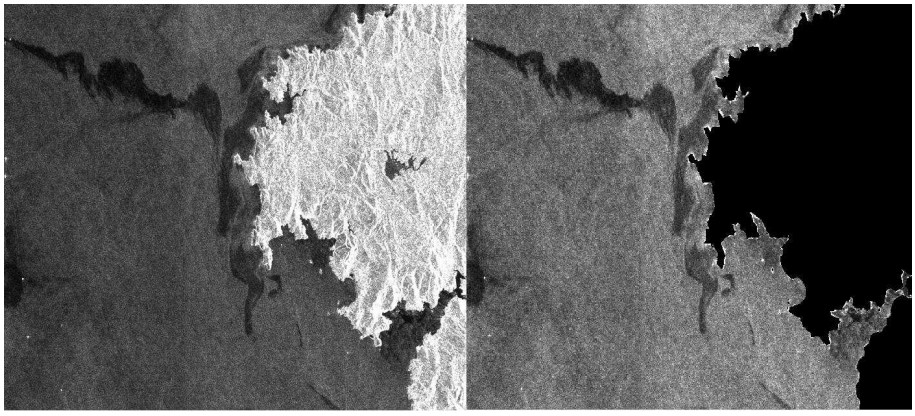

Figure 2.17: Zoom of the Prestige Envisat image, showing the effect of applying a land-mask based on the GSHHS coastline.

3. **Segmentation** We have applied one of the algorithms developed in this work, namely Algorithm-1, described in detail in Section 3.5. This algorithm has been chosen only due to practical reasons, as it is the fastest of the proposed segmentation methods.

4. **Feature Extraction** We have developed and run different matlab routines in order to calculate a number of features, both from the dark patch as also from a 150x150 pixel window from the surroundings. The complete list of extracted features is given in detail bellow, Section 2.4.2. The features have then been stored in the database, together with other complementary information.

5. **Classification** In the classification step, we have run all the classifiers described in detail in Section 4.2.

2.4.2 Extracted Features

A total of 35 features have been considered for classification. An attempt was made to extract as many features as possible from those referred in the state-of-the-art literature (see Section 1.2).
These features have all been extracted from the segmented dark patch, except for the wind information, which was overtaken from the ancillary data or from the operators reports. The used features can be divided in four main groups: geometrical, backscatter, texture and ancillary data features.

1. Geometrical Features:

   - Area (A)
   - Perimeter (P)
   - Complexity (C), defined as

     $$\mathrm{C} := \frac{\mathrm{P}}{2\sqrt{\pi \mathrm{A}}}.$$

     This feature will take a small numerical value for regions with simple geometry and larger values for complex geometrical regions.
   - Length (L): sum of skeleton edges (obtained by Delaunay triangulation), that build the main line.
   - Width (W): mean value of Delaunay triangles which are crossed by main line.
   - Length To Width Ratio (LWR), defined as

     $$\mathrm{LWR} := \frac{\mathrm{L}}{\mathrm{W}}.$$

   - Compactness (Comp), defined as

     $$\mathrm{Comp} := \frac{\mathrm{LW}}{\mathrm{A}}.$$

   - First Invariant Planar Moment (FIPM), defined as

     $$\mathrm{FIPM} := \frac{1}{n^2} \sum_{i=1}^{n} \left[(x_i - x_c)^2 + (y_i - y_c)^2 \right]$$

     with

     $$x_c := \frac{1}{n} \sum_{i=1}^{n} x_i \ , \ y_c = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

     and n the number of points in the dark patch contour.
   - Ellipse-Length (EL): value of main axe of an ellipse fitted to the data.
   - Ellipse-Width (EW): value of minor axe of an ellipse fitted to the data.
   - Ellipse-Asymetry (EA), defined as

     $$\mathrm{EA} := 1 - \frac{\mathrm{EW}}{\mathrm{EL}}.$$

- Form Factor (FF): represents the dispersion of the dark patch pixels from its longitudinal axis, when a line is fitted; it is calculated as norm of residuals after line fit to the dark patch pixels.
- Spreading (S): this feature is derived from the principal component analysis of the vectors whose components are the coordinates of the pixels belonging to the dark patch (see [24] for details). Feature S will be low for long and thin objects and high for objects closer to a circular shape.

2. Backscatter Features:

- Inside Slick Radar Backcatter (μ_{obj})
- Inside Slick Standard Deviation (σ_{obj})
- Outside Slick Radar Backscatter (μ_{sce})
- Outside Slick Standard Deviation (σ_{sce})
- Intensity Ratio ($\frac{\mu_{obj}}{\mu_{sce}}$)
- Intensity Standard Deviation Ratio ($\frac{\sigma_{obj}}{\sigma_{sce}}$)
- Intensity Standard Deviation Ratio Inside (ISRI), defined as

$$\text{ISRI} := \frac{\mu_{obj}}{\sigma_{obj}}.$$

- Intensity Standard Deviation Ratio Outside (ISRO), defined as

$$\text{ISRO} := \frac{\mu_{sce}}{\sigma_{sce}}.$$

- ISRI ISRO Ratio
- Min Slick Value (MinObj)
- Max Slick Value (MaxObj)
- Max Contrast (ConMax): difference (in dB) between the background mean value and the lowest value inside the object, defined as

$$\text{ConMax} := \mu_{sce} - \text{MinObj}.$$

- Mean Contrast (ConMe): difference (in dB) between the background mean value and the object mean value, defined as

$$\text{ConMe} := \mu_{sce} - \mu_{obj}.$$

- Max Gradient (GMax): maximum value (in dB) of border gradient magnitude, calculated using Sobel operator.
- Mean Gradient (GMe): mean border gradient magnitude (in dB).
- Gradient Standard Deviation (GSd): standard deviation (in dB) of the border gradient magnitudes.

3. Texture Features:

   Texture is a combination of repeated patterns with a regular frequency and texture analysis has often proved to be effective for oil spill classification (see [28], [20] and [34]). We have used gray level co-occurrence matrices (GLCM) to specify texture measures (see [63]) and have used the following measures:

   - GLCM Homogeneity
   - GLCM Contrast
   - GLCM Entropy
   - GLCM Correlation
   - GLCM Dissimilarity

4. Ancillary Data Features:

   - Wind Speed

A fifth group, of so-called context features, is also referred to in the literature, but was not used by us. This group includes features like:

- Distance from land
- Distance to large dark areas
- Distance to next bright spot (vessel, platform)
- Number of bright spots nearby
- Number of dark Objects in the vicinity

We note that many of the extracted features are related to the same slick characteristics and are correlated. For example the features "Spreading", "Compactness" and "LengthToWidthRatio", both describe how long and thin objects are. In the same way, the features "First Invariant Planar Moment", "Form Factor" and "Ellipse Assimetry" can all be considered to give an indication to the general shape of the object. In our work we wanted to examine as much of the features referred to in the literature as possible.

As a remark, it is important to say that all features have been standardized before being further used in the classification step. This is necessary because features can have different scales although they refer to comparable objects [64]. Consider for instance, a pattern $x = [x_1, x_2]$ where x_1 is a width measured in meters and x_2 is a height measured in centimeters. Both can be compared, added or subtracted but it would be unreasonable to do it before appropriate normalization. We have used the following classical centering and scaling procedure of the data : $x\,\mathrm{norm}_i = (x_i - \mu_i)/\sigma_i$, where μ_i and σ_i are the mean and the standard deviation of feature x_i over the training examples.