

ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA (GBYP - 2010)

Data Recovery Plan – Elaboration of 2010 Data from sst and the Aerial Survey on Spawning Aggregations

Final Report 03 December 2010

Ana Cañadas, Philip Hammond & José Antonio Vázquez
Anilam Research and Conservation Ltd
Cándamo 116, La Berzosa, 28240 Hoyo de Manzanares, Madrid, Spain

Background

The comprehensive ICCAT Atlantic Wide Research Programme on Bluefin Tuna (GBYP) aims to improve basic data collection, understanding of key biological and ecological processes, and assessment models and management. An important element of this programme is to carry out aerial line transect surveys of the spawning population in the Mediterranean when and where schools can traditionally be sighted close to the surface to support development of fishery-independent indices.

Under the GBYP Data Recovery Framework it is desired to include an evaluation of the importance of environmental covariates, such as sea surface temperature data, in the aerial survey design. Density surface modelling is an approach that uses physical and environmental data to help explain variation in distribution and density and predict areas that are important for the focal species. When combined with line transect sampling (called the model-based method; Hedley et al. 1999), it is an alternative technique to conventional line transect sampling (design-based method; Hiby and Hammond 1989; Buckland et al. 2001).

Objectives

To fit spatial models, using methods (density surface modelling) described in Cañadas & Hammond (2006; 2008), to explore the relationship between bluefin tuna density and environmental covariates.

To provide maps of the predicted densities of bluefin tuna in the survey blocks.

Data

Data availability

The 2010 aerial survey data were already available to the authors from a previous contract with Hammond, Cañadas & Vázquez (2010). Sea surface temperature (sst) data were made available by ICCAT in electronic format at a resolution of $0.25^{\circ} \times 0.25^{\circ}$ for May, June and July 2010. Figure 1 shows the survey areas and the effort transects and Figure 2 shows the bluefin tuna sightings. Figures 10 to 20 (in Annex) show the mean sea surface temperature for the months of June and July and for each week in those months.

Data processing

Environmental data

A grid of cells was built for the whole Mediterranean with the same resolution as the sst data provided ($0.25^{\circ} \times 0.25^{\circ}$). These cells were populated with the sst as well as other potential covariates available to the proposers (see Table 1 for a complete list).

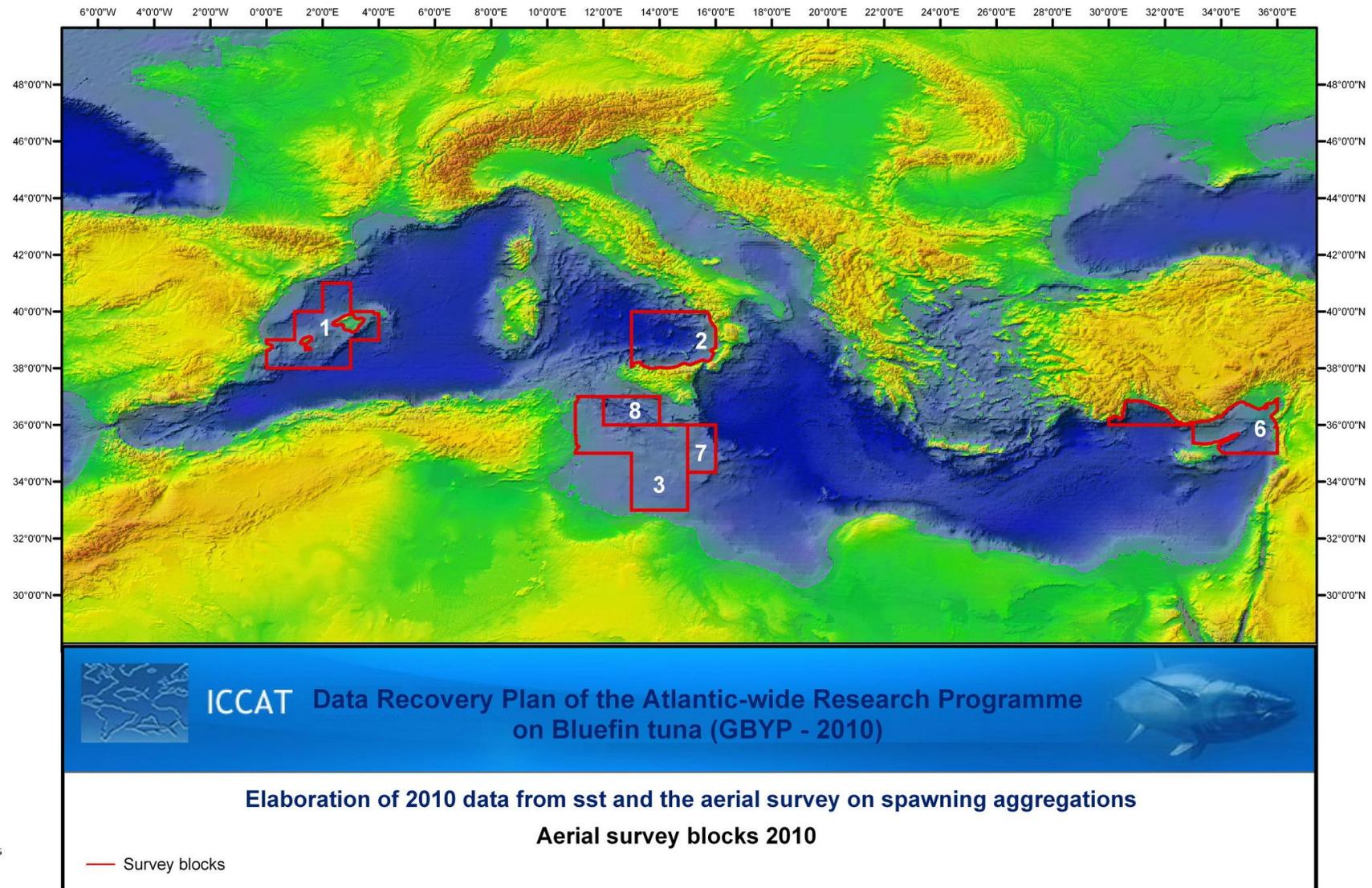


Figure 1a. Aerial survey blocks considered in the analysis

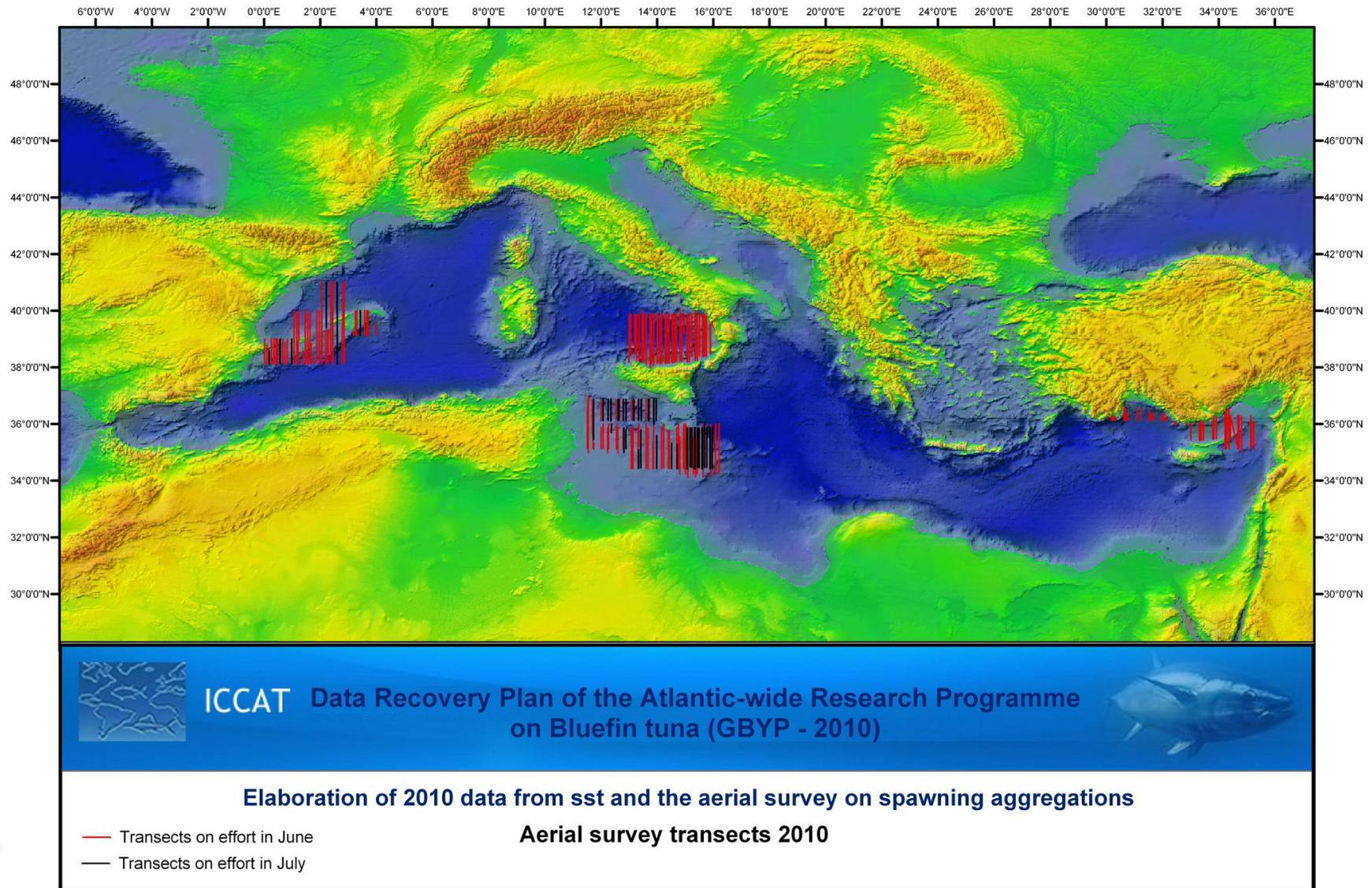


Figure 1b. Aerial survey transects. Transects in June are shown in red and transects in July are shown in black.

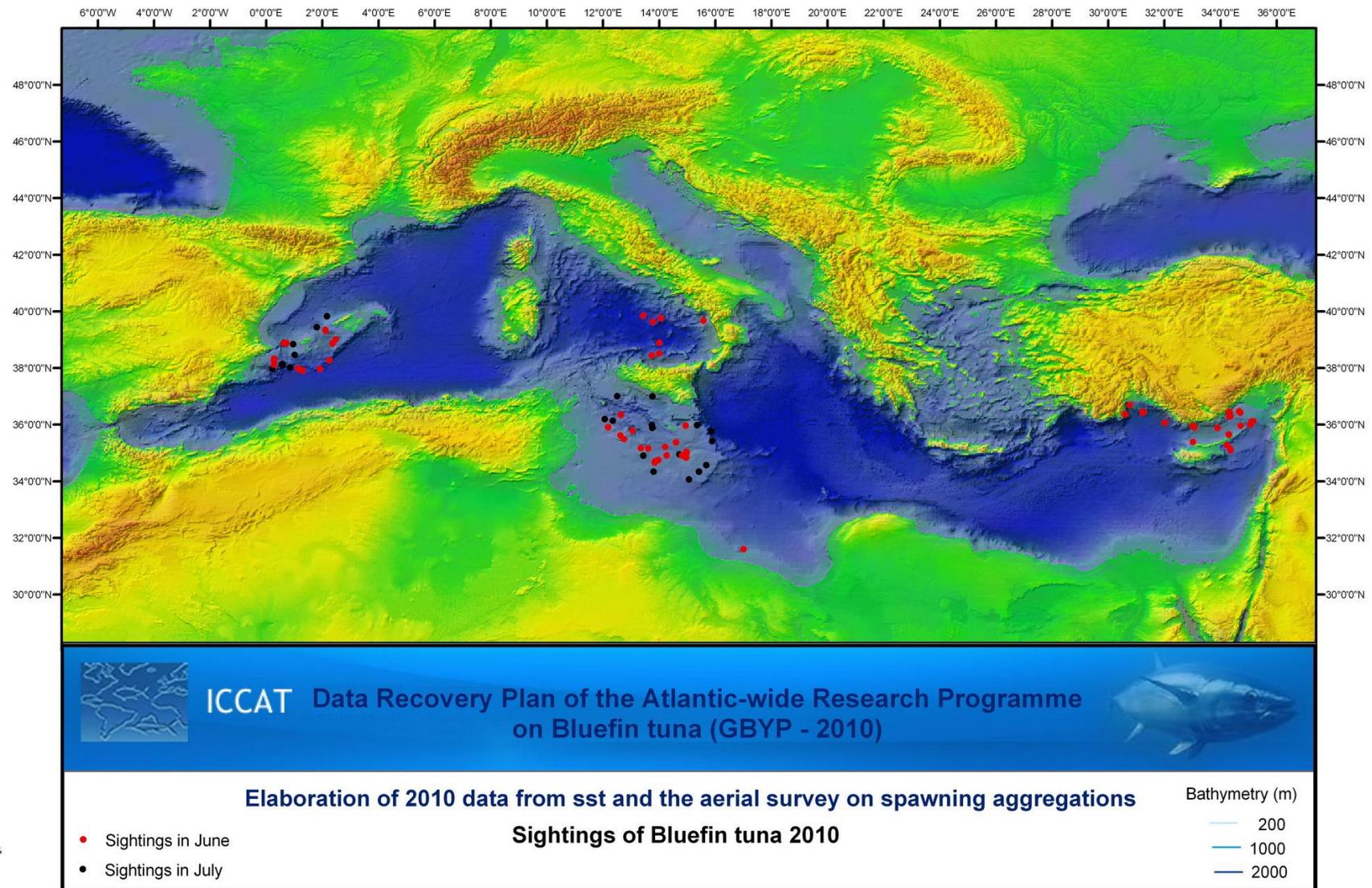


Figure 2. Sightings of bluefin tuna. Sightings in June are shown in red and sightings in July are shown in black.

Table 1. List of covariates tested in the spatial models for significance in their contribution to explain spatial distribution of sightings of bluefin tuna

Covariate	Description	Origin
Latitude	Latitude in decimal degrees	
Longitude	Longitude in decimal degrees	
Depth_mean	Mean depth within the grid cells	ETOPO 2v2 (http://www.ngdc.noaa.gov/mgg/fliers/01m_gg04.html)
Depth_sd	Standard deviation of depth within the grid cells: measure of complexity of sea floor	Derived from ETOPO 2v2 data
Depth_CV	Coefficient of variation of depth within the grid cells: measure of complexity of sea floor	Derived from ETOPO 2v2 data
Ci	Contour Index: combined measure of depth and slope	$(\text{max_depth} - \text{min_depth}) * 100 / \text{max_depth}$
Dist200	Distance from the centre of the grid cell to the nearest point on the 200m depth contour	GIS
Dist1000	Distance from the centre of the grid cell to the nearest point on the 1000m depth contour	GIS
Dist2000	Distance from the centre of the grid cell to the nearest point on the 2000m depth contour	GIS
Aspect	Orientation of the sea floor relative to North	GIS
Sst_day	Sst in the grid cell on the day the segment of effort occurred	Derived from sst provided by ICCAT
Sst_week	Mean sst in the grid cell in the week the segment of effort occurred	Derived from sst provided by ICCAT
Sst_month	Mean sst in the grid cell in the month the segment of effort occurred	Derived from sst provided by ICCAT
Sst_mean2	Running average of sst on the day and the day before the segment of effort occurred	Derived from sst provided by ICCAT
Sst_mean3	Running average of sst on the day and the two days before the segment of effort occurred	Derived from sst provided by ICCAT
Sst_mean5	Running average of sst on the day and the four days before the segment of effort occurred	Derived from sst provided by ICCAT

Effort segments and sightings

Aerial survey effort data were organized into segments of similar length and searching conditions, which comprised the sampling units for spatial modelling. This process was done in a way so that each segment fit exactly inside one grid cell (not sharing grid cells). This process yielded a total of 1351 segments (i.e. sampling units) of average length 24 nmi.

All segments were associated with the covariates in Table 1, according to the grid cell in which they fell.

The estimated numbers of groups (obtained through the Horvitz-Thompson estimator, see equation 1) were associated to their corresponding segment of effort (assigning 0 to the remaining segments), and this value was used as response variable for the models. Of the 1351 segments of effort, 61 (4.5%) had associated bluefin tuna sightings, for a total of 110 sightings.

Data analysis

Initial exploration of data

The data were explored initially in two ways: (a) frequency distributions of the data for some covariates were produced for all segments and also only for segments containing sightings, and (b) the two-sample Kolmogorov-Smirnov non-parametric test was applied to all covariates comparing all segments with only those segments containing sightings, assuming the null hypothesis that the samples are drawn from the same distribution.

Spatial modelling

Generalised Additive Models (GAMs) were used to model bluefin tuna density as a function of the available covariates.

The response variable used to formulate a spatial model of abundance of groups was the estimated number of groups (\hat{N}) in each segment, rather than the actual counts (Hedley et al. 1999). They were estimated through the Horvitz-Thompson estimator (Horvitz & Thompson 1952), where the probability of detection was obtained from the detection function fitted to the data:

$$\hat{N}_i = \sum_{j=1}^{n_i} \frac{1}{\hat{p}_{ij}} \quad (1)$$

where n_i is the number of detected groups in the i^{th} segment, and \hat{p}_{ij} is the estimated probability of the j^{th} detected group in segment i , obtained from the detection function.

The abundance of groups was modeled using a Generalized Additive Model (GAM) with a logarithmic link function. A Poisson error distribution was not considered appropriate for the response variable due to over-dispersion. Therefore, a quasi-poisson family was used, with variance proportional to the mean. The general structure of the model was:

$$\hat{N}_i = \exp \left[\ln(a_i) + \theta_0 + \sum_k f_k(z_{ik}) \right] \quad (2)$$

where the offset a_i is the searched area for the i^{th} segment (calculated as the length of the segment multiplied by two times the truncation distance), θ_0 is the intercept, f_k are smoothed functions of the explanatory covariates, and z_{ik} is the value of the k^{th} explanatory covariate in the i^{th} segment.

Models were fitted using package ‘mgcv’ version 1.6-2 for R (Wood 2001). Automated model selection by a stepwise procedure was not yet implemented in the version of R used (2.11.1) (<http://cran.r-project.org>). Therefore, manual selection of the models was done using three indicators: (a) the GCV

(General Cross Validation score) which is in practice an approximation to AIC (Wood 2000) and in which smoothing parameters (in terms of number of knots and degrees of freedom) are chosen by the software to minimize the GCV score for the model, unless they are directly specified; (b) the percentage of deviance explained; and (c) the probability that each variable is included in the model by chance. The decision to drop a term from the model was adopted following the criteria proposed by Wood (2001). In all models, a visual inspection of the residuals was also made, especially to look for trends.

The best model was used to predict bluefin tuna distribution, in a stratified fashion, within all the survey blocks. As an exploratory "experiment", a prediction was also produced for the whole Mediterranean Sea. The final model was predicted for each week from 29 May to 1 August to show potential variability in the predicted densities as the sst changes with time in the Mediterranean.

Attempts were made also to model the weight of the schools as a function of the environmental covariates available, but no relationship could be found. Therefore, the estimated mean weight of bluefin tuna per block obtained from the distance sampling analysis was used.

To obtain the final prediction of bluefin tuna weight in the survey blocks, the predicted abundance of groups in each block was multiplied by the mean weight of the block. The same was done for the whole Mediterranean using the mean weight of all sightings across blocks.

The predictions produced by the spatial models were saved in the same grid of cells, and plotted in a G.I.S.

Results

Initial exploration of data

Visual inspection of the histograms showed the difference in frequency distribution between all segments and segments containing sightings for the sst covariates, and the lack of difference for depth-related covariates. As examples, Figure 3 shows histograms for covariates sst_day, depth_mean and dist2000.

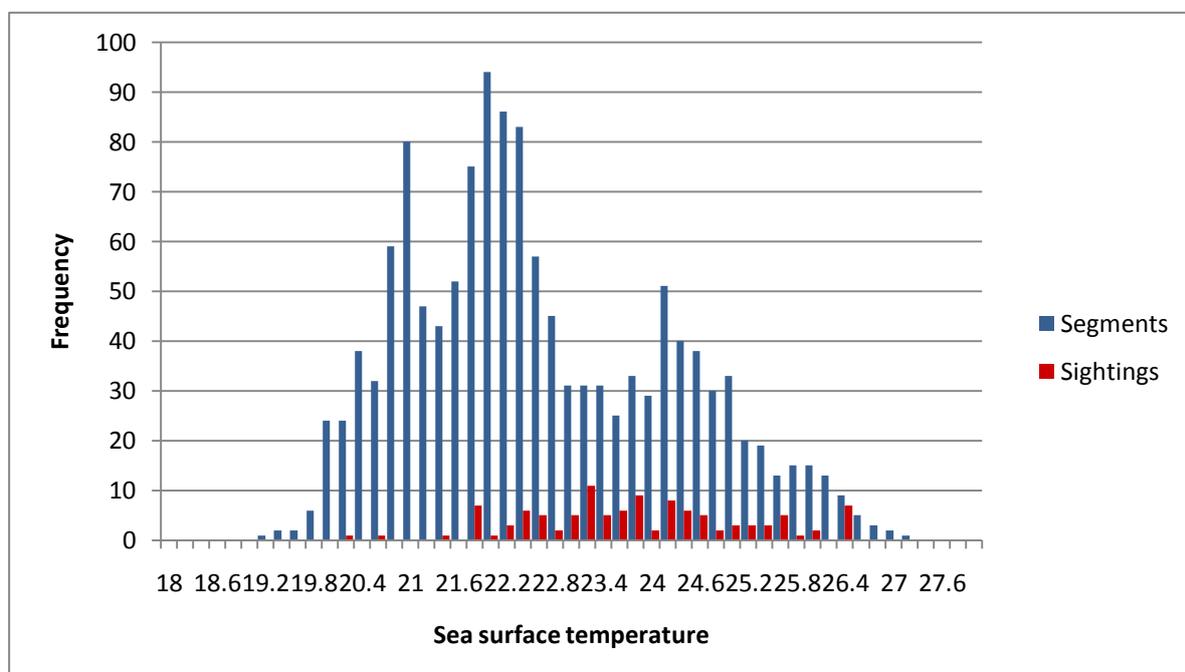


Figure 3a. Frequency distribution of daily sea surface temperature (sst_day) for all segments (in blue) and segments containing sightings (red).

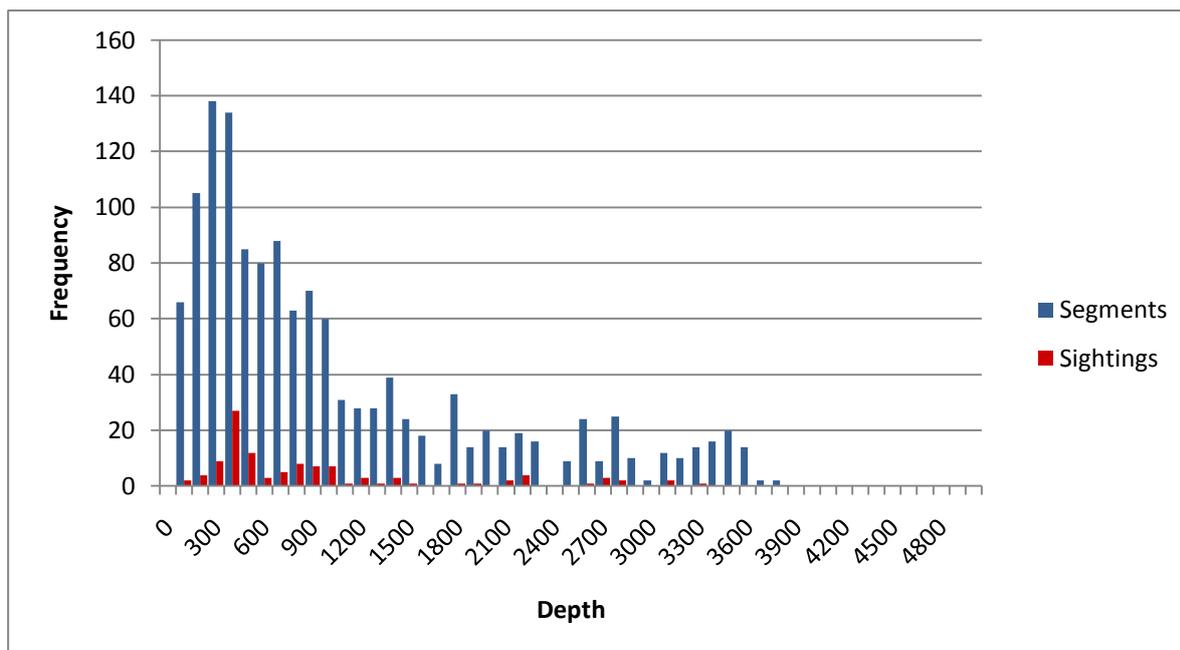


Figure 3b. Frequency distribution of mean depth (depth_mean) for all segments (in blue) and segments containing sightings (in red).

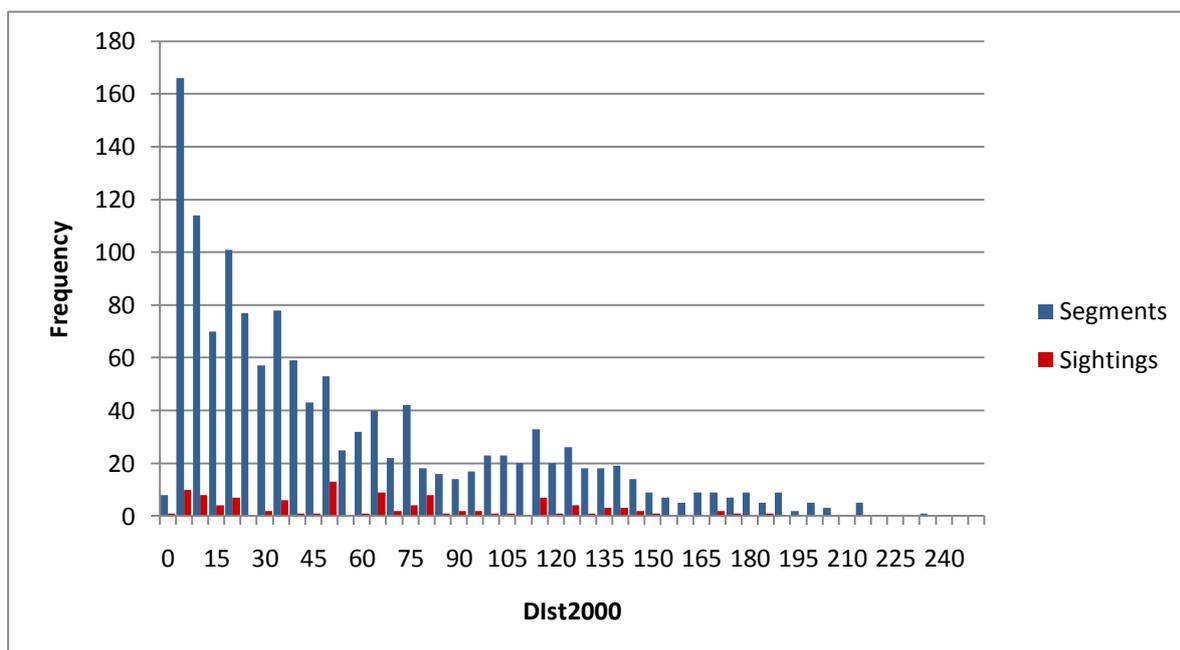


Figure 3c. Frequency distribution of distances to the 2000m depth contour (Dist2000) for all segments (in blue) and segments containing sightings (in red).

Table 2 shows the results of the Kolmogorov-Smirnov tests for all covariates. All the sst covariates were highly significant, but only two of the depth related covariates were significant (distance from the 1000m and the 2000m depth contours) and much less so than the sst covariates.

Table 2. Results of Kolmogorov-Smirnov tests to compare the distribution of all segments with that of only those segments with sightings. In bold, significant differences (rejection of null hypothesis) at $\alpha=0.01$.

Covariate	Parameter D	P value
Depth_mean	0.1081	0.1856
Depth_sd	0.1164	0.1268
Depth_CV	0.1236	0.0896
Ci	0.1055	0.2074
Dist200	0.1169	0.1242
Dist1000	0.1792	0.0029
Dist2000	0.2085	0.0003
Aspect	0.1508	0.0195
Sst_day	0.4023	<0.0001
Sst_week	0.3236	<0.0001
Sst_month	0.2469	<0.0001
Sst_mean3	0.4005	<0.0001
Sst_mean5	0.3966	<0.0001

Spatial modelling

Figures 4 to 8 show the smooth functions for the individual sea surface temperature related covariates. The daily, 2-day running average and weekly means show a very similar pattern, while the 3-day and 5-day running averages show a different pattern. Nevertheless, in all cases the trend is for higher densities towards higher temperatures, although there is a high response also around 23° for the mean for 3 and 5 day running averages. All these covariates were highly significant, but the one that better fits the data is that of sst_day (deviance explained= 17.5%, GCV=0.992), closely followed by sst_mean5 (DE=16.7%, GCV=1.000), sst_mean3 (DE=16.0%; GCV=1.008), sst_mean2 (DE=15.9%, GCV=1.011), and sst_week (DE=14.3%, GCV=1.029).

The best model included two covariates: sst_day (sea surface temperature during the day the segments of effort were carried out) and depth_mean. This model explained 21.4% of the deviance and both covariates were highly significant. Figure 9 shows the smooth functions for these two covariates fitted in the same model.

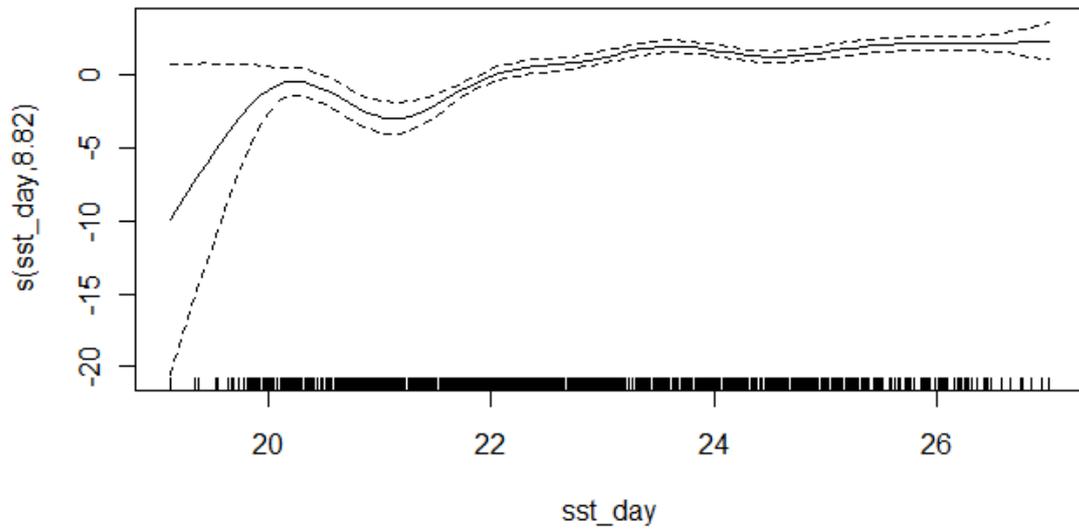


Figure 4. Smooth function for the daily sea surface temperature (*sst_day*). The ticks on the x axis show the distribution of the samples used in the model (the effort segments) for each covariate. The dashed lines represent ± 1 se. When the line of the smooth function goes above 0 in the y axis (showing a relative index of density), it means that the covariate has a positive effect on the response variable (estimated number of groups), and *vice versa*.

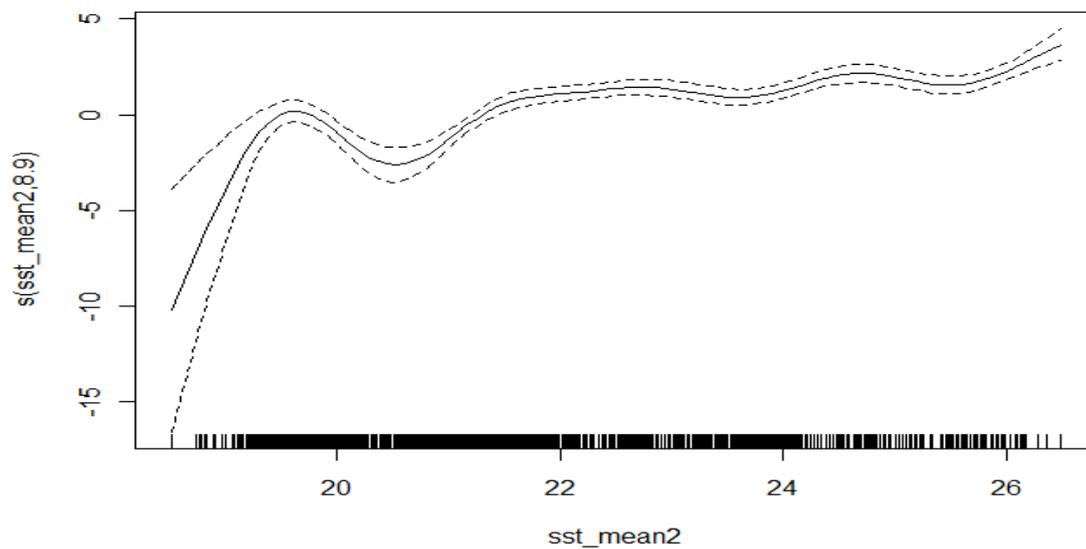


Figure 5. Smooth function for the 2-days running average of surface temperature (*sst_mean2*).

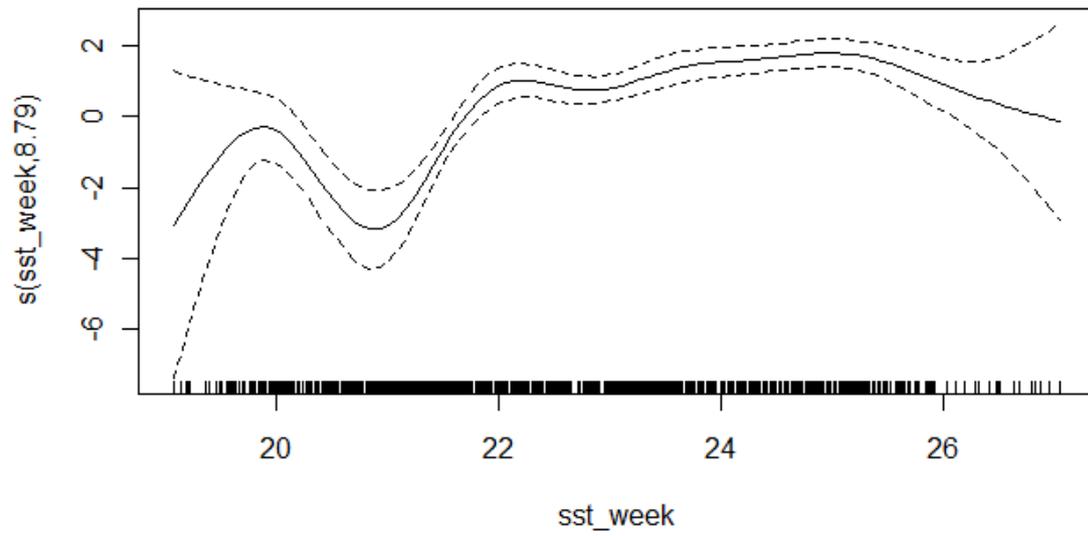


Figure 6. Smooth function for the weekly average of surface temperature (sst_week).

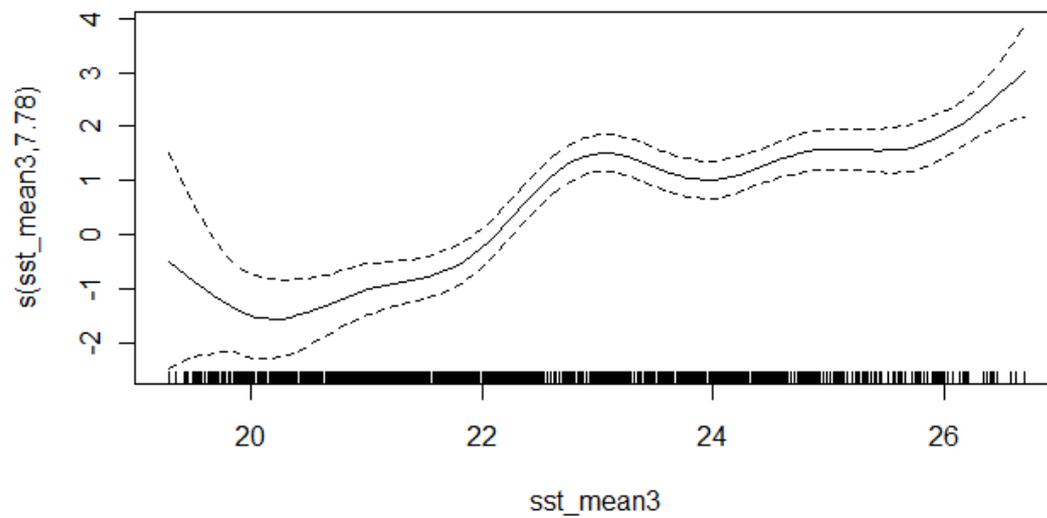


Figure 7. Smooth function for the 3-days running average of surface temperature (sst_mean3).

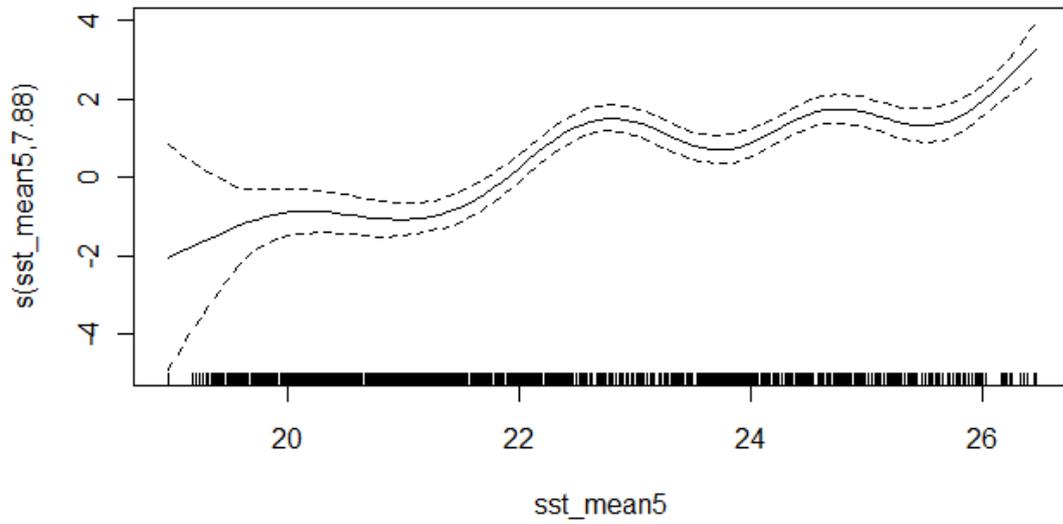


Figure 8. Smooth function for the 5-days running average of surface temperature (sst_mean5).

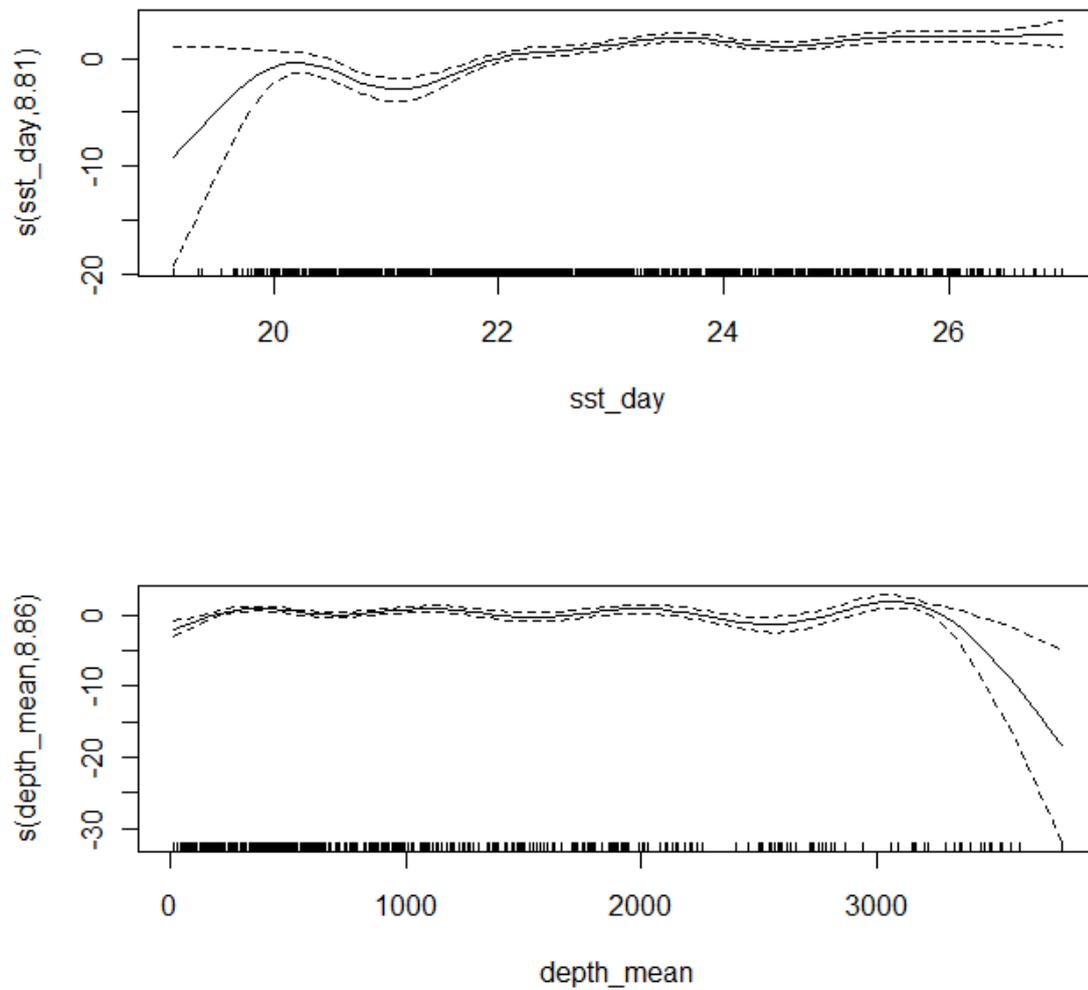


Figure 9. Smooth functions for the daily sea surface temperature (sst_day) and the depth of the sea floor (depth_mean) fitted in the same model.

Figures 21 to 31 show the predictions of density of weight from the fitted model for the 11 weeks comprising this study for the six blocks of the study area. Note that the density scale differs among these figures. Figures 32 to 42 (in Annex) show the exploratory predictions from the fitted model for these same 11 weeks for the whole Mediterranean Sea.

Table 3 show the estimated total weight of bluefin tuna in each of the survey blocks predicted from the models: mean weight for the whole period, for the months of June and July, and for each of the 9 weeks of the study period. For comparison, the estimated density per block from the conventional distance sampling analysis provided in the previous contract (Hammond, Cañadas & Vázquez 2010) is also given.

Table 3. Predicted total weight of bluefin tuna in each survey block from spatial modelling and from conventional distance sampling (CDS).

Block	June	July	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Mean	CDS
1	227	14,781	639	296	254	721	13,445	12,331	18,061	18,910	19,165	9,314	1,244
2	946	17,363	1,068	2,846	8,939	530	12,134	13,938	20,481	30,753	17,470	12,018	1,540
3	240	5,243	187	194	752	806	3,657	4,680	6,898	8,466	5,298	3,438	2,336
6	5,841	11,405	4,965	7,020	6,665	5,280	6,671	9,692	13,356	18,302	18,408	10,040	10,434
7	18	923	26	15	251	169	808	591	1,120	1,722	830	615	131
8	1,000	12,088	0	969	401	514	2,838	9,498	11,556	13,133	10,733	5,516	2,474
Total	8,271	61,804	6,885	11,341	17,261	8,019	39,553	50,730	71,471	91,287	71,905	40,939	18,158

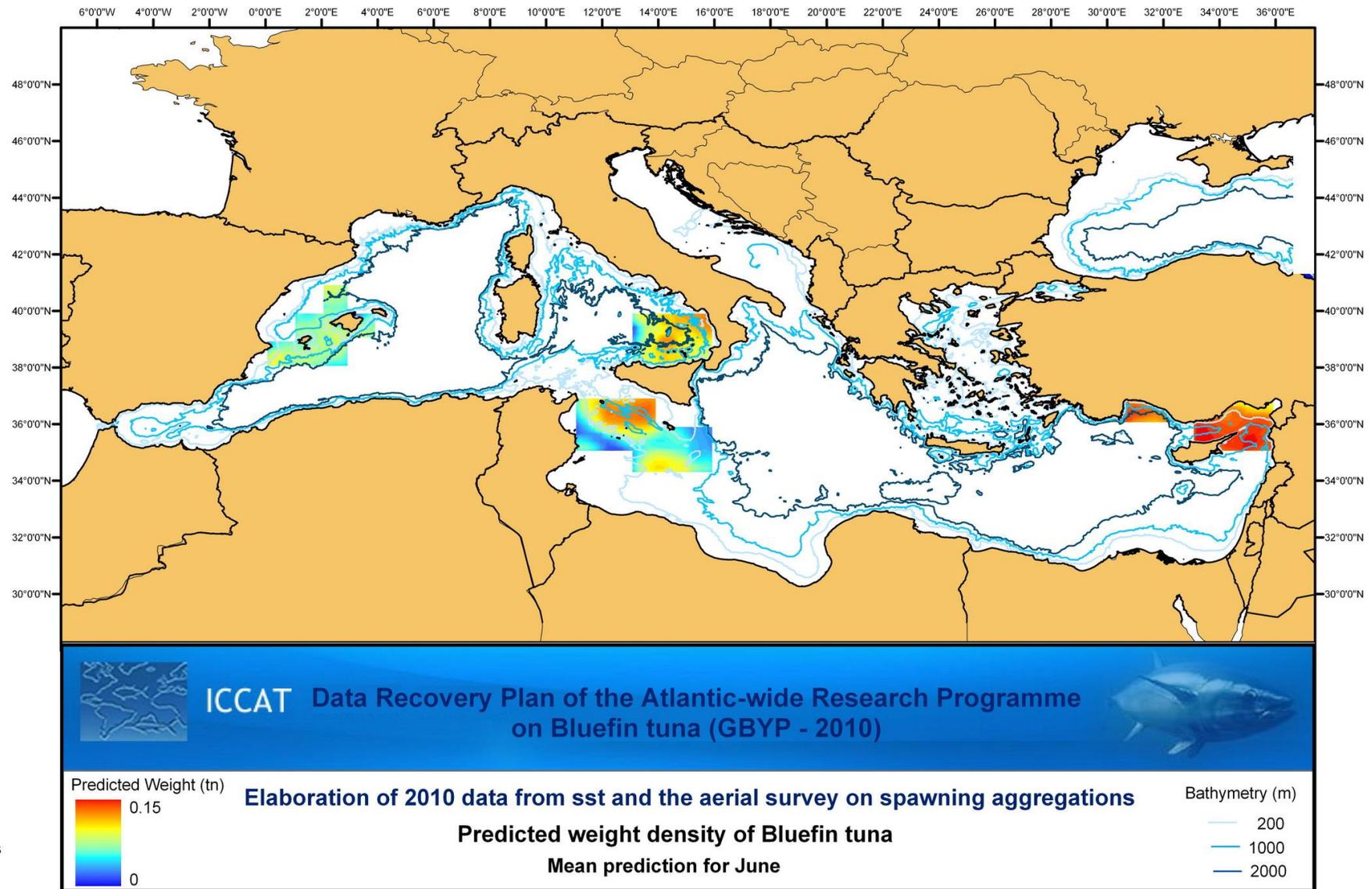


Figure 21. Predicted density of weight in tonnes of bluefin tuna in June 2010

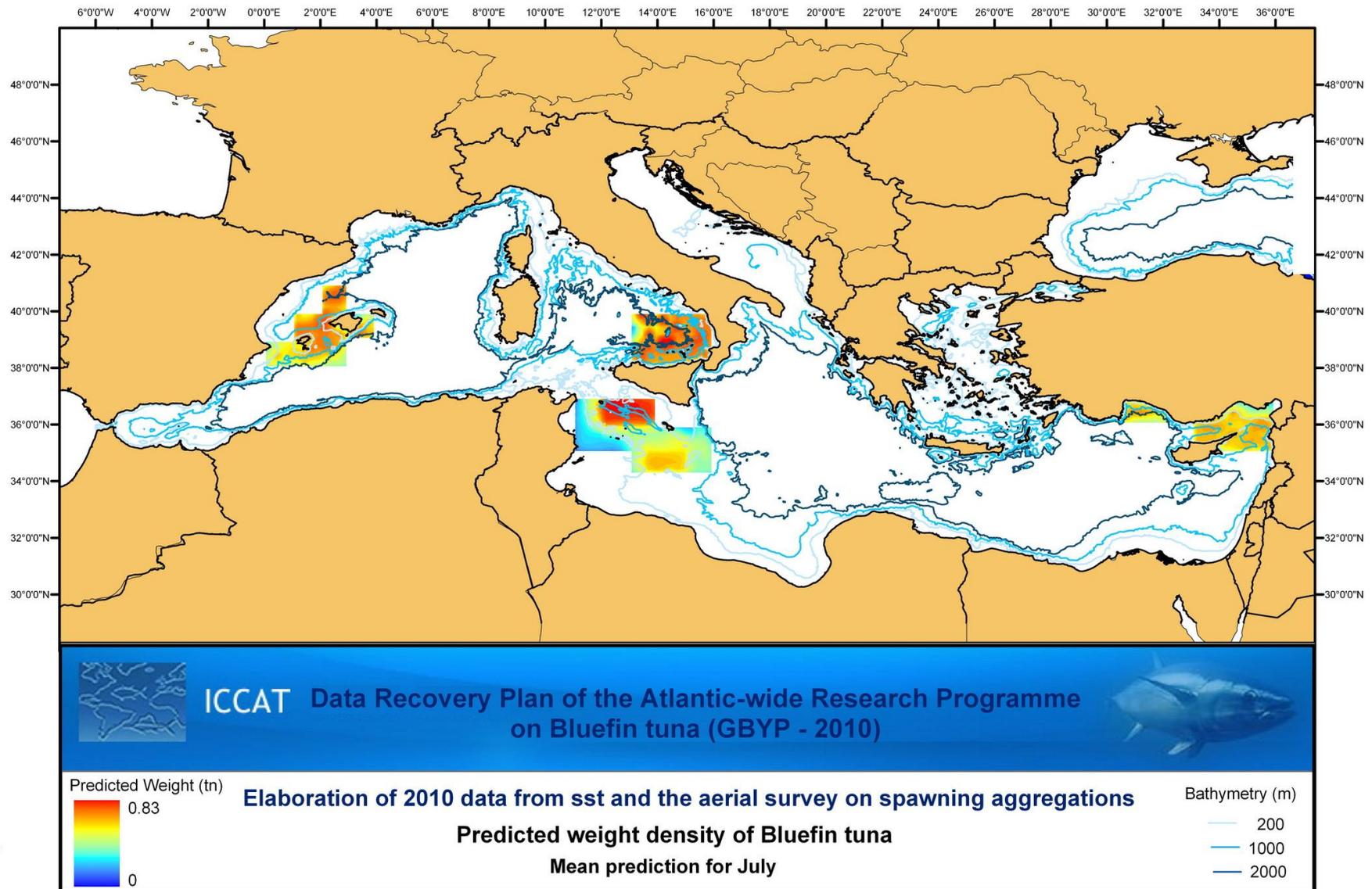


Figure 22. Predicted density of weight in tonnes of bluefin tuna in July 2010

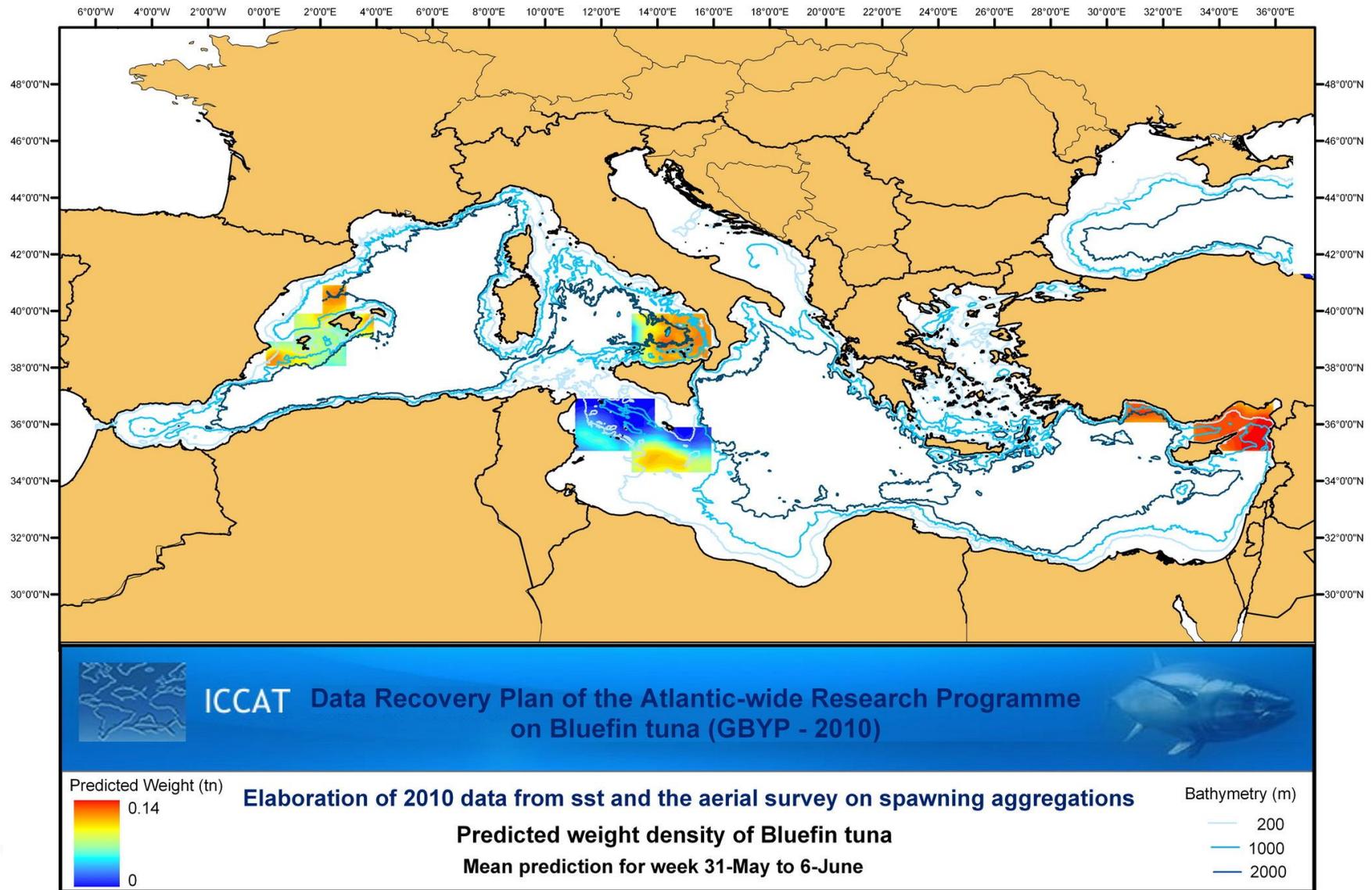


Figure 23. Predicted density of weight in tonnes of bluefin tuna for 31-May to 6-June 2010

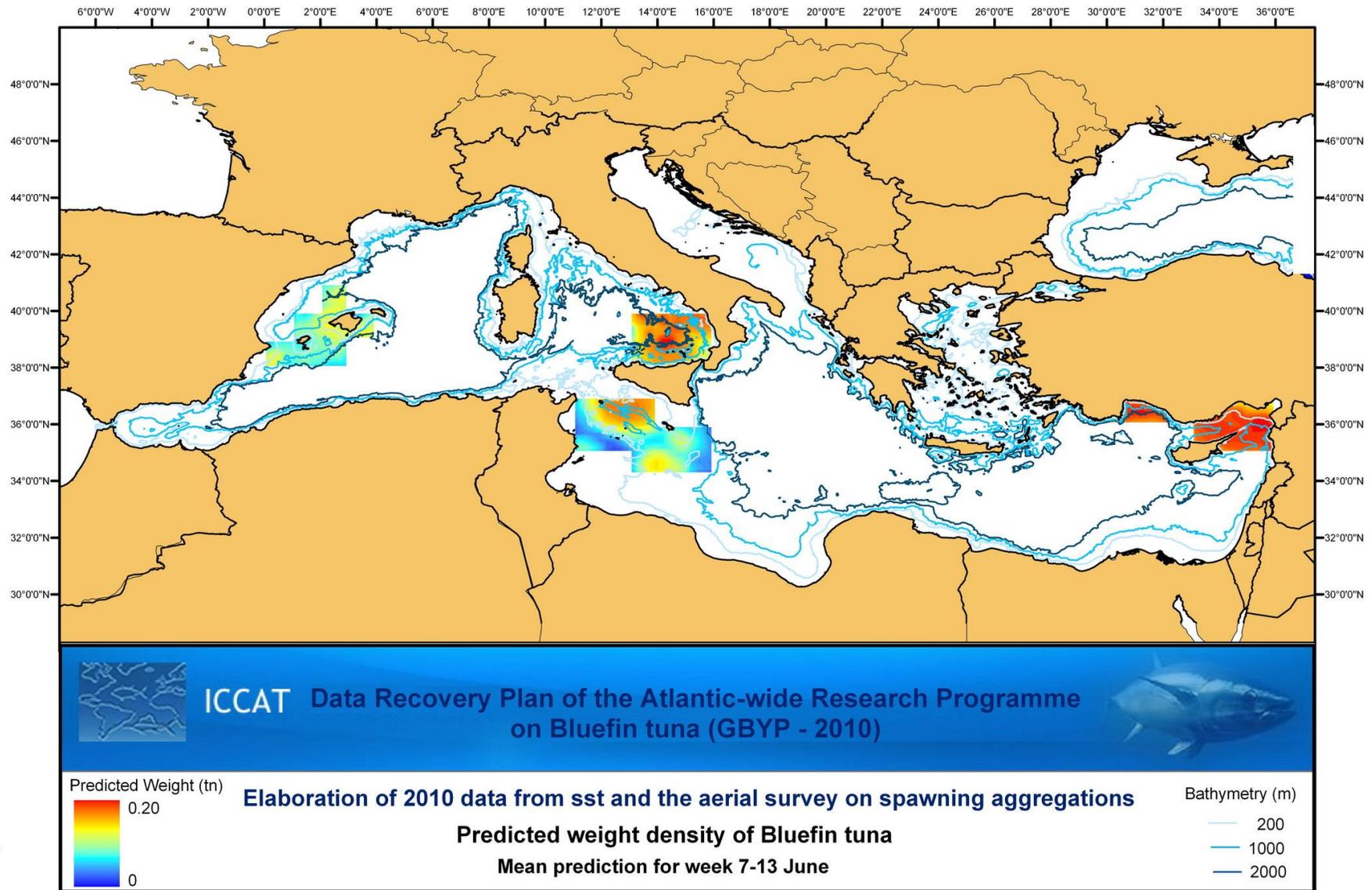


Figure 24. Predicted density of weight in tonnes of bluefin tuna for 7-13 June 2010

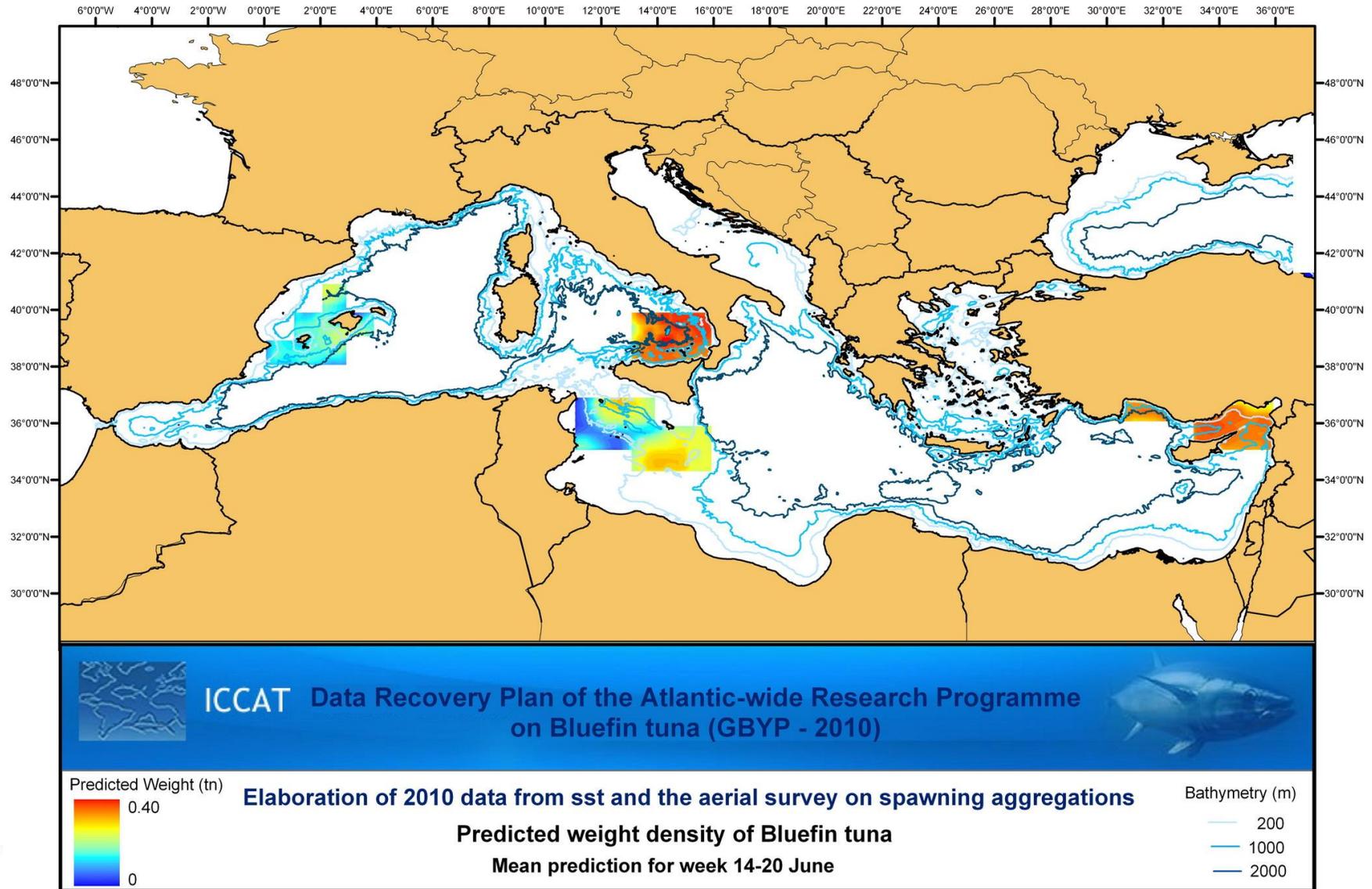


Figure 25. Predicted density of weight in tonnes of bluefin tuna for 14-20 June 2010

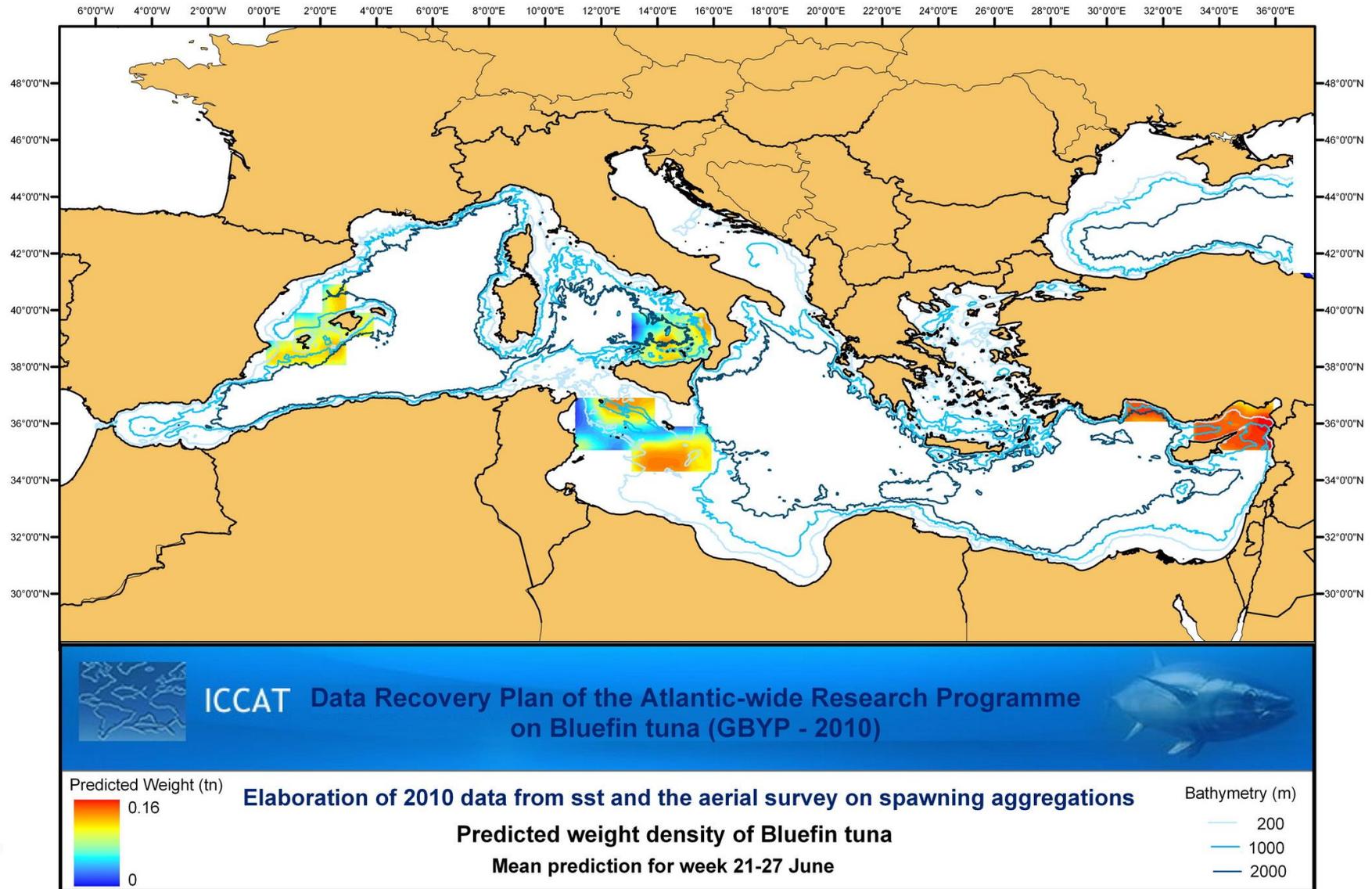


Figure 26. Predicted density of weight in tonnes of bluefin tuna for 21-27 June 2010

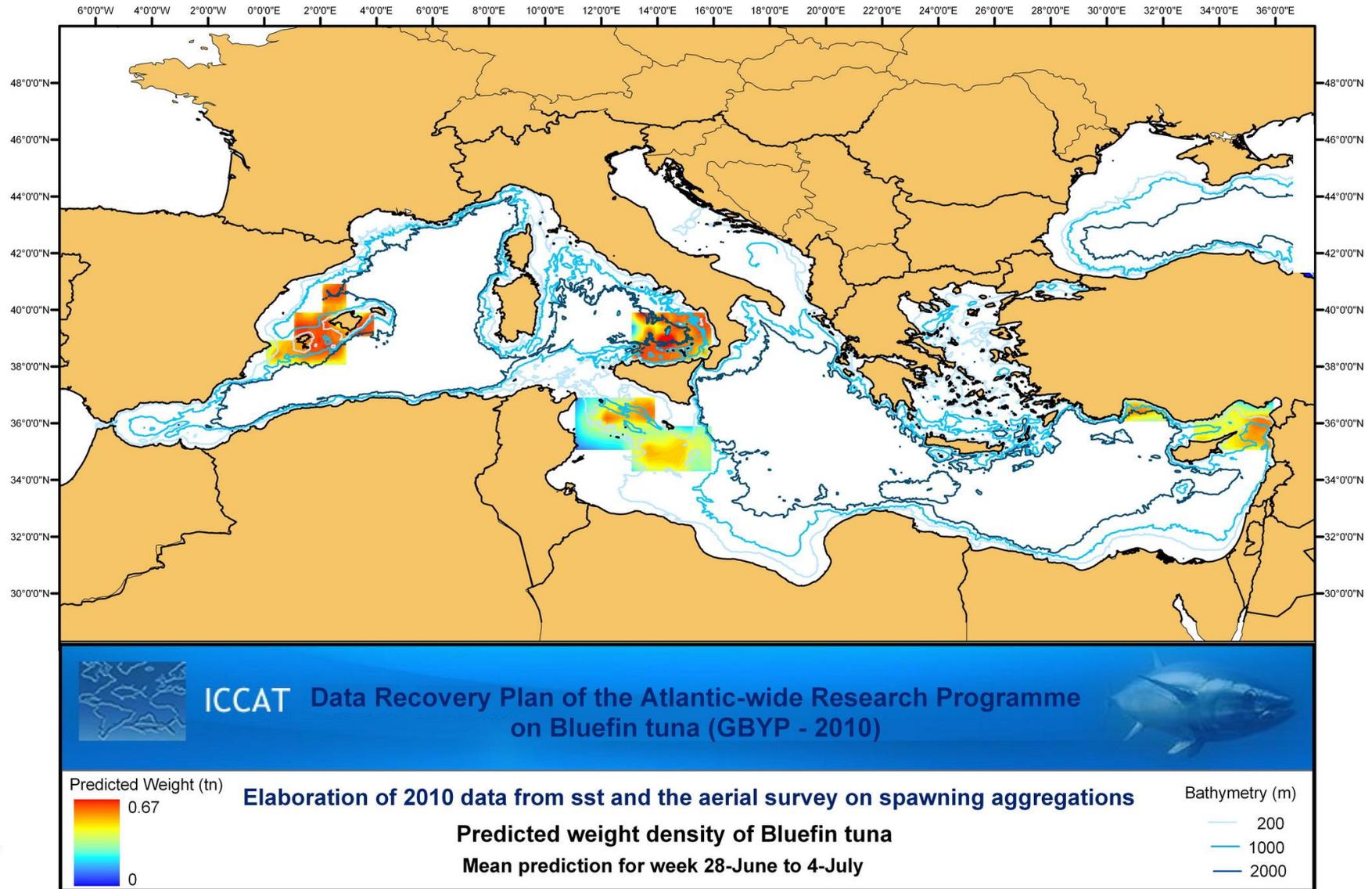


Figure 27. Predicted density of weight in tonnes of bluefin tuna for 28-June to 4 July 2010

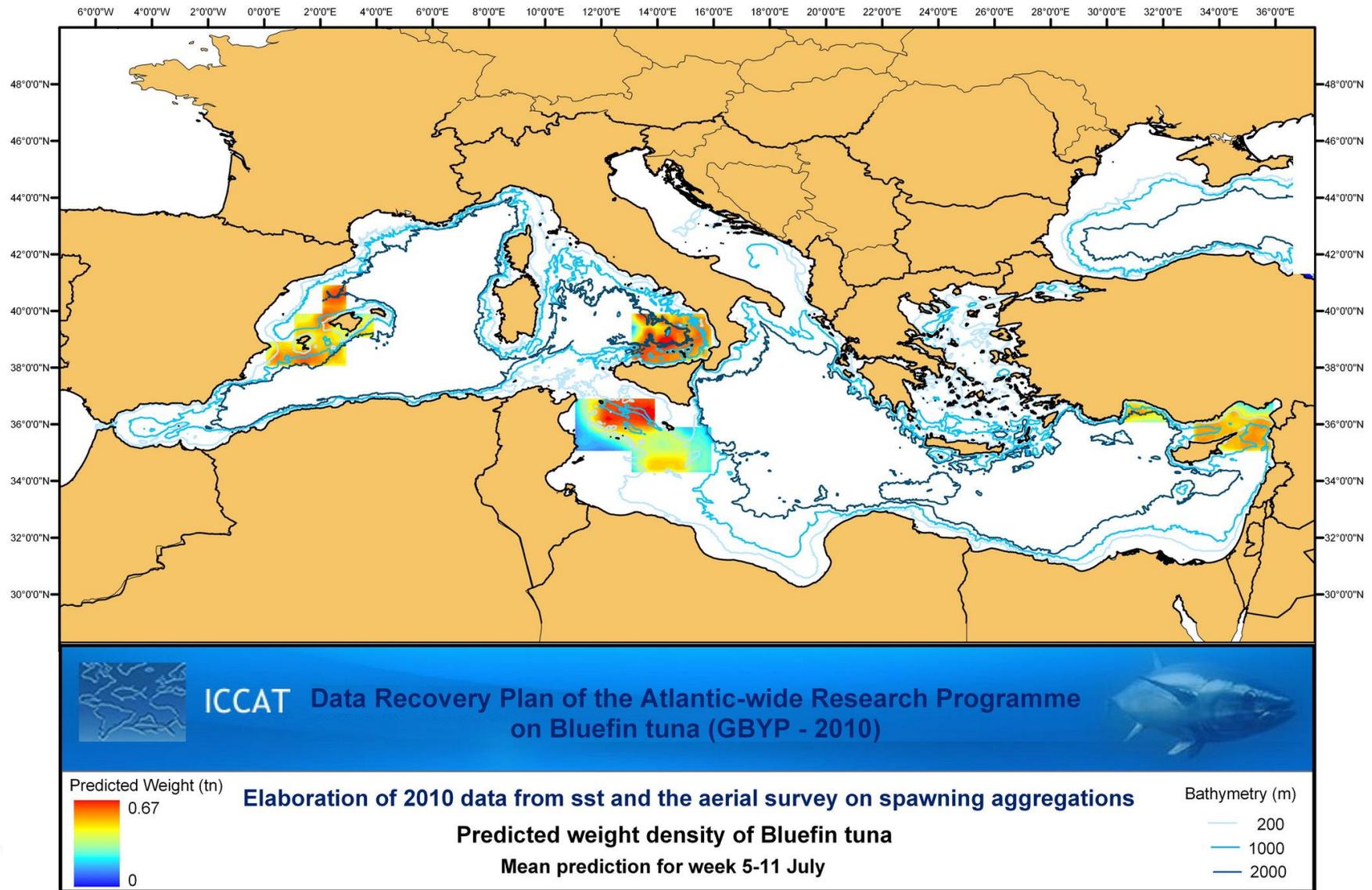


Figure 28. Predicted density of weight in tonnes of bluefin tuna for 5-11 July 2010

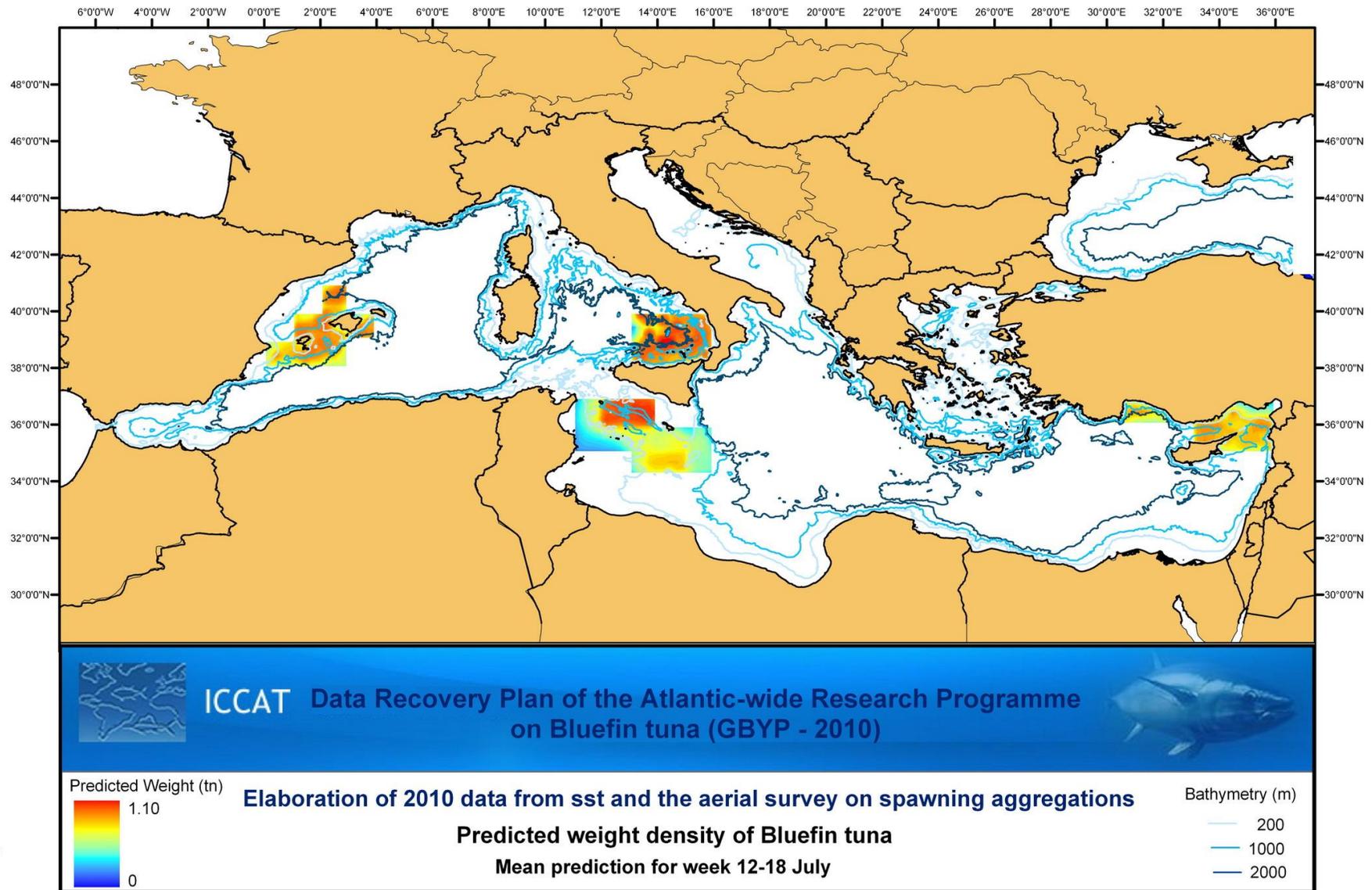


Figure 29. Predicted density of weight in tonnes of bluefin tuna for 12-18 July 2010

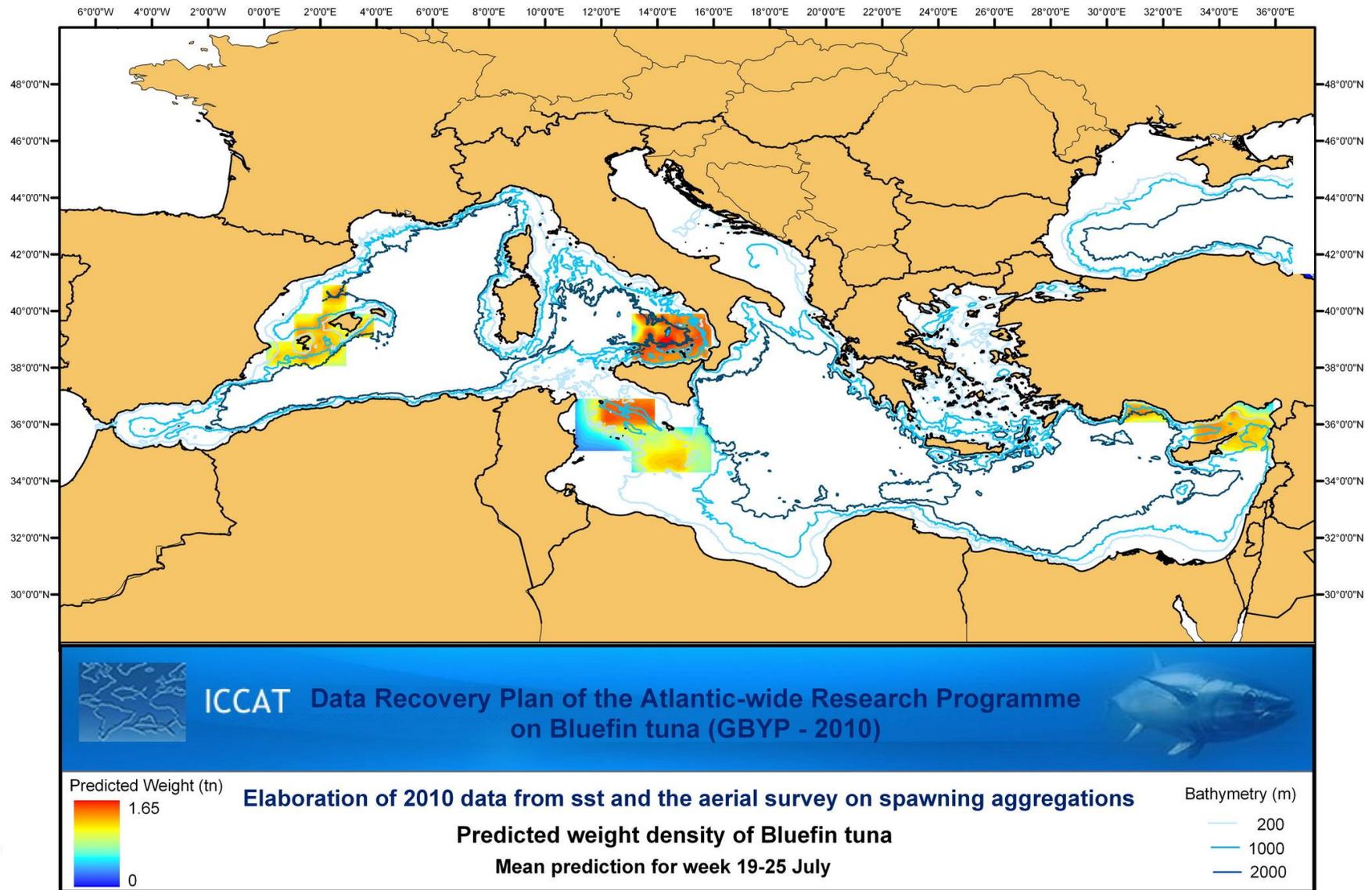


Figure 30. Predicted density of weight in tonnes of bluefin tuna for 19-25 July 2010

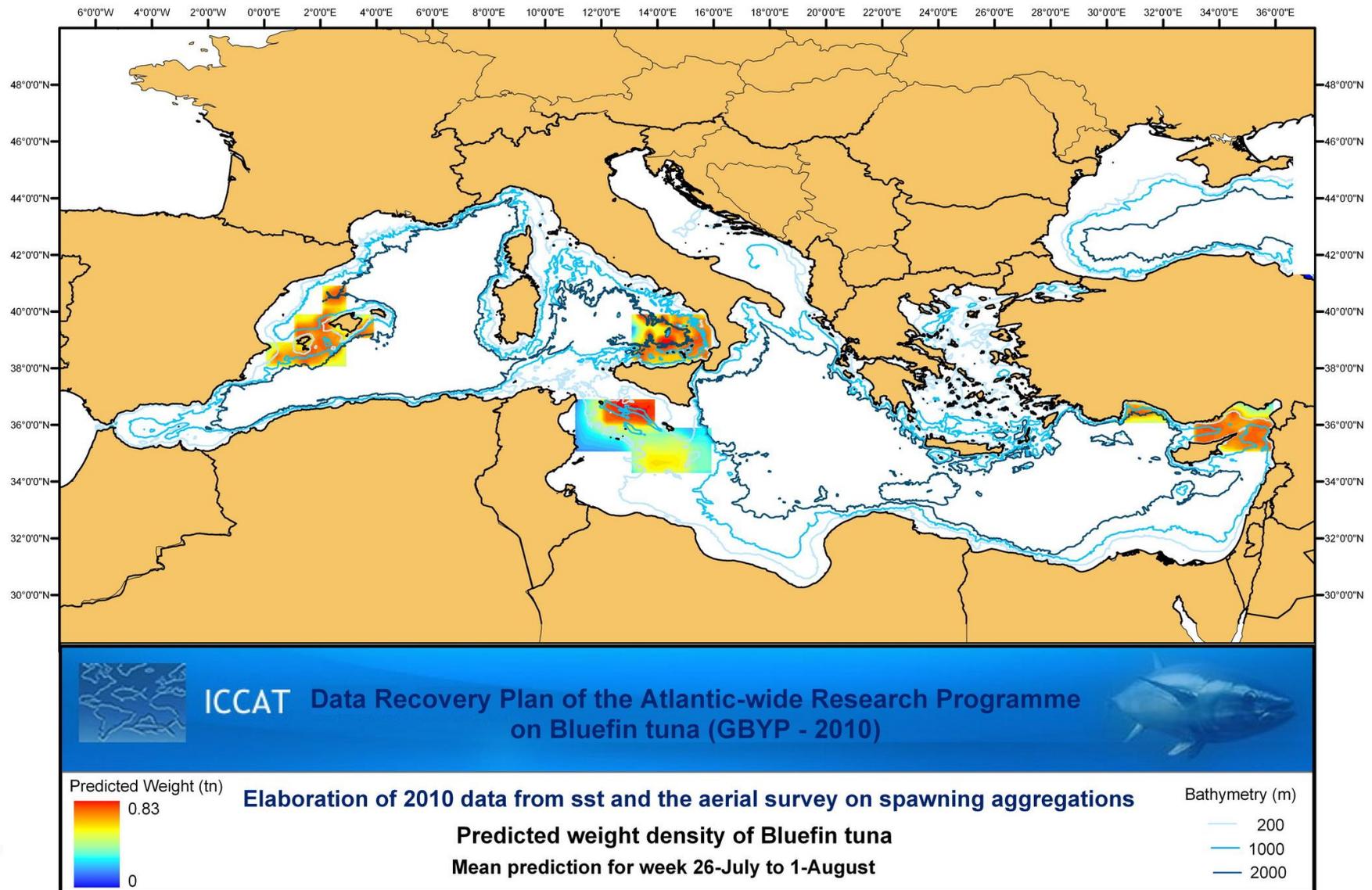


Figure 31. Predicted density of weight in tonnes of bluefin tuna for 26-July to 1-August 2010

Discussion

Spatial modelling to predict distribution and abundance is potentially a valuable analytical tool but its usefulness depends on the quality of the survey data. The more the survey data can be improved, the more value would be derived from the spatial modelling. For example, the only measure of school size we could use consistently was weight but it would be interesting to use number of fish (including of different size ranges) if these data could be collected consistently. Also, the greater the spatial coverage of the survey, the greater the reliability and applicability of the model results for the whole Mediterranean Sea.

All figures and Table 3 show a temporal pattern of an increase in density and total weight predicted from June to July, and in general from the earlier weeks to the later ones. This is driven by the general trend of increasing density with increasing sea surface temperature. Most of the survey (79%) was carried out in June, and this would explain the difference in predicted density between spatial modelling and conventional distance sampling (CDS), given that CDS does not take into account the spatial or temporal variation in density due to environmental covariates. This, and the higher proportion of survey effort in June, would also explain why the CDS estimates are closer to predicted density in June than in July.

A consistent pattern shown also by the maps is the shift of higher densities from the East to the West over the weeks, as sst increases in the western areas.

The results presented here do not include estimates of coefficients of variation (CV) or confidence limits. These need to be generated through non-parametric bootstrapping of the data and refitting the models; there has been insufficient time in this contract to accomplish this. Additionally, the limited quantity and quality of the available survey data do not warrant undertaking this task at this stage. Nevertheless, estimates of precision are essential if results are to be used for management and this should be done in future spatial modelling analyses of aerial survey data.

Abundance or weight can be predicted as an exploratory exercise over any area of the Mediterranean or over the whole basin, as shown in Figures 32 to 42. These predictions are an exploration and for illustration only. Nevertheless, they could be useful in comparison of predicted high density areas with, for example, fishing areas, and also to help direct the spatial coverage of future surveys. These Mediterranean-wide predictions should not be used for management advice.

In conclusion, although the results presented here are limited by the available data (coverage, quantity and quality), they are of value in indicating the kind of results that can be generated by spatial modelling of survey and environmental data and in helping to plan the survey for 2011 in terms of timing and areas.

References

- Buckland, ST, Anderson, DR, Burnham, KP, Laake, JL, Borchers, DL & Thomas, L (2001). *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University Press, Oxford.
- Hammond, Cañadas & Vázquez (2010). Atlantic-wide research programme on bluefin tuna (GBYP - 2010) - Design for aerial line transect survey in the Mediterranean Sea. Final Report to ICCAT. May 2010.
- Hedley, S.L., Buckland, S.T. & Borchers, D.L. 1999. Spatial modelling from line transect data. *Journal of Cetacean Research and Management*, 1 (3): 255-260.
- Hiby, L. & Hammond, P.S. 1989. Survey techniques for estimating abundance of cetaceans. *Reports of the International Whaling Commission (Special Issue 11)*: 47-80.
- Wood, S.N. 2000. Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties. *J.R.Statist.Soc.B* 62(2):413-428
- Wood, S. N. 2001. "mgcv: GAMs and Generalized Ridge Regression for R." *R News* 1(2): 20-25.

Annex

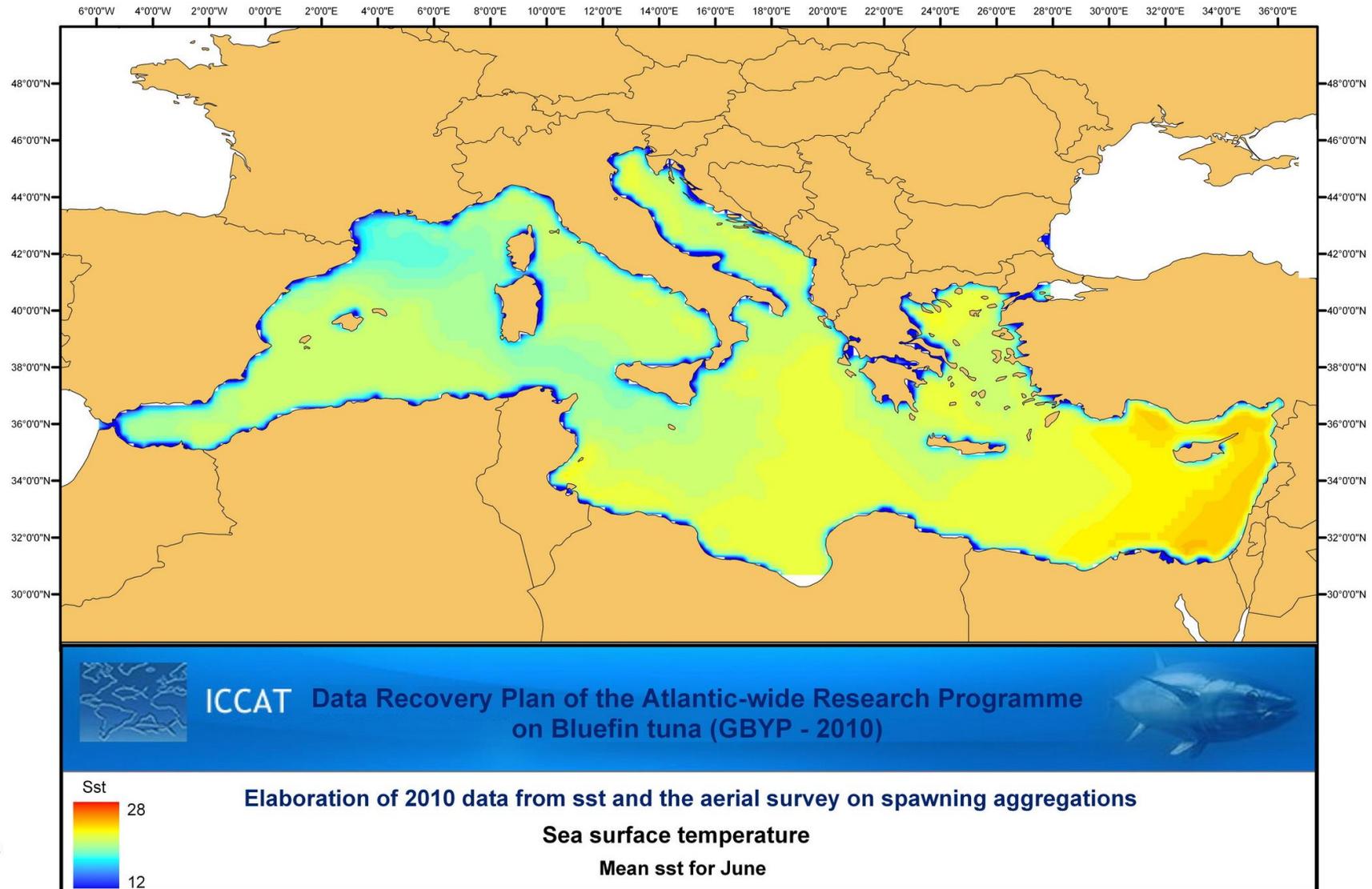


Figure 10. Mean sea surface temperature in June 2010

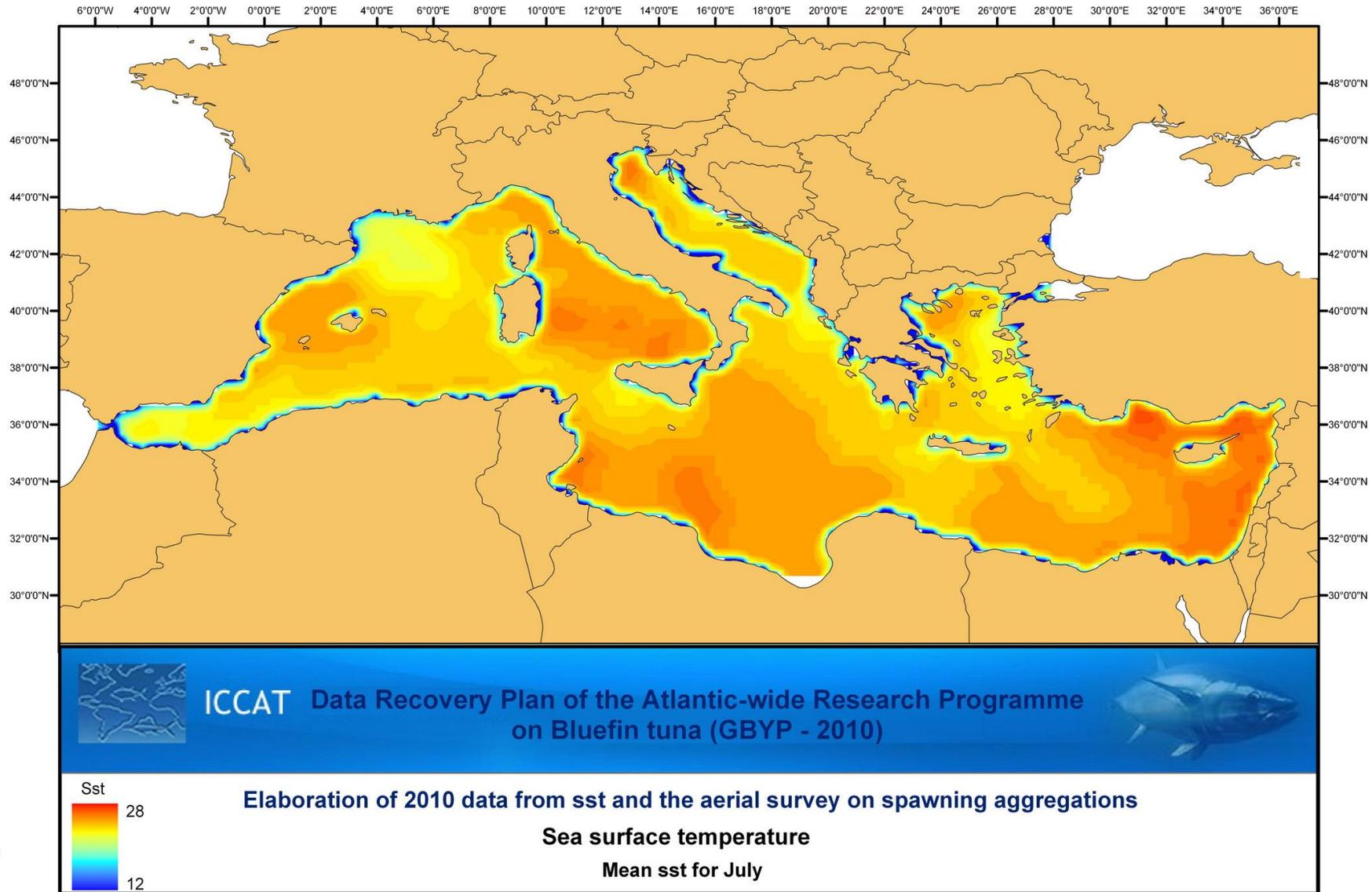


Figure 11. Mean sea surface temperature in July 2010

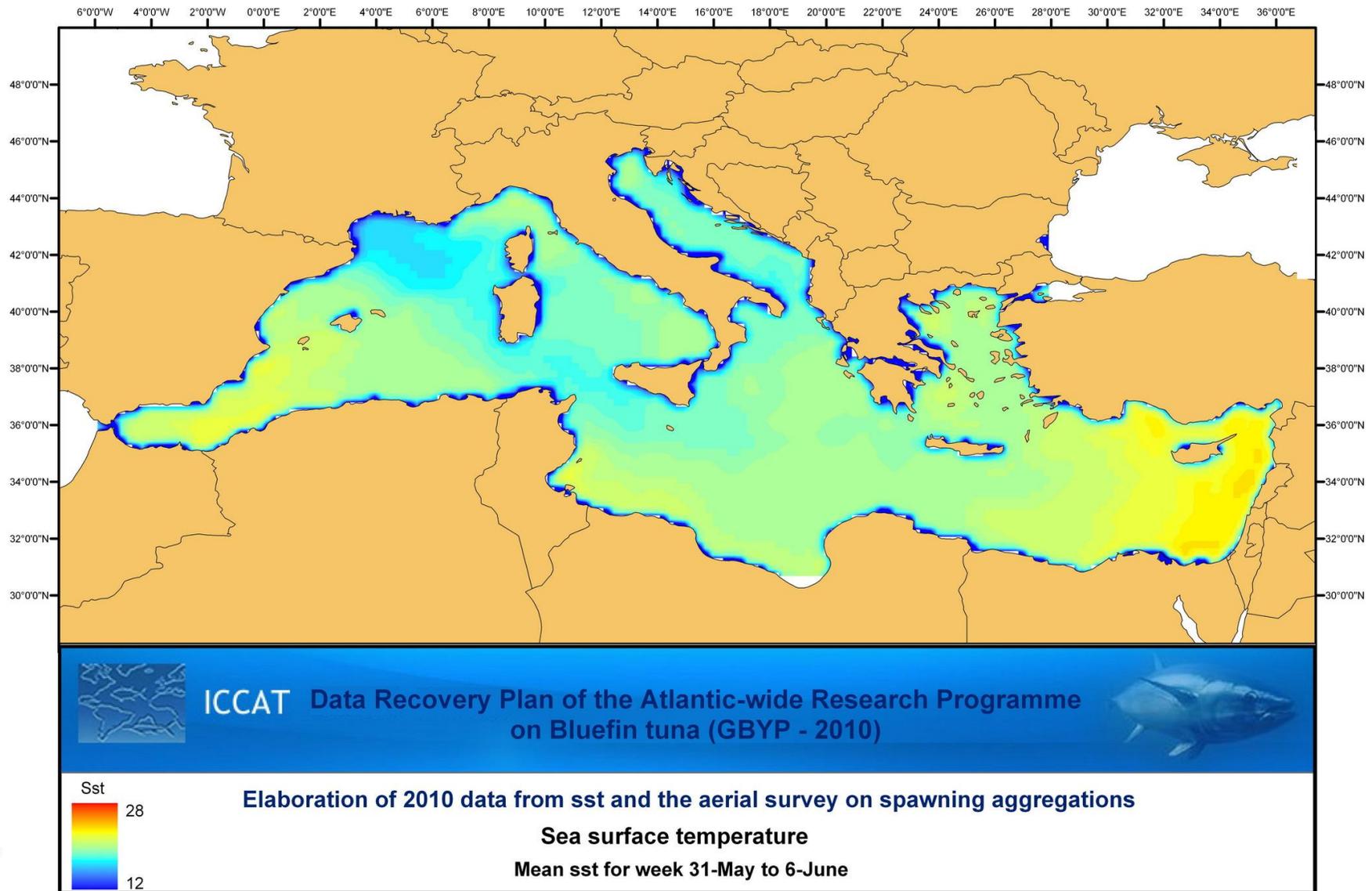


Figure 12. Mean sea surface temperature for 31-May to 6-June 2010

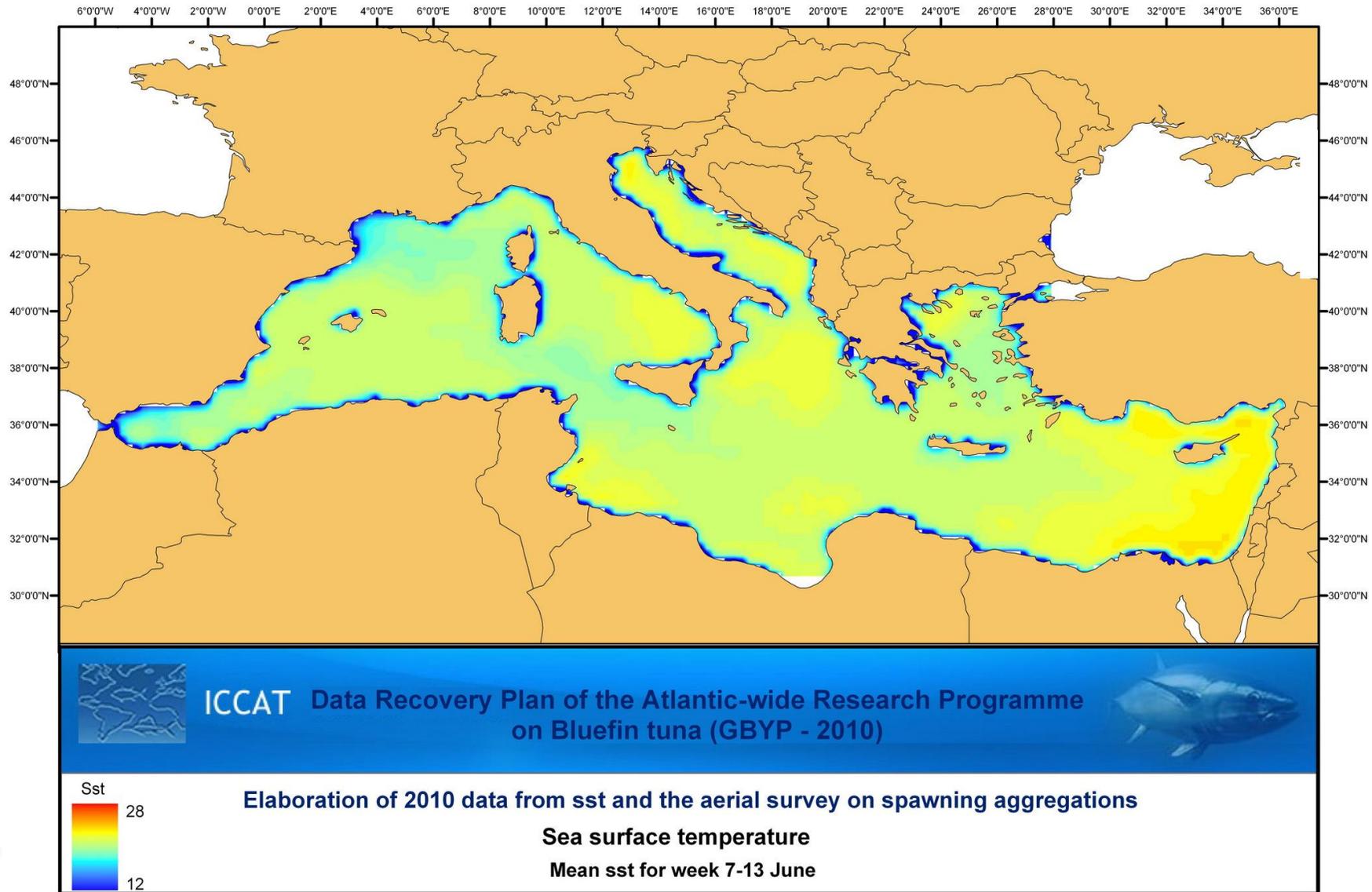


Figure 13. Mean sea surface temperature for 7-13 June 2010

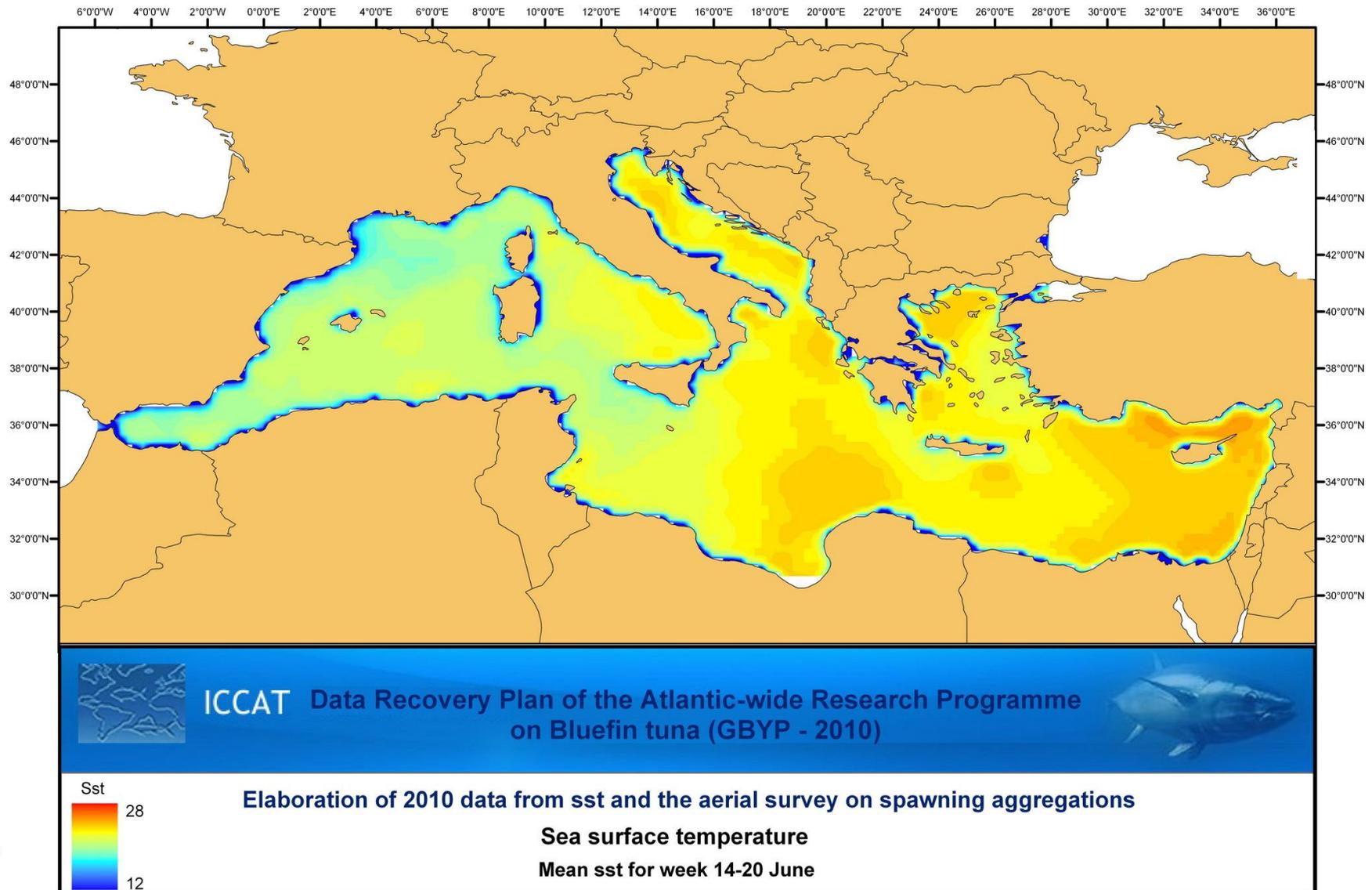


Figure 14. Mean sea surface temperature for 14-20 June 2010

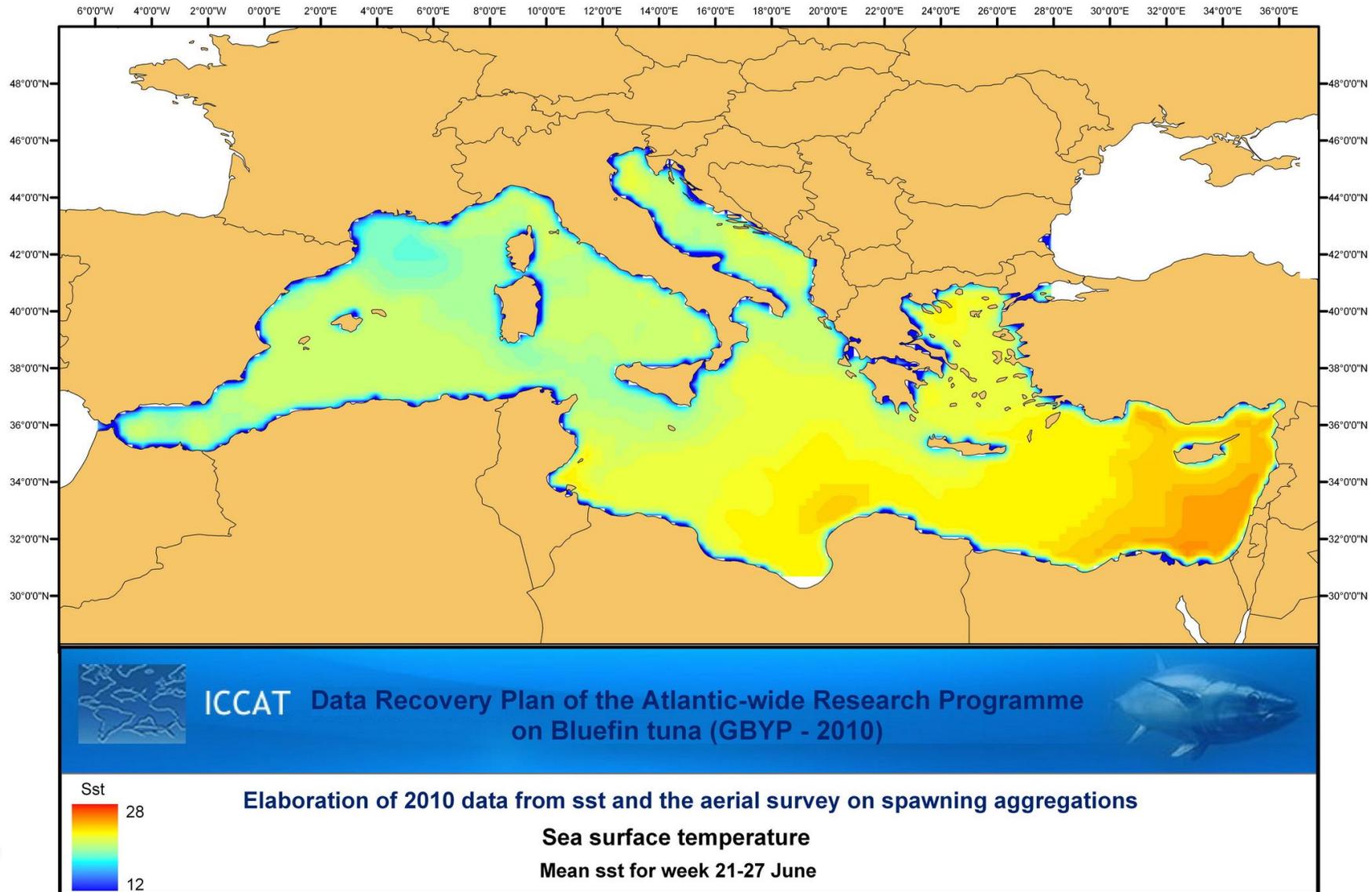


Figure 15. Mean sea surface temperature for 21-27 June 2010

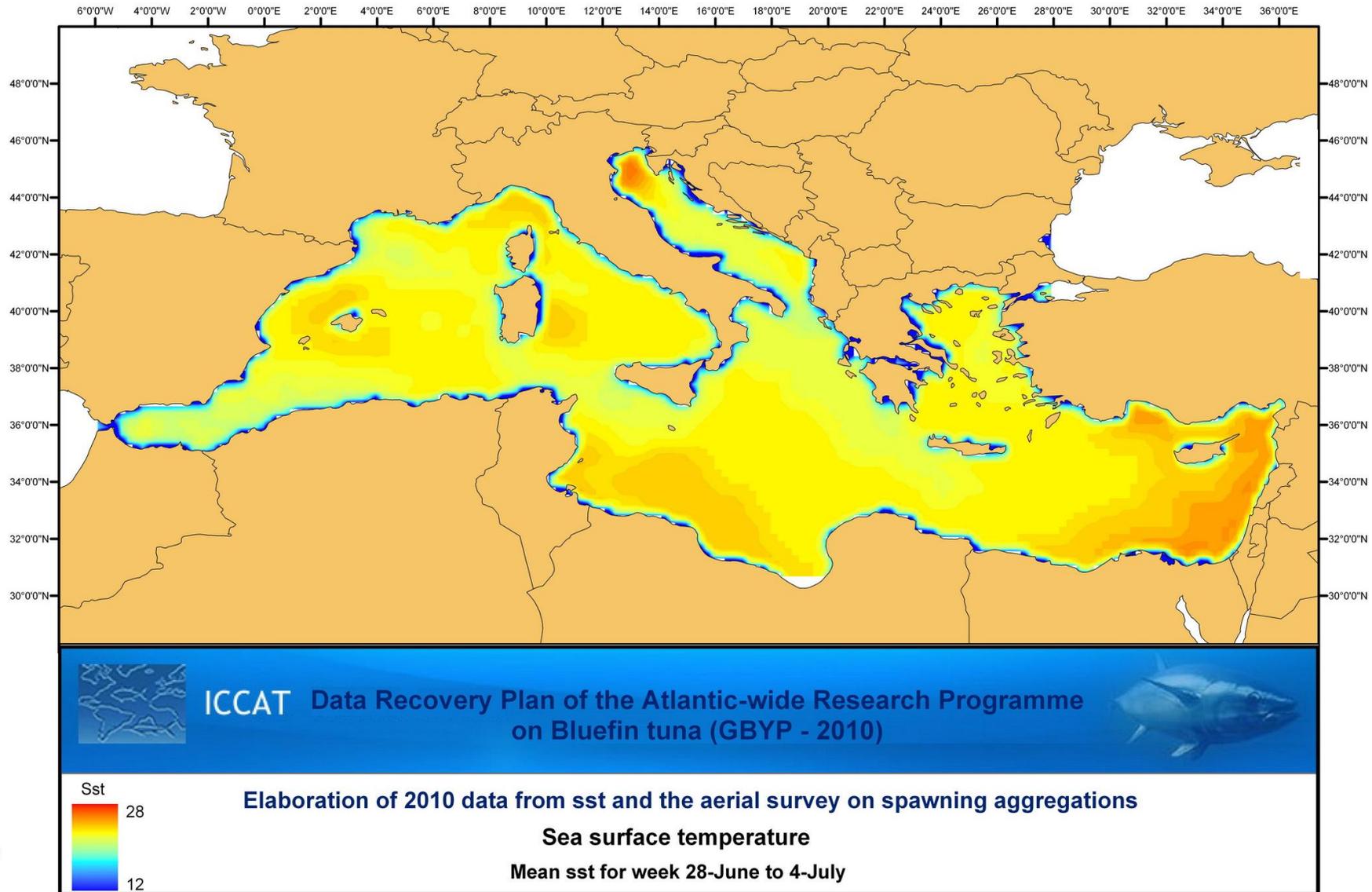


Figure 16. Mean sea surface temperature for 28-June to 4 July 2010

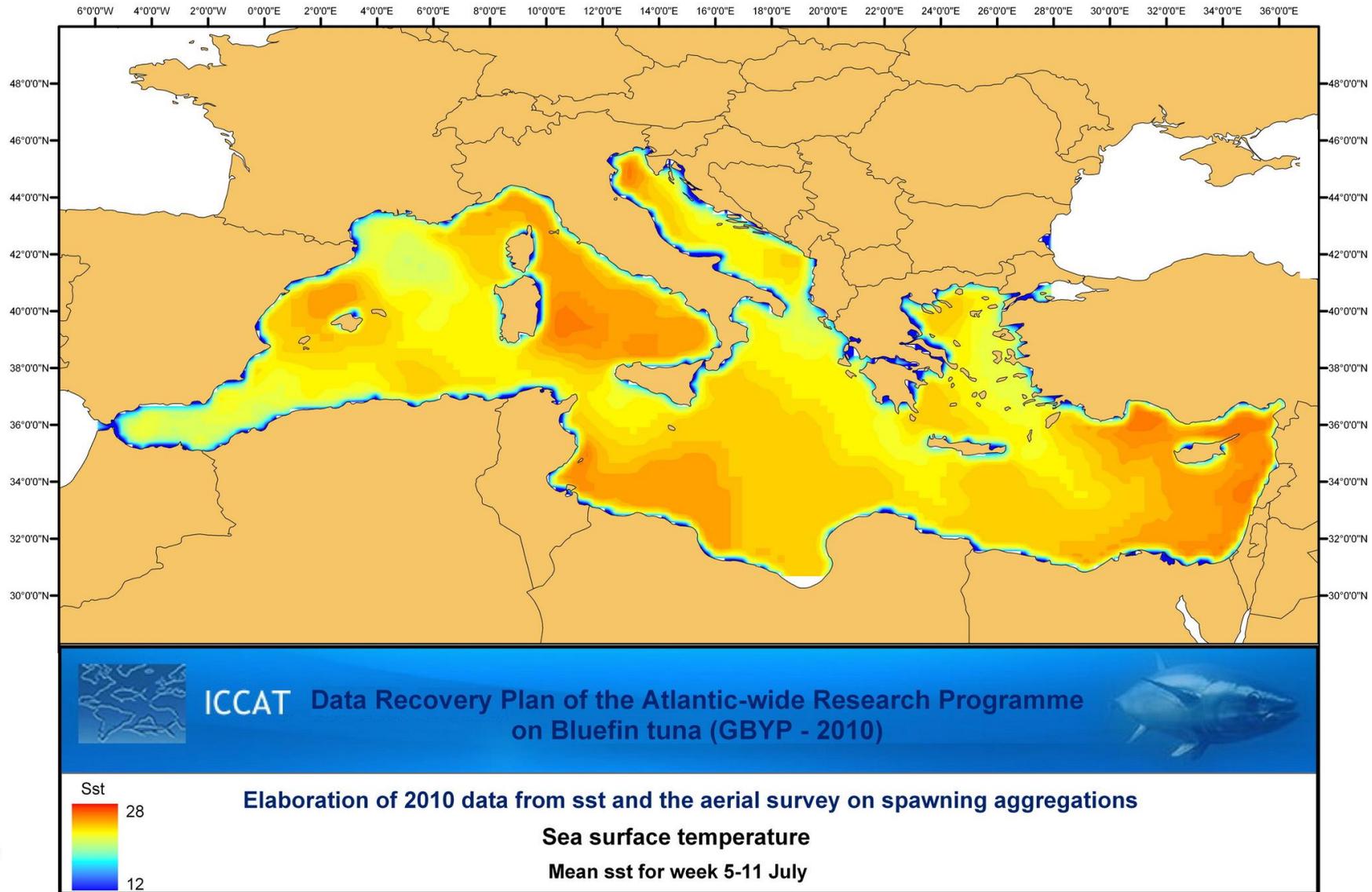


Figure 17. Mean sea surface temperature for 5-11 July 2010

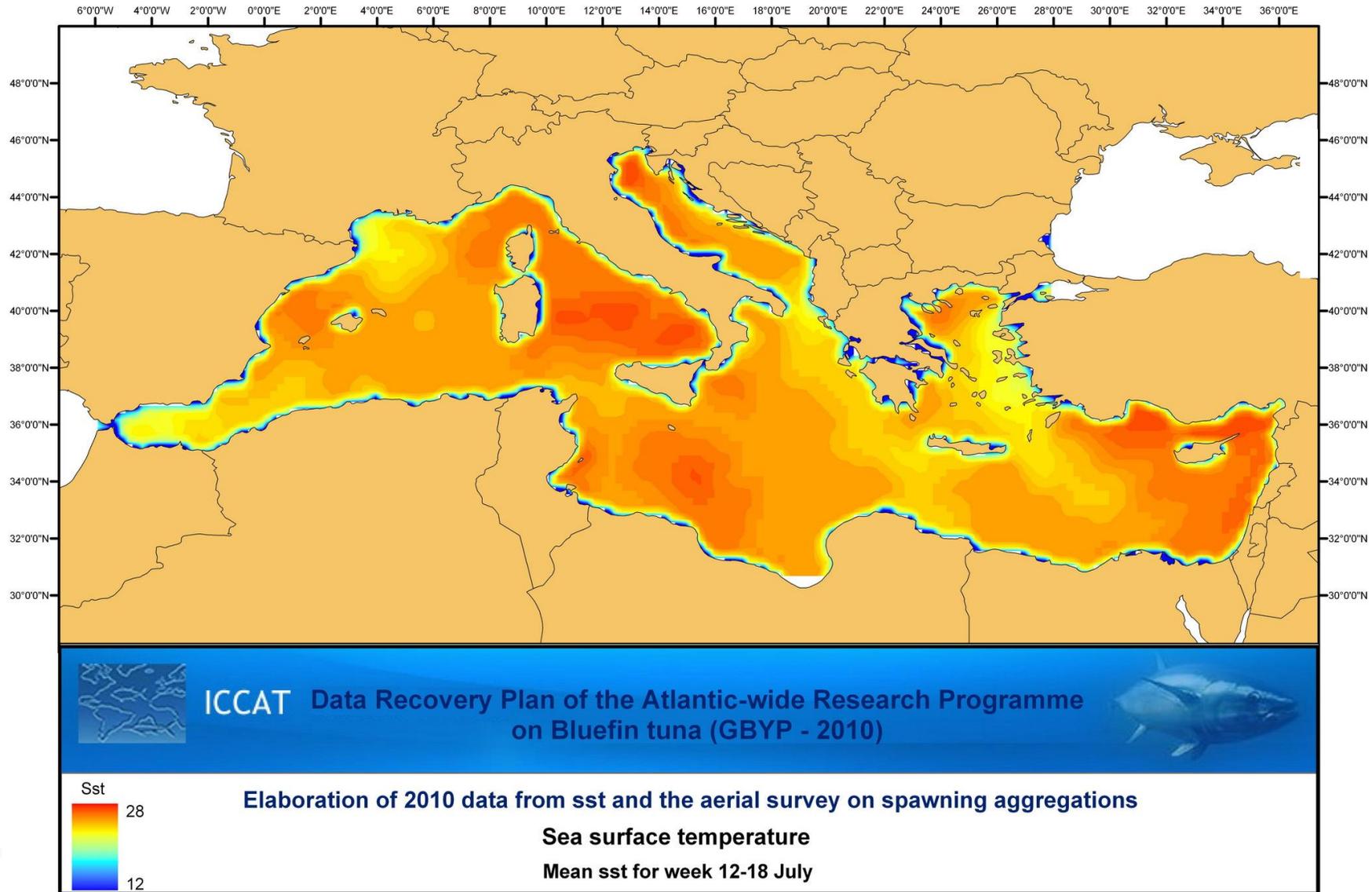


Figure 18. Mean sea surface temperature for 12-18 July 2010

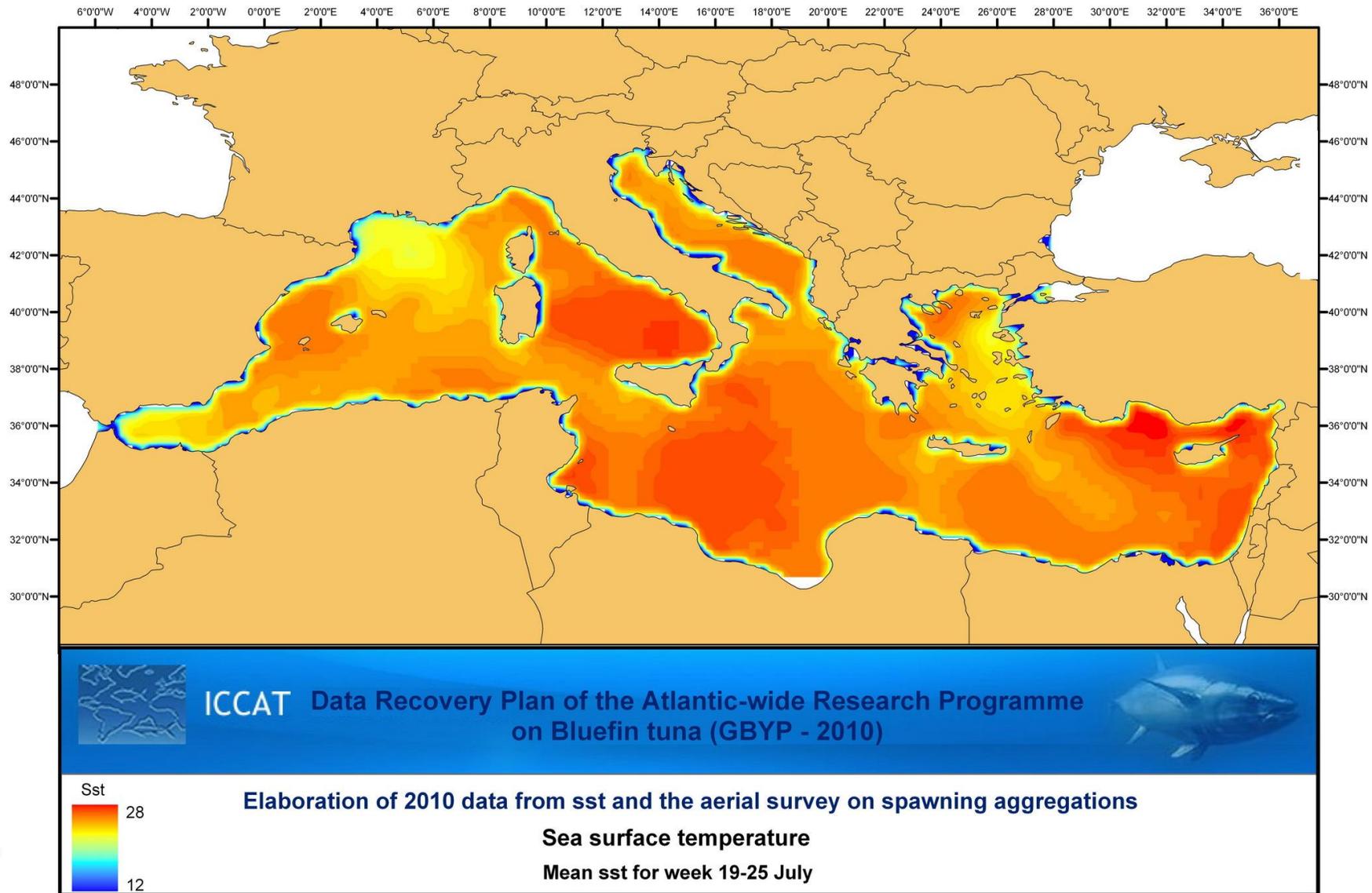


Figure 19. Mean sea surface temperature for 19-25 July 2010

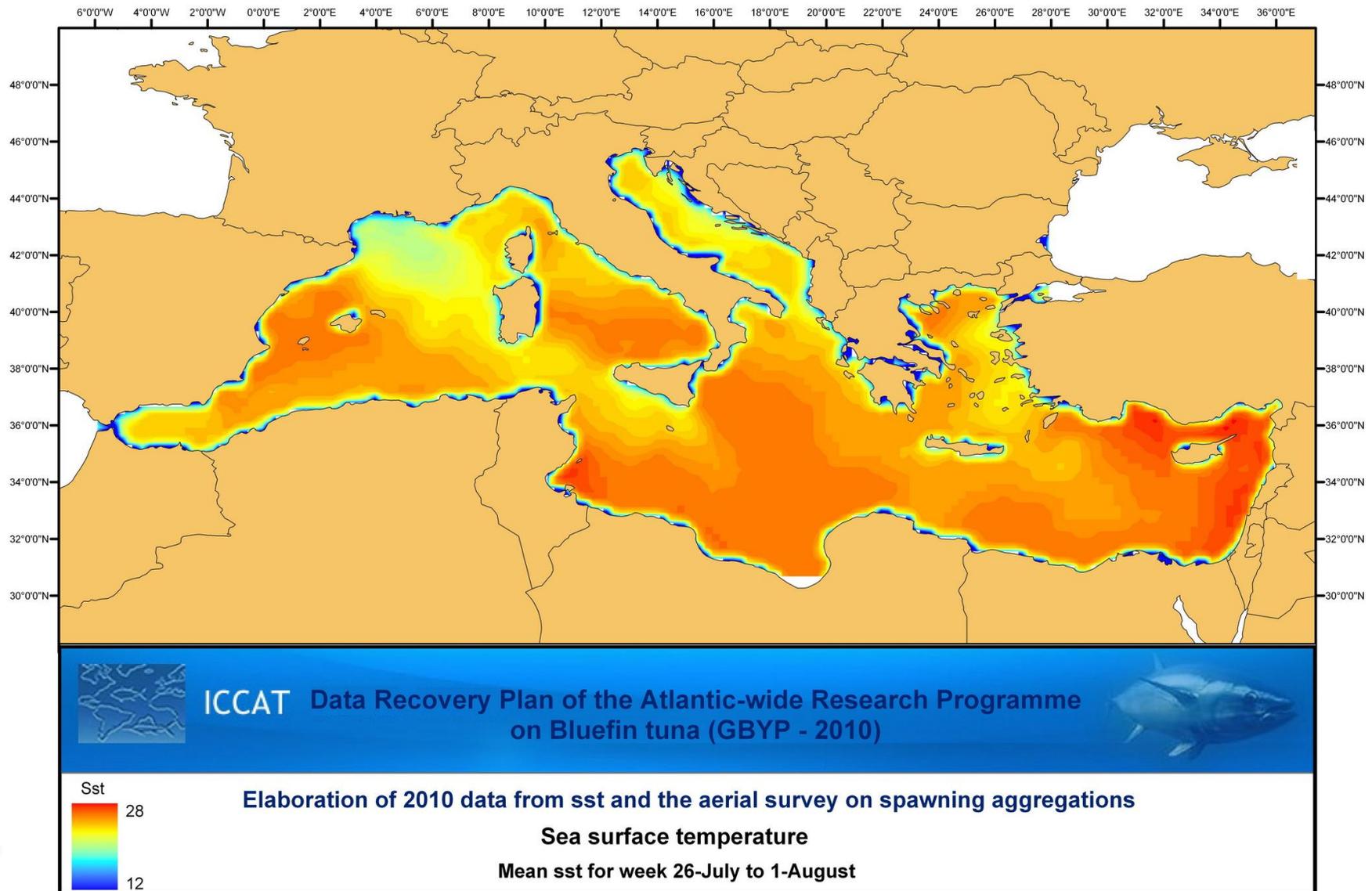


Figure 20. Mean sea surface temperature for 26-July to 1-August 2010

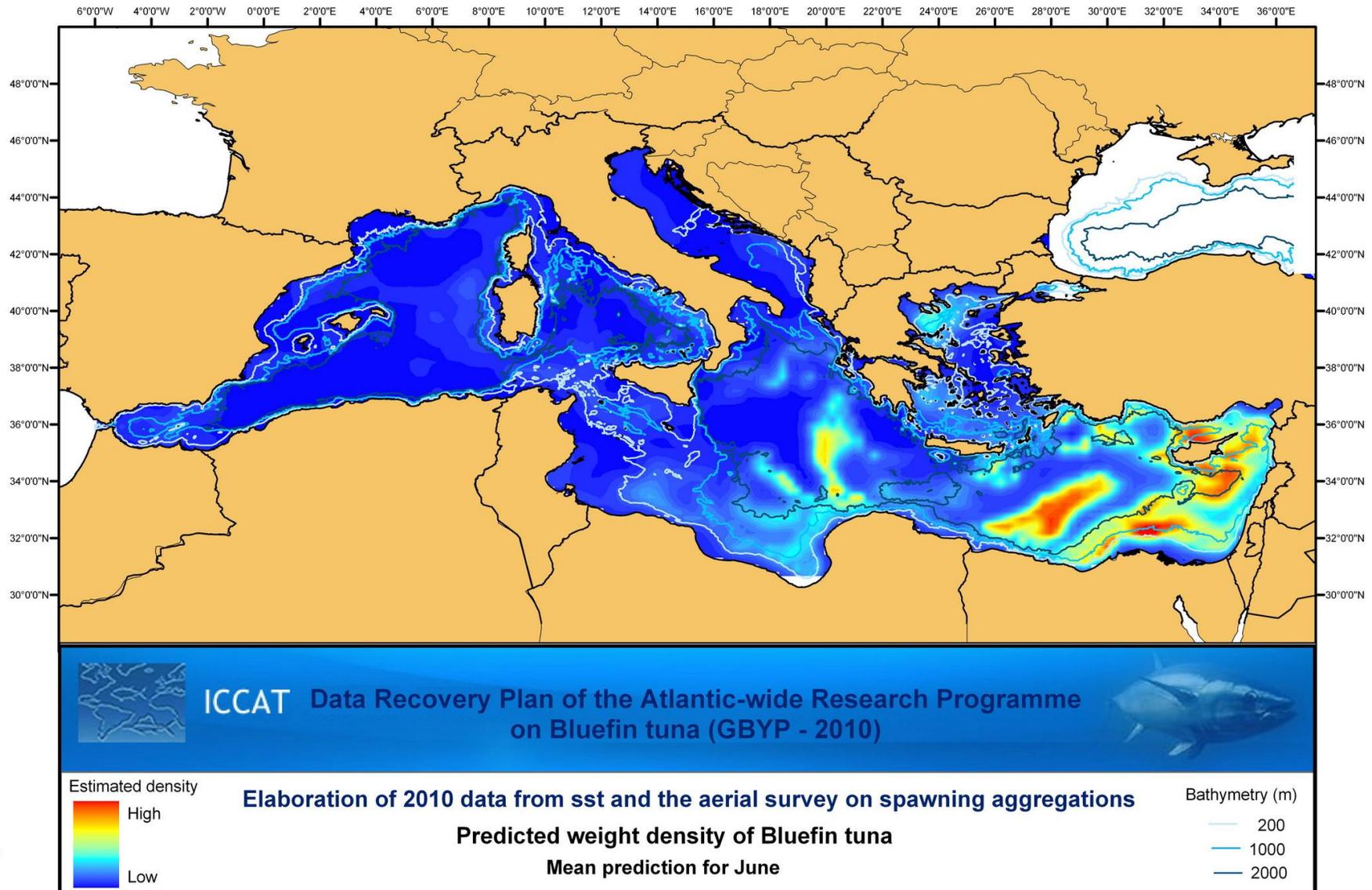


Figure 32. Predicted density of bluefin tuna in June 2010

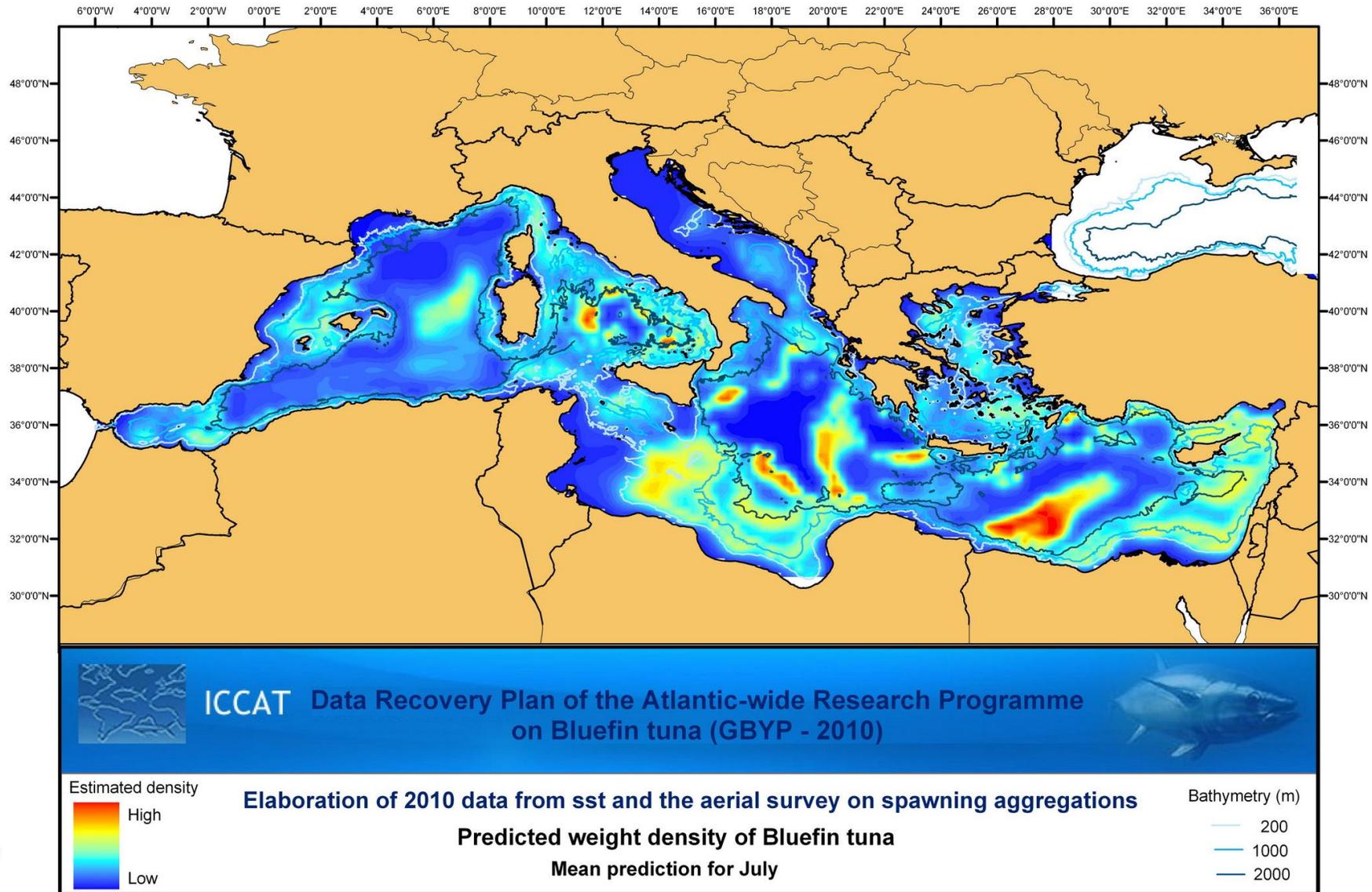


Figure 33. Predicted density of bluefin tuna in July 2010

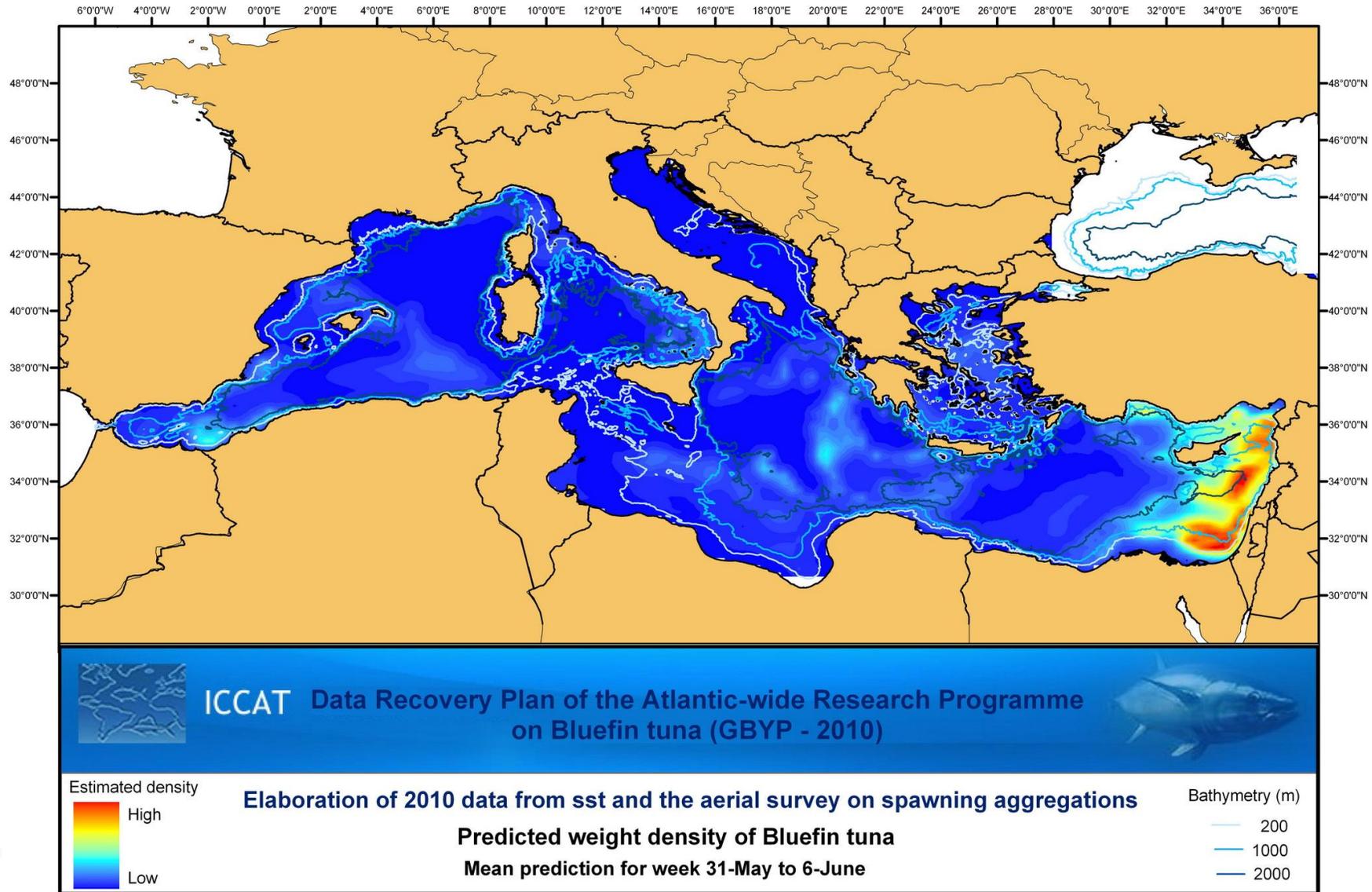


Figure 34. Predicted density of bluefin tuna for 31-May to 6-June 2010

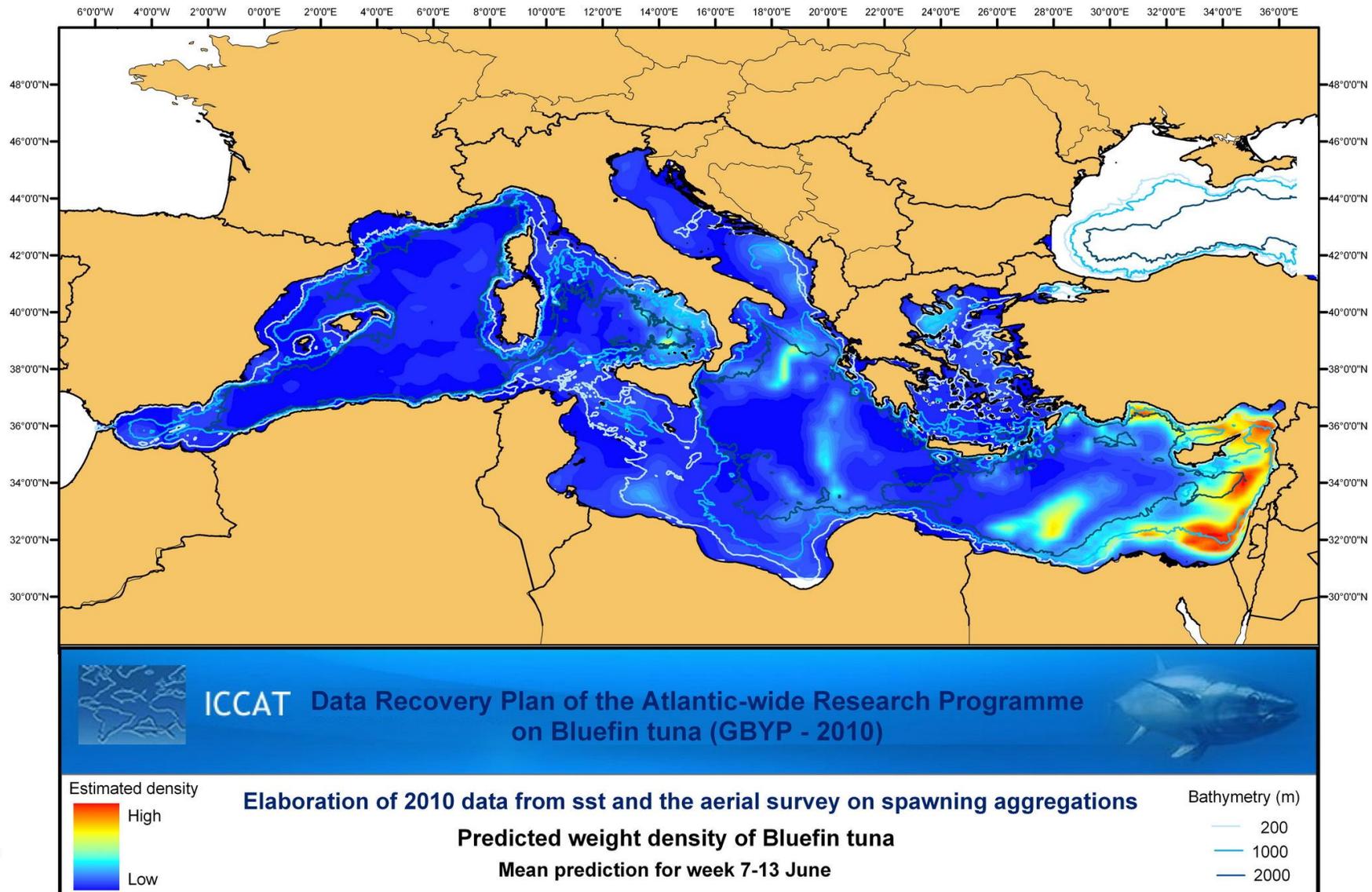


Figure 35. Predicted density of bluefin tuna for 7-13 June 2010

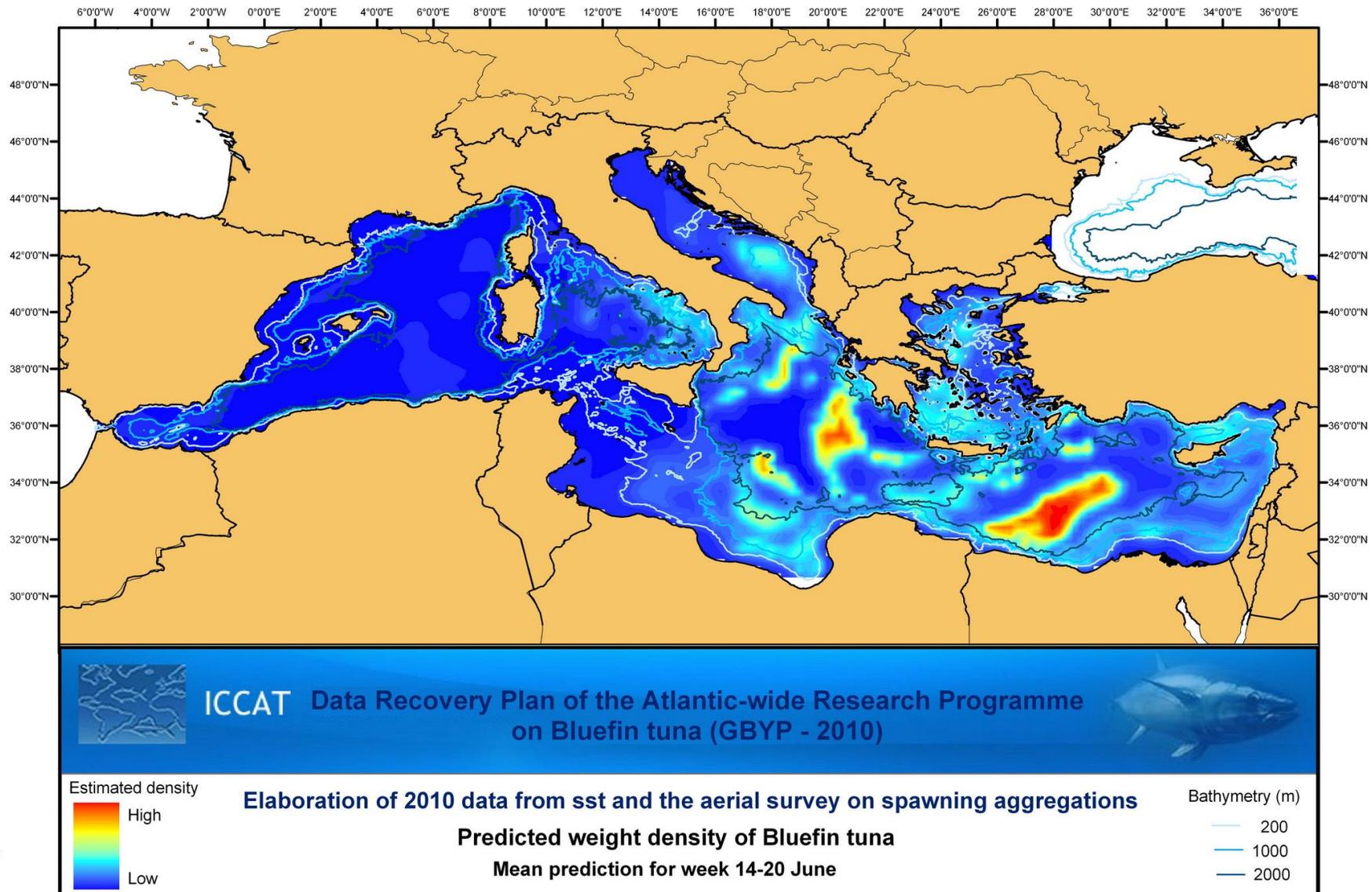


Figure 36. Predicted density of bluefin tuna for 14-20 June 2010

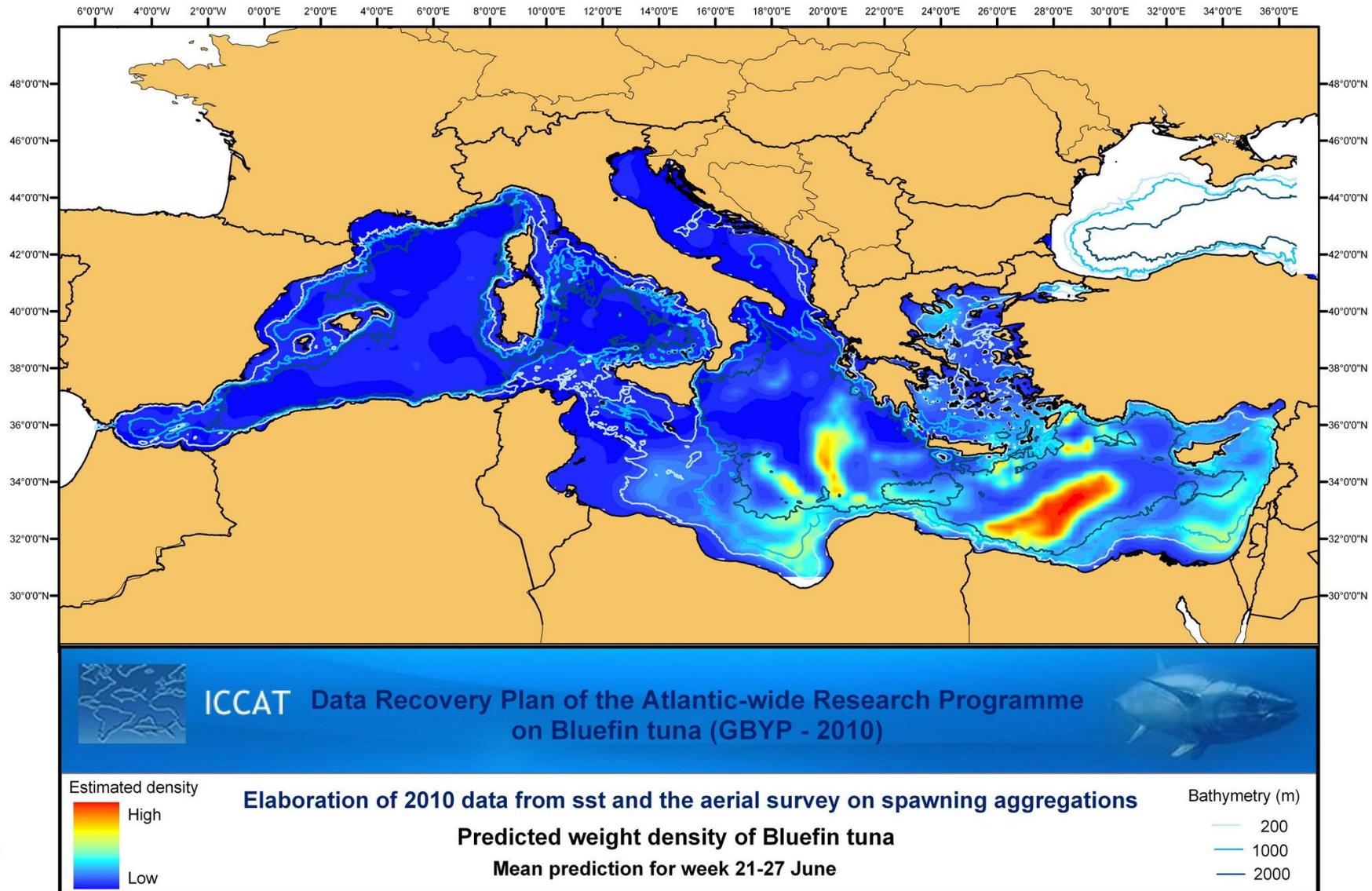


Figure 37. Predicted density of bluefin tuna for 21-27 June 2010

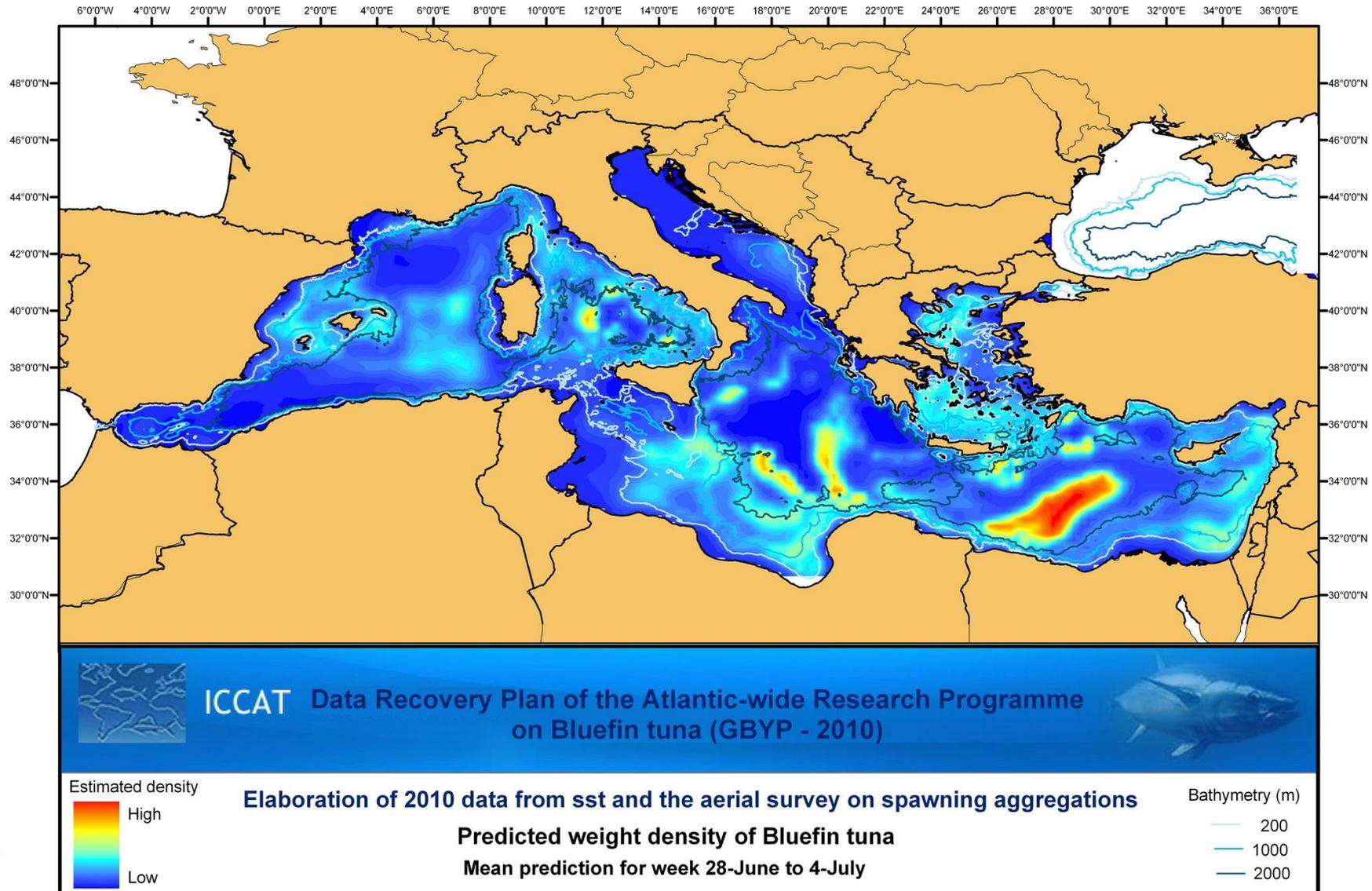


Figure 38. Predicted density of bluefin tuna for 28-June to 4 July 2010

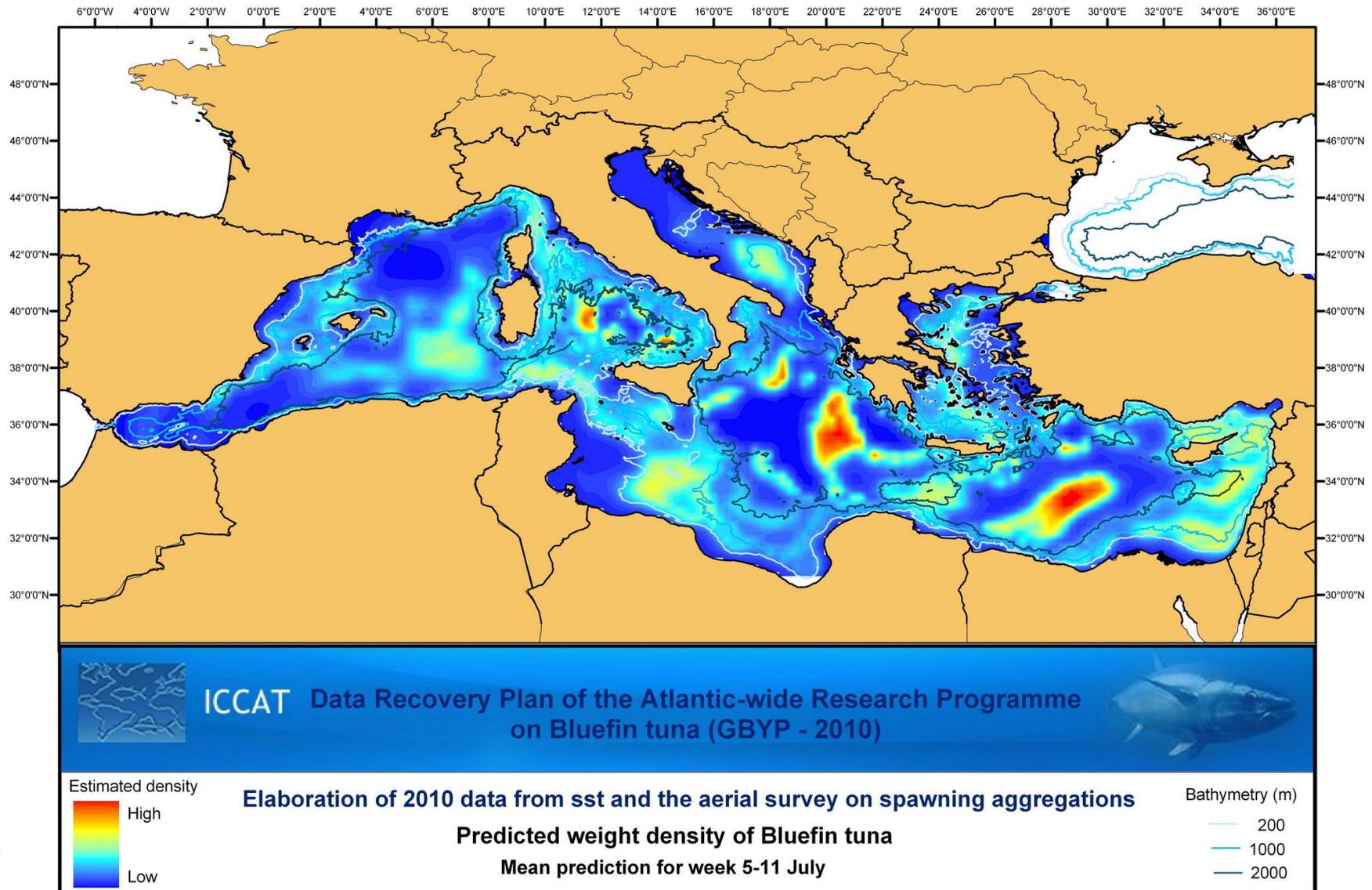


Figure 39. Predicted density of bluefin tuna for 5-11 July 2010

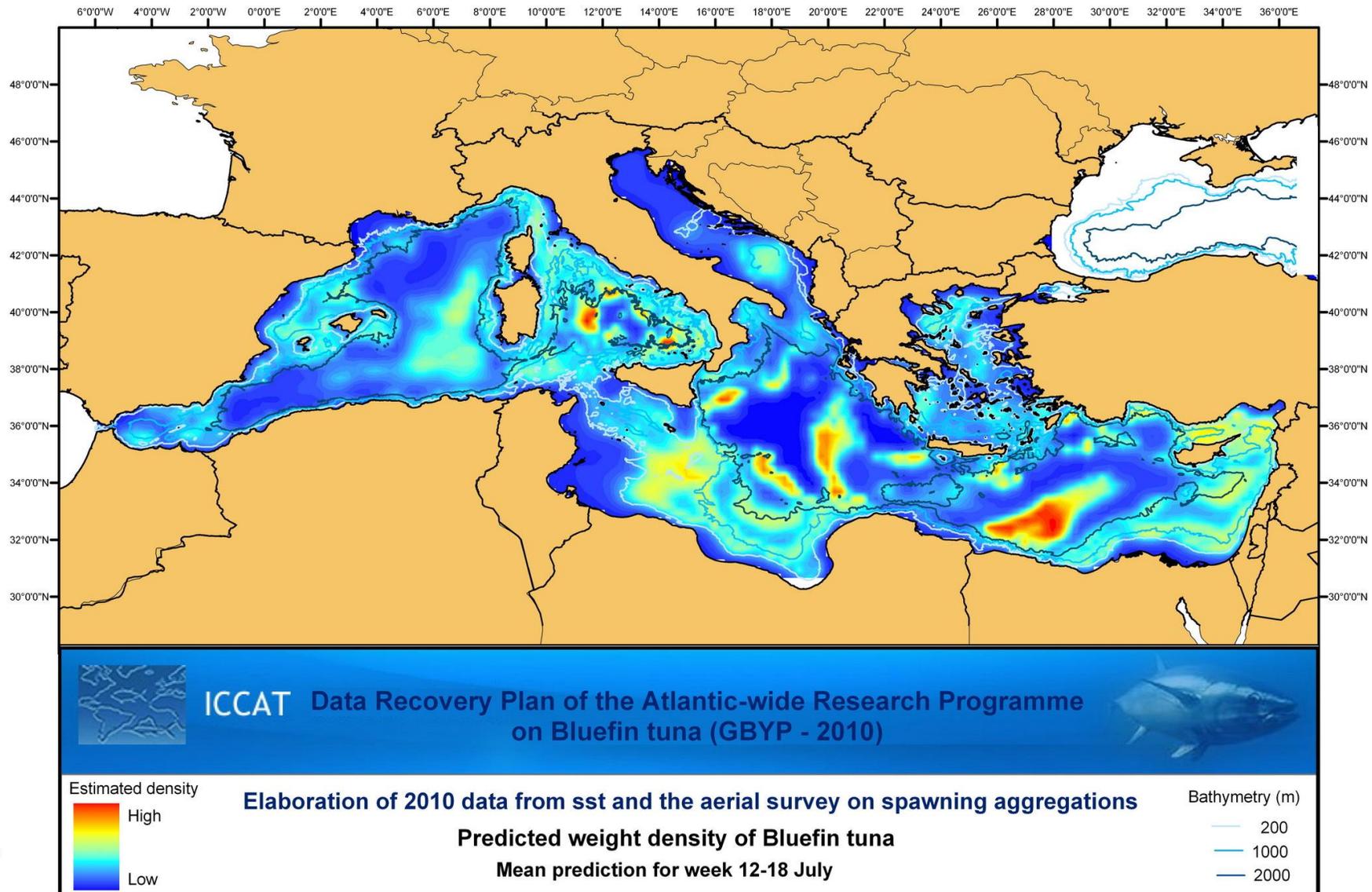


Figure 40. Predicted density of bluefin tuna for 12-18 July 2010

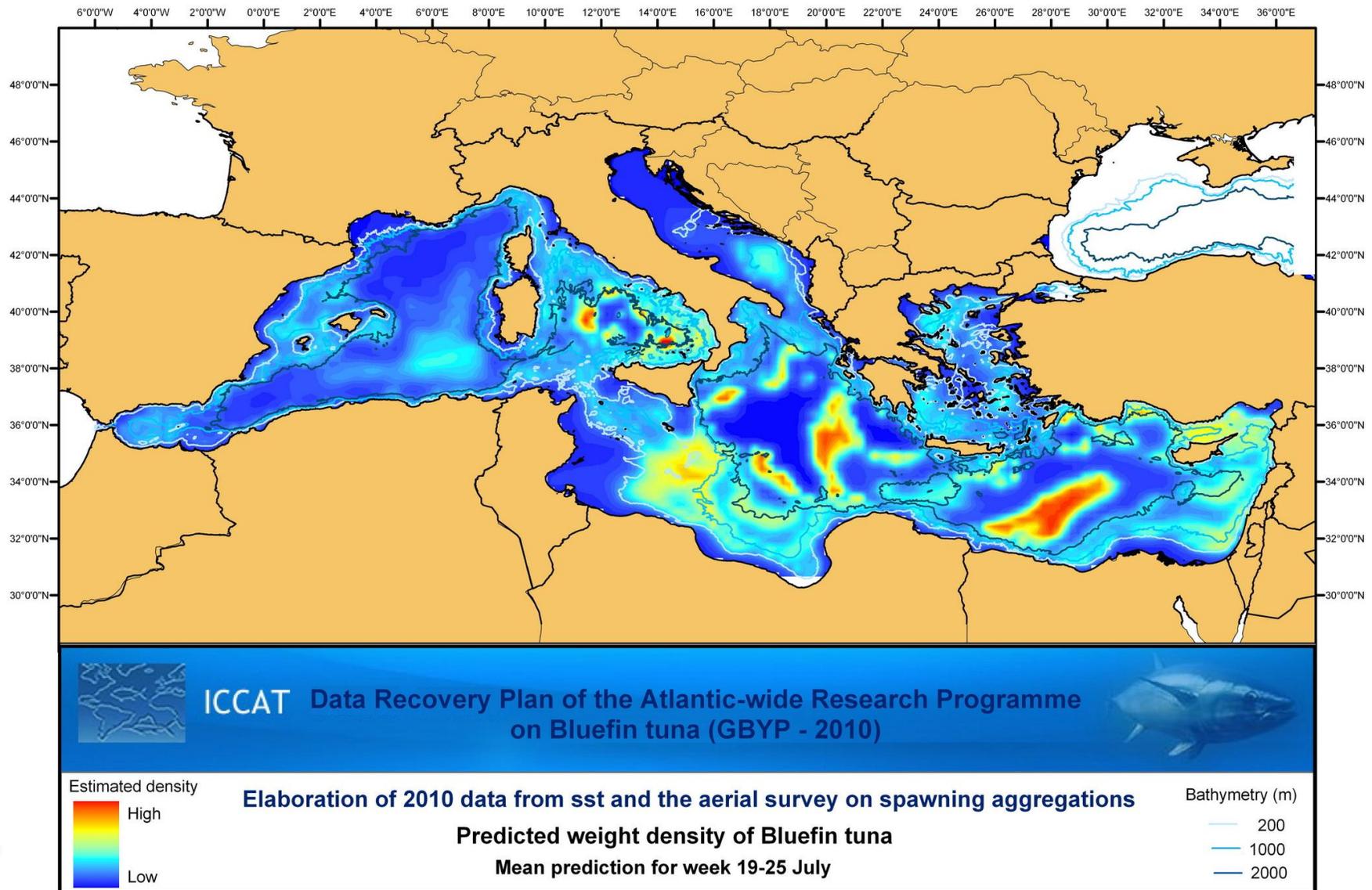


Figure 41. Predicted density of bluefin tuna for 19-25 July 2010

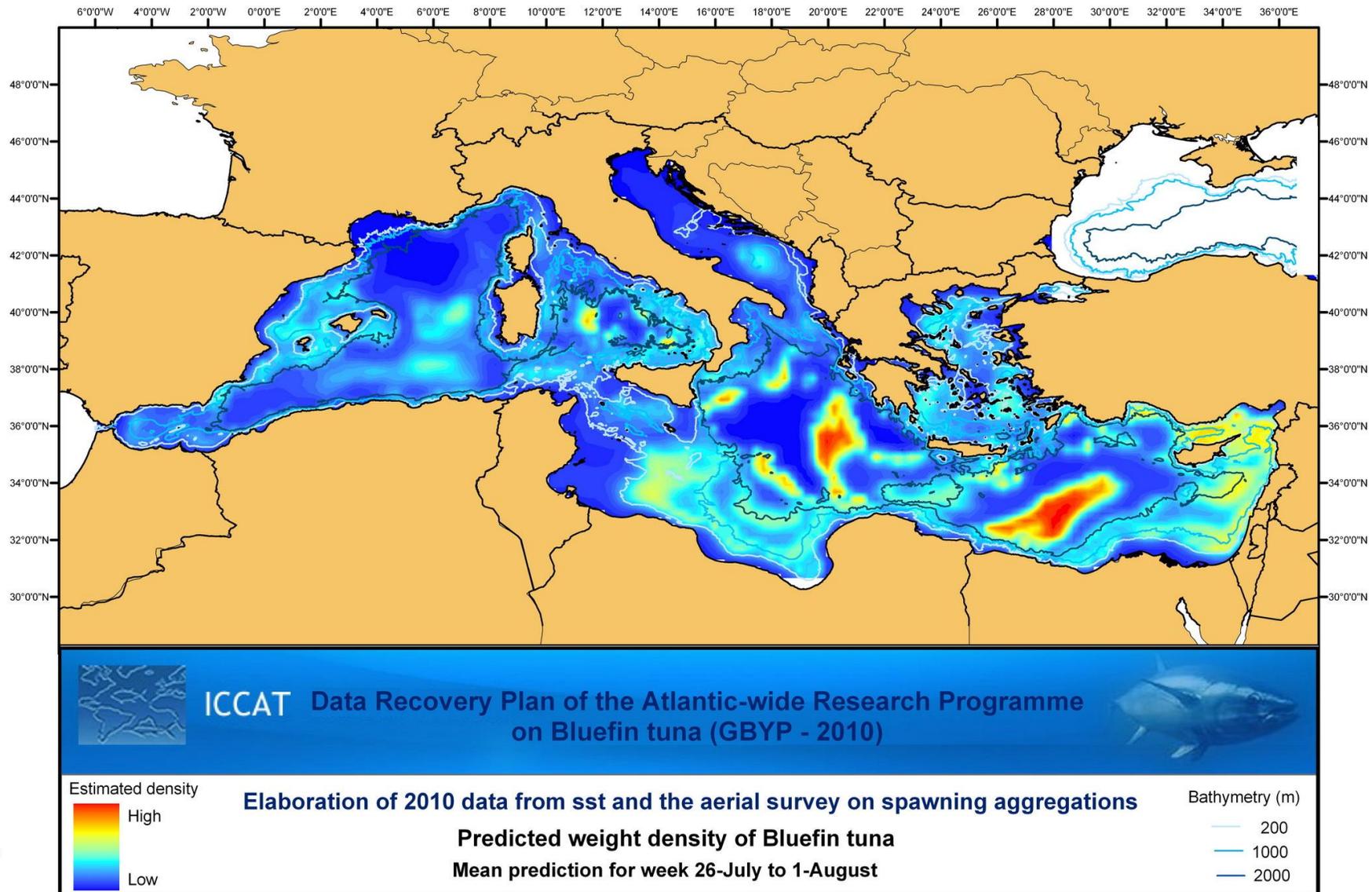


Figure 42. Predicted density of bluefin tuna for 26-July to 1-August 2010