike in the value of threshold needed to maintain the frequency bias of 1 during mid-May.
- Examining the raw fields at this time shows that the AMM7v8 forecasts have initiated blooms in the North Sea and southwest approaches, where there are no signs of one in the L4 observations, and the general background concentration of chlorophyll appears to be higher.
- The AMM7v11 BIO DA analysis is far less biased than the AMM7v8 forecasts and much closer to the L4 product.



Threshold used with AMM7 forecast fields to ensure a frequency bias of 1 (equivalence in proportion of observed threshold exceedances)

MODE – Method for Object-based Diagnostic Evaluation

COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

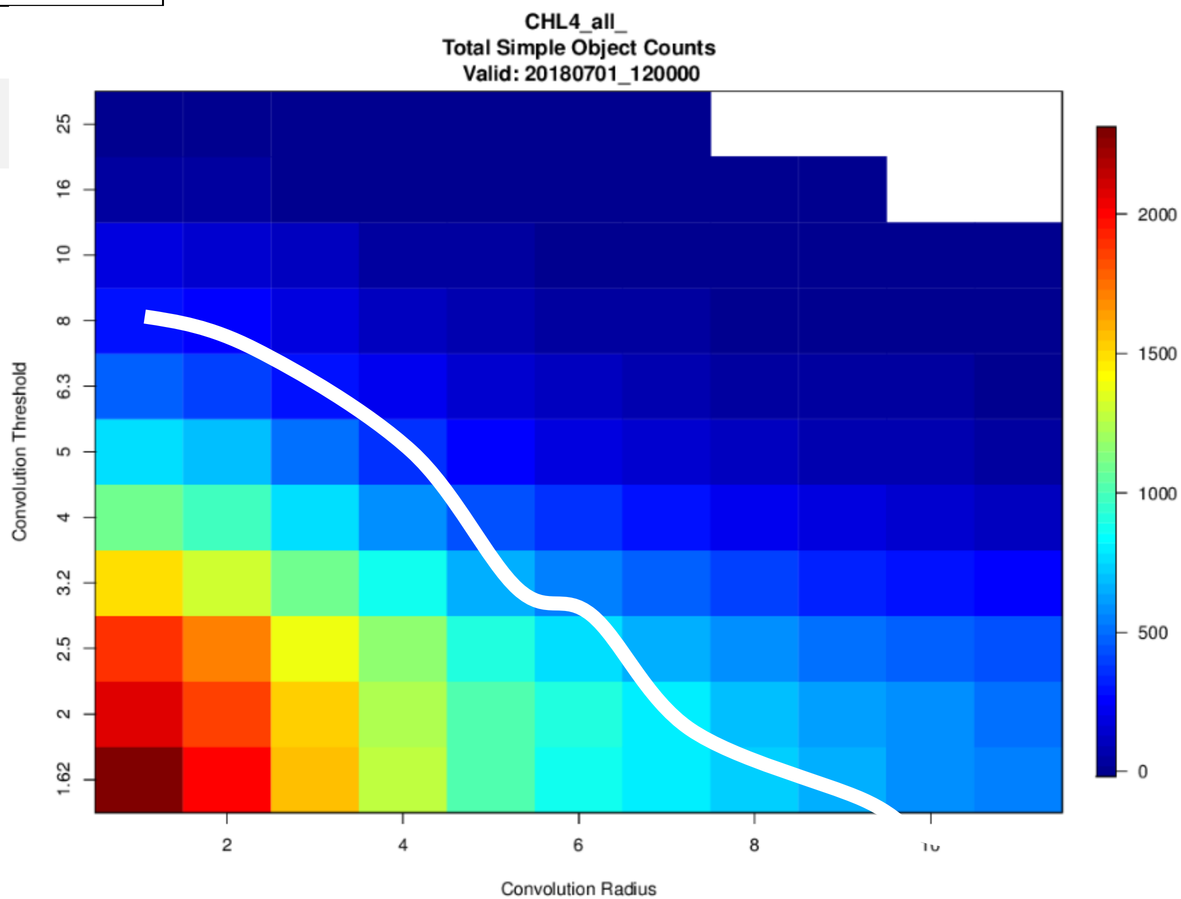
Quilt plots

Explore the relationship of threshold and smoothing radius.

This helps in selecting what the appropriate smoothing radius is for each threshold.

It suggests that for the larger thresholds there are few objects anyway without smoothing so the number of objects may be manageable without smoothing.

For the lowest thresholds you may need to use the highest convolution radius to get the number of objects under control.



Quilt plot for sensitivity analysis: number of objects identified as a function of convolution radius R (number of grid squares) and threshold T (mg/m^3)

MODE – Method for Object-based Diagnostic Evaluation

Impact of biases illustrated

Standard MODE graphical output for 10 July 2018 comparing the AMM7v8 analysis to the L4 product.

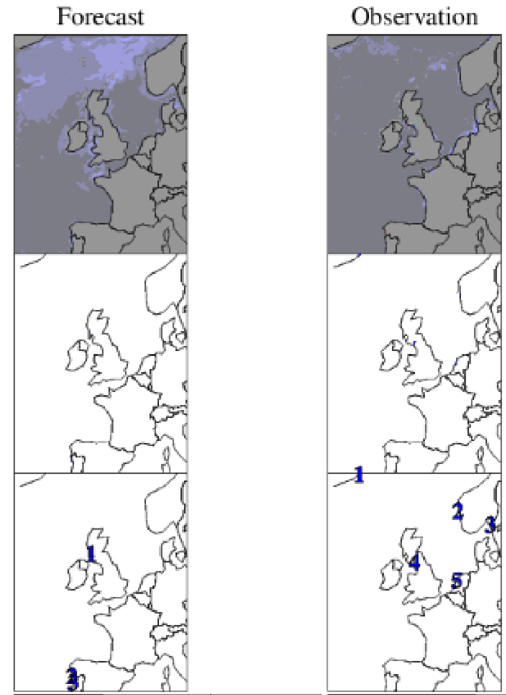
(a) No smoothing for the threshold of 8.0 mg/m^3, showing only 3 objects in the AMM7v8, and a larger number in the L4 product.

(b) Observed and forecast objects are matched based on a range of criteria, to create an “interest” score (higher score = stronger match).

(c) MODE finds one match (forecast object 1 and observed object 4) but with an interest value of 0.5858 it falls below the 0.7 threshold where MODE thinks it is a reasonable match.

(d) The maps in the top panel show the difference in the biases at lower thresholds.

MODE: CHL at 0,0,*,* vs CHL at 0,0,*,*



Fcst	Obs	Interest
1	4	0.5858
2	1	NA
3	1	NA
1	2	NA
2	2	NA
3	2	NA
1	3	NA
2	3	NA
3	3	NA
1	1	NA
2	4	NA
3	4	NA
1	5	NA
2	5	NA
3	5	NA

	Forecast	Observation
Model	AMM7	
Field	CHL	CHL
Level	0,0,*,*	0,0,*,*
Units	mg C/m^3	milligram m-3
Initial	2018 07 10 12:00:00	392495 06 07 13:37:04
Valid	2018 07 10 12:00:00	392452 07 30 00:00:00
Accum	00:00:00	00:00:00

Centroid/Boundary	2.00	4.00
Convex Hull/Angle	0.00	1.00
Aspect/Area	0.00	1.00
Int Area/Curvature	2.00	0.00
Complexity/Intensity	0.00	0.00
Total Interest Thresh	0.70	

	Forecast	Observation
Mask M/G/P	on/off/off	on/off/off
Conv Radius	1	1
Conv Thresh	>=8.0	>=8.0
Obj Filters	1	1
Inten Perc	p50	
Merge Thresh	>=6.5	>=6.5
Merging	thresh	thresh
Matching	match/merge	
Simple/M/U	3/0/3	5/0/5
Area	55	61
Area M/U	0/55	0/61
Cluster	0	0
MMI	0.0000	0.0000
MMI (F+O)	0.0000	

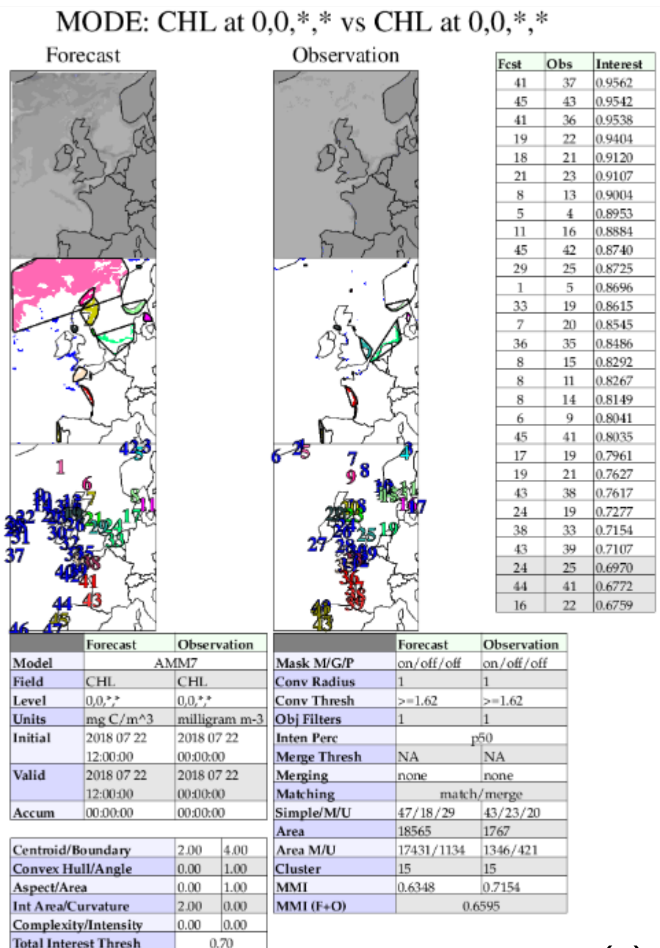
MODE – Method for Object-based Diagnostic Evaluation

Impact of convolution radius

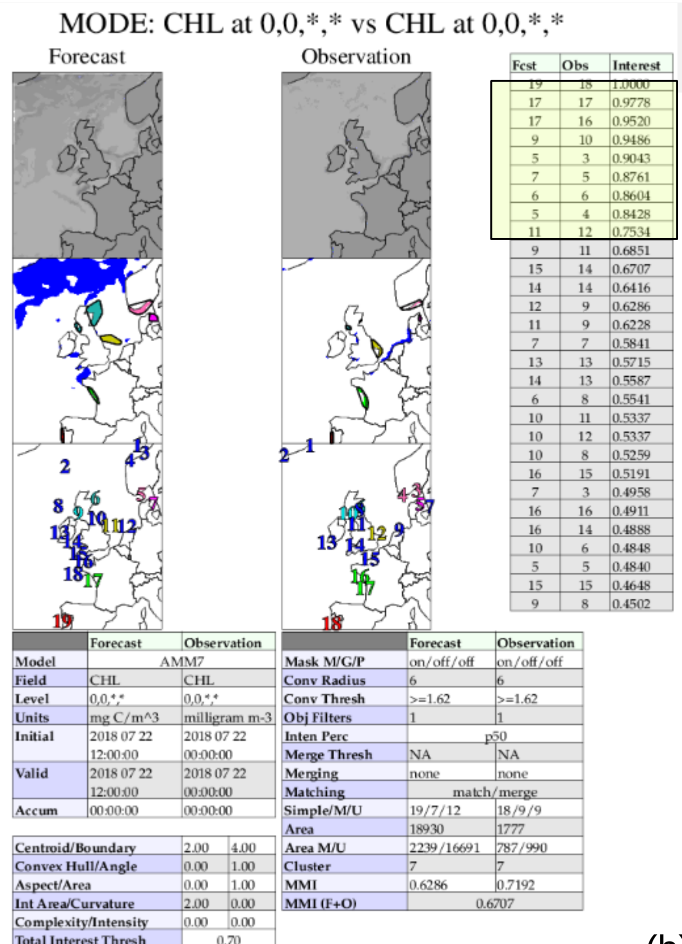
To be able to make a sensible analysis one has to find the balance between threshold and smoothing. Too many objects makes it difficult to analyse. Too much smoothing may mean events become less distinct.

(a) Shows no smoothing for the lowest threshold of 1.62 $mg.m^{-3}$, showing a large number (too many) of objects.

(b) A convolution radius of 6 is applied, reducing the number of objects. This is becoming more manageable.



(a)



(b)

MODE – Method for Object-based Diagnostic Evaluation

Met Office

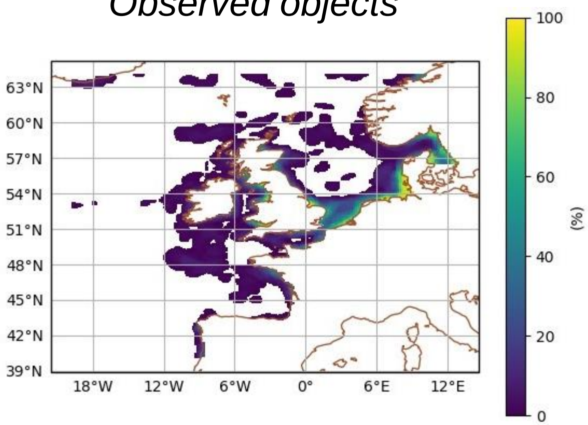
COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

European Union Programme

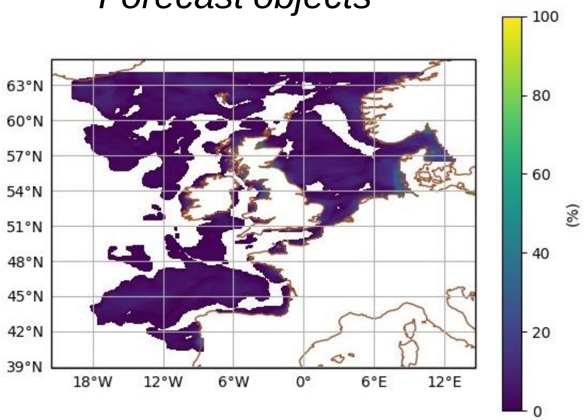
Copernicus Europe's eyes on Earth

MERCATOR OCEAN
OCEAN FORECASTERS
Entrusted entity by mandate of the European Commission

Observed objects



Forecast objects

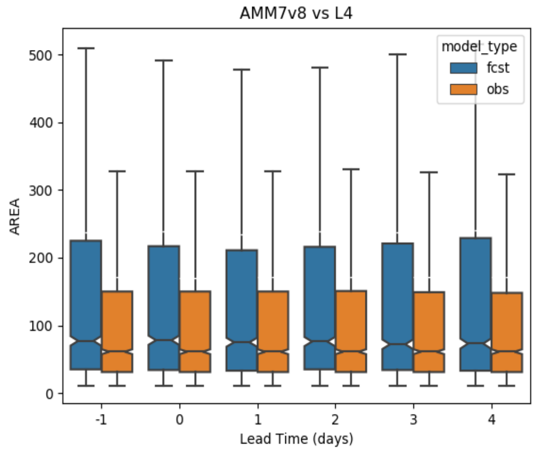


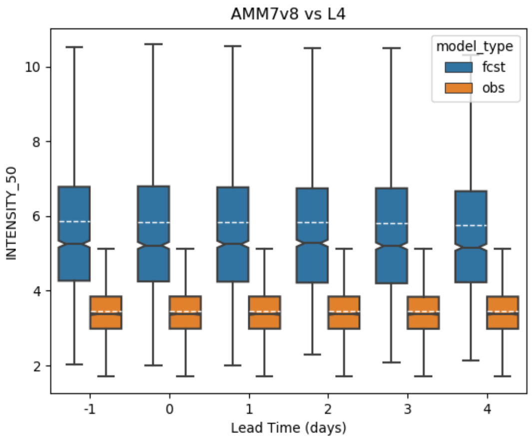
Percentage of objects identified over the season

- Composite spatial coverage of objects identified through the 2019 bloom season for both observed and forecast objects.
- The maps show the proportion of time (in days) that an object occurs at that grid point.
- For the observed composite, the near continuous presence of chlorophyll near the coasts stands out.
- The AMM7v8 forecasts in display a greater spread of locations; there are more points where there is an object identified between 0 and 20% of the time.
- Of note is a patch in the central North Sea where the L4 product has no objects identified but the AMM7v8 forecasts have objects identified for a low proportion of the time.
- There are also areas, for example the south-west approaches, where there appears to be a good level of consistency between the forecast and observed object frequencies.

MODE – Method for Object-based Diagnostic Evaluation

Object attribute distributions



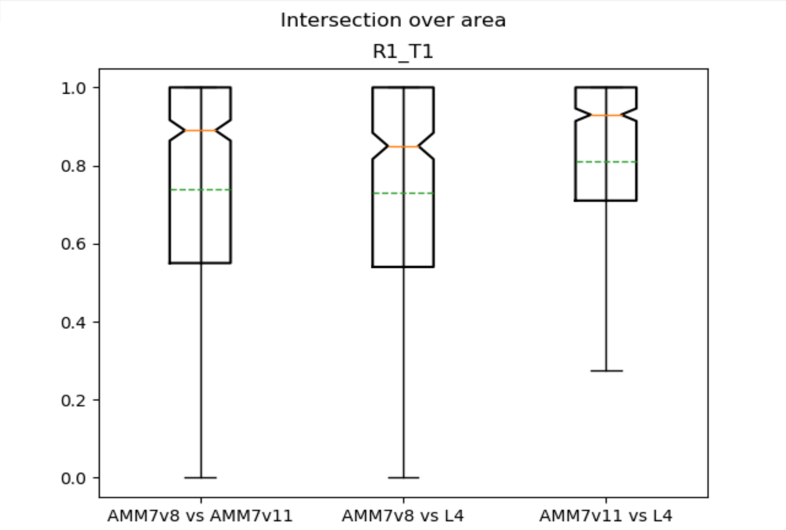


The object areas in grid squares (right) show the distribution of 50th (median) percentile values from all the identified objects in the period.

- There is very little variation with lead time
- AMM7v8 forecasts have a broader distribution in size and are bigger

Paired Object attribute

- The intersection-over-area gives a measure of how much the paired forecast-observed objects overlap in space
- If the objects do not intersect, this metric is 0
- Here many of the matched areas overlap perfectly (it is easy for smaller L4 areas to be completely enveloped by the model analyses).
- However, there is a very long whisker which shows that there are instances where this is 0.
- It is clear that the AMM7v11 BIO DA analysis is closest to the L4 product, with all pairs overlapping in some way. There is quite a difference between the median (notch) and the mean (dashed line).



Paired object attributes for the Day 0 results at the 2.5 mg.m^{-3} threshold and a smoothing radius of 5 grid squares. Ratio of the intersection area over the largest of the forecast or observed object area.

Met Office

COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

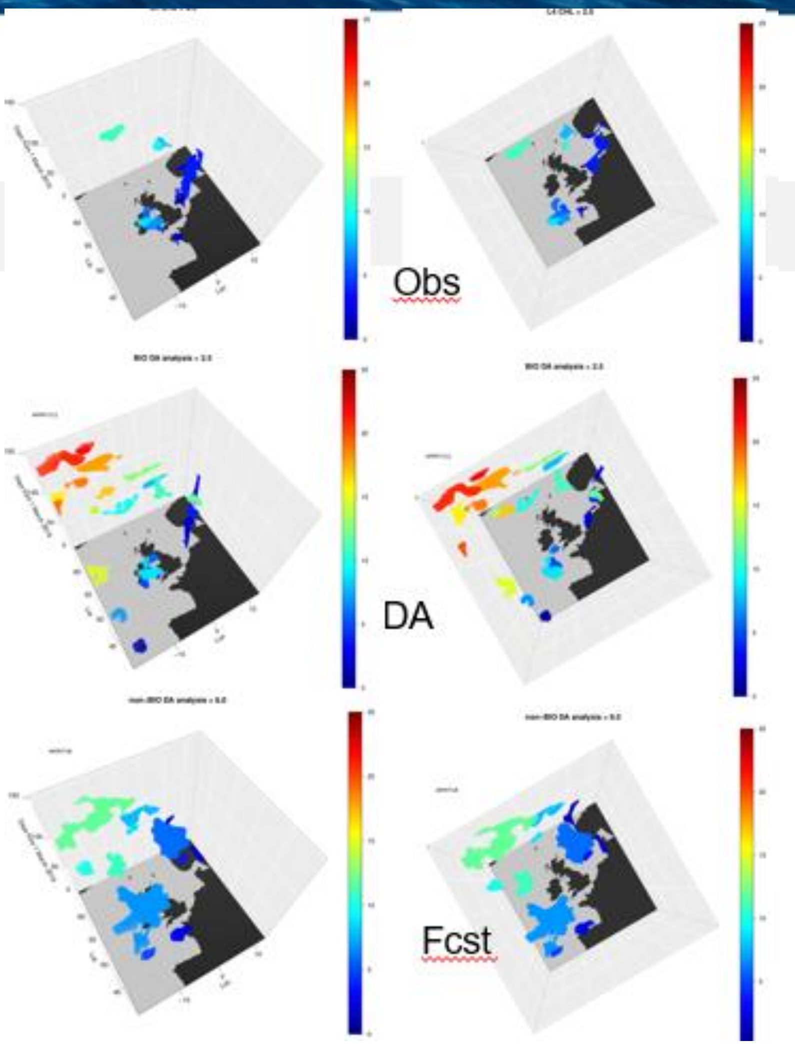
European Union Programme

Copernicus Europe's eyes on Earth

MERCATOR OCEAN
OCEAN FORECASTERS
Entrusted entity by mandate of the European Commission

MODE TD – Method for Object-based Diagnostic Evaluation Time Domain

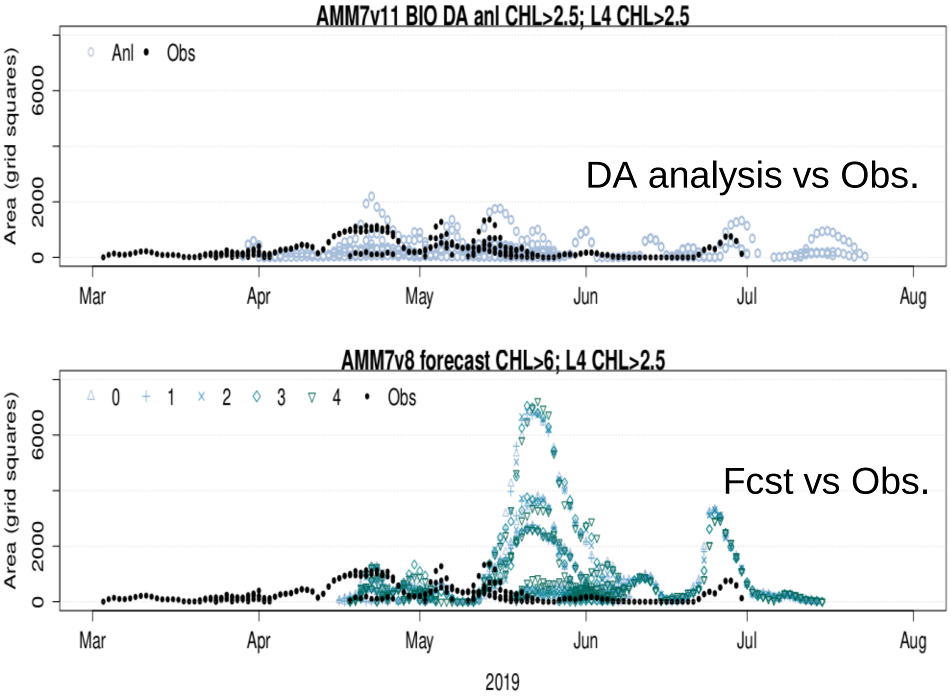
- MODE Time Domain (MTD) provides the ability to connect the objects at a given time to those coming before and after, to gain an understanding of how the object (which represents a feature or event of interest) evolves
- The figures show a summary of the temporal evolution of the chlorophyll blooms during the 2019 season, showing three different analyses: the satellite based L4 product (top), the AMM7v11 BIO DA analysis (middle) which assimilates chlorophyll, and the AMM7v8 analysis without any biogeochemical assimilation (bottom).
- Both a side-on and top-down (or bird's eye) view are provided
- The assimilation of chlorophyll is sufficient to ensure that the analysis is relatively unbiased (in terms of concentration) compared to the L4 product
- The AMM7v11 BIO DA analysis is far more "active" providing many more blooms/episodes
- The colours show how the bloom migrates north and west as the season progresses (based on the progression of object identifiers and colours)
- The non-DA analysis on the other hand produces far fewer blooms/episodes but those that are produced are far too large, suggesting that in addition to the concentration bias there is also a spatial extent bias.



Temporal evolution of identified chlorophyll space-time objects. Colours represent the object numbers, which increase with time. Thresholds in mg.m^{-3}.

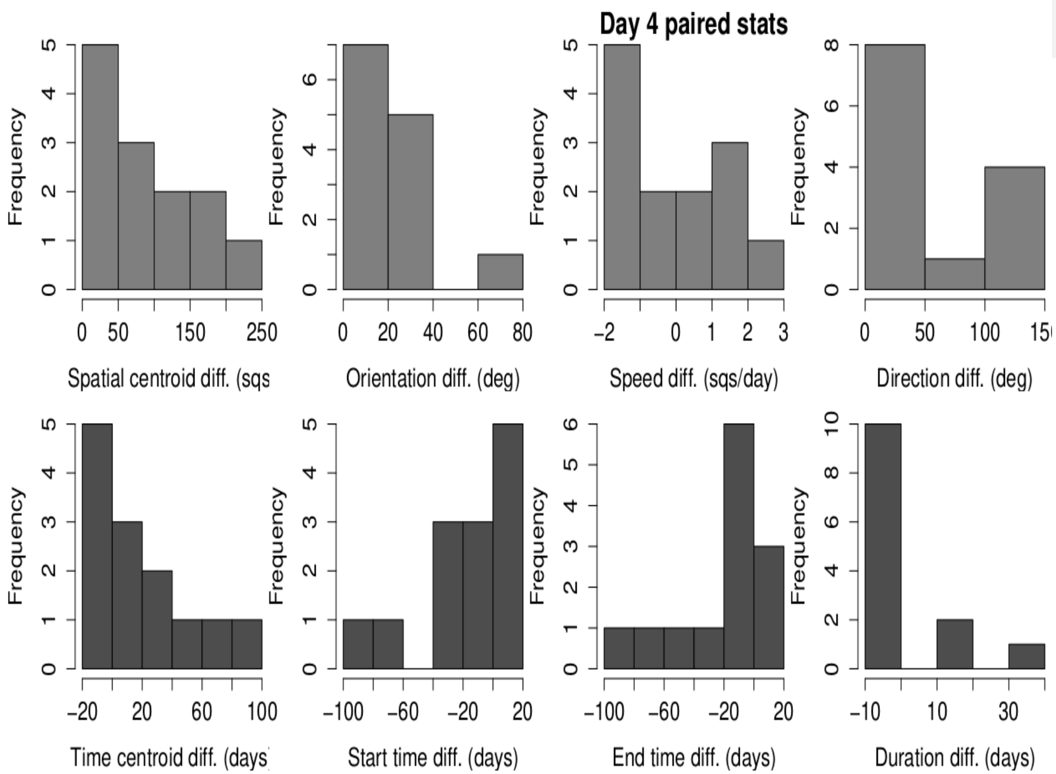
MODE TD – Method for Object-based Diagnostic Evaluation Time Domain

- The first identifiable chlorophyll bloom in the AMM7v11 BIO DA analysis was identified late in March 2019, around 10 days later than in the L4 product
- The AMM7v8 forecasts only picked up the first event of the season in April 2019, almost a month after the first identified chlorophyll object was identified in the L4 product
- Subsequent peaks are better aligned but the mid-May peak is completely over-estimated by AMM7v8
- The overlap between the different forecast lead times indicates that there is very little difference in the forecasts as a function of lead time
- Equally it would appear as though the model does not capture the end of the bloom season, stopping too soon by at least a week
- Forecasts are currently driven from the AMM7v8 non-BIO DA analysis and despite the concentration bias removal still show a large spatial extent bias for some episodes.



Time series of all identified MTD object areas.

MODE TD – Method for Object-based Diagnostic Evaluation Time Domain



- The spatial centroid (centre of mass) differences can be extensive, but the majority are within 0-50 grid squares apart (i.e. up to ~350 km).
- Generally, the orientation of objects is within 40 degrees.
- The vast majority of paired objects have time centroid differences +/- 20 days of the observed, with a preference for the forecasts being later (difference being defined as forecast time minus observed time).
- This is better illustrated by the distribution of start and end times.
- Forecast blooms are generally too short.

MODE TD – Method for Object-based Diagnostic Evaluation Time Domain

Met Office

COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

European Union Programme

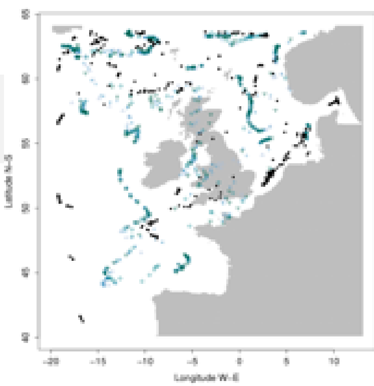
Copernicus Europe's eyes on Earth

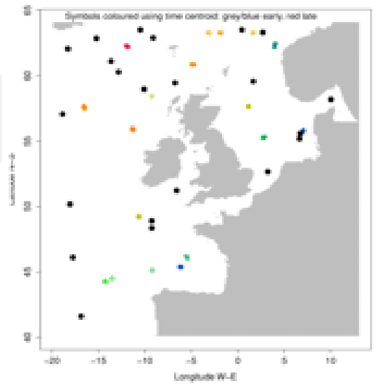
MERCATOR OCEAN
OCEAN FORECASTERS
Entrusted entity by mandate of the European Commission

- Centroids in space and time can be examined spatially (in terms of their location).
- The figures here give a visual comparison of the impact of using the L4 satellite product or the AMM7v11 BIO DA analysis as the verifying analysis.
- The different symbols indicate the different forecast lead times.
- The black dots indicate the observed centroid.
- There are some clear differences between the AMM7v11 BIO DA analysis and the L4 product, especially in the southern North Sea and also in the north and west.
- There are some areas/times where black dots and forecast centroids can be found in in reasonable spatial proximity (though this may not indicate temporal proximity).

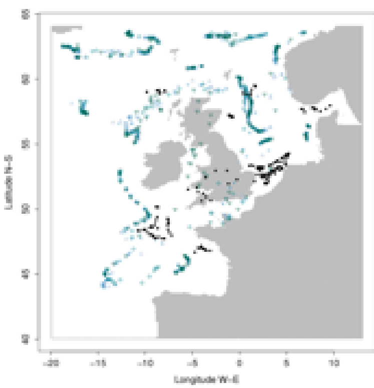
- The time centroid is derived from these spatial centroids.
- Again, the black dots represent the observed time centroid whereas the colours indicate the forecast centroids.
- The colours represent the relative position within the season, with blue colours early in the season, and the reds and pinks towards the end of the season.
- The forecast time centroids for the different lead times are essentially on top of each other showing there is no change with lead time.
- The impact of using the AMM7v11 BIO DA analysis and L4 product is evident in the observed centroids, with the AMM7v11 BIO DA analysis producing many more objects in deeper waters to the north and west.

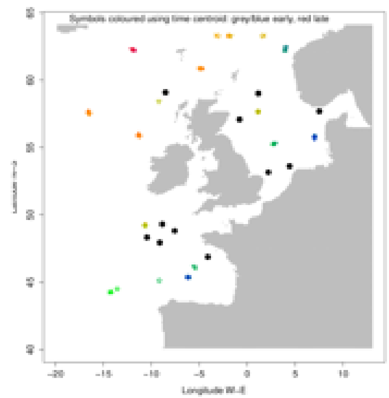
AMM7v11 BIO DA analysis





L4 satellite observations





Space centroids *Time centroids*

Composites of the entire 2019 season

MODE TD – Method for Object-based Diagnostic Evaluation Time Domain

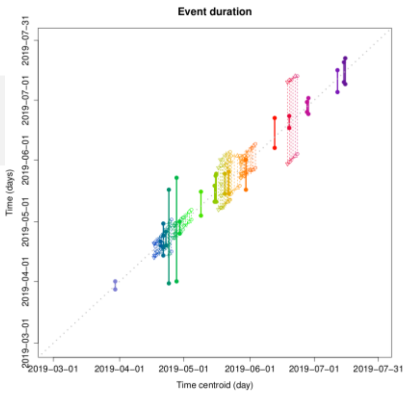
COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

European Union Programme

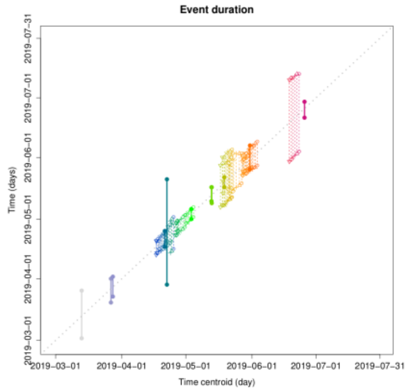
- The progression of the bloom is shown on the right, where the x-axis represents elapsed time
- Vertical lines on given dates indicate the location of a time centroid
- The initial identification and end of a given object/event are indicated by the start and end of the vertical lines
- Solid lines represent the observed events whereas dashed lines are the forecast events

- From this the difference in the onset of the 2019 season is clear
- The AMM7v11 BIO DA analysis manages to detect something fairly early on but does not detect the initial events in the L4 product
- There is evidence of the influence of the model in the AMM7v11 BIO DA analysis at both the start and end of the season, where the BIO DA analysis follows the model in continuing to find events beyond where the L4 product detected anything
- AMM7v8 forecast blooms/episodes are too short in duration, though there may be a bias to be too long-lived later in the season.
- Overall, most groups of forecast objects have some association with an observed object around about the same time (though this does not mean they are close in space)

AMM7v11 BIO DA analysis



L4 satellite observations



Duration of time objects

COPERNICUS
MARINE ENVIRONMENT MONITORING SERVICE

Conclusions

- *Bias* – there is a significant concentration bias in the forecast compared to the observations, which must be mitigated against before using MODE or MTD; the bias may be improved using the DA analysis to drive the forecast.
- *Timing issues* – MTD analysis has shown that the initial onset of the bloom is almost a month late (25 days) in AMM7v8. Subsequent events are handled somewhat better.
- *Location* – beyond the timing issues the model does generally produce chlorophyll objects (blooms) in the right areas, but not necessarily at the right time.
- *Evolution with lead time* – there is very little change in results between the analysis and the forecasts at all lead times. Predicting the onset of the bloom seems problematic.
- *Benefit of AMM7v11 BIO DA* – it would be good to see whether forecasts initialised using the BIO DA analysis will improve the timing errors and the bias. The BIO DA analysis is an improvement compared to the one without (which the forecasts in this study were initiated with).
- *Size of objects* – as for MODE, the objects are generally too large. This spatial extent bias is in addition to the previously mentioned concentration bias.
- *Number of objects* – AMM7v8 produces fewer objects compared to observed and these are too large. Many of the coastal objects identified in the L4 product cannot be resolved by the model due to the coarseness of the coastline in the 7 km model. This situation would improve should the model resolution increase from 7 km to 1.5 km.

Questions and discussion

Further detail at:

- Ocean Science paper, in preparation, May 2020 -

Mittermaier M., North R., Pequignet C., Maksymczuk J., Ford, D.
Using object-based spatial verification methods for the evaluation of forecasts of the 2019 chlorophyll bloom season over the European North-west Shelf

This work has been carried out as part of the Copernicus Marine Environment Monitoring Service (CMEMS) HiVE project. CMEMS is implemented by Mercator Ocean International in the framework of a delegation agreement with the European Union.

Verification was performed using the Model Evaluation Tools (MET) verification package, that was developed by the National Center for Atmospheric Research (NCAR), and which can be configured to generate MODE/MODE-TD results. MET is free to download from GitHub at https://github.com/NCAR/MET.