Chapter 11
Ongoing Refinement

The previous chapters of this book have built up models for the limited available measurement data and for aircraft motion, and then used these models to produce a pdf of the final aircraft latitude and longitude. In principle, one could use this pdf to direct a search and then the very act of searching would provide further measurement data. This chapter describes how the Bayesian method can be used to adapt the position pdf to account for data collected after the accident flight. Two data sources are discussed: the sonar imagery data collected in the search and the discovery of the flaperon on Reunion Island.

11.1 Updating the Distribution Using Search Results

The measurements collected as part of the search can be treated in the same mathematical framework as the communications data. This was used, for example, in the search for AF447 [40]. In this case the aircraft is no longer moving so the prediction stage becomes degenerate and the predicted pdf is the same as the previous posterior pdf. The result of the search can be summarised by the probability that the cumulative search effort would have detected the aircraft at any particular location, $P_D(\mathbf{x})$. The posterior pdf given the search effort (denoted $\mathcal{S}$) then becomes:

$$p(\mathbf{x}_{\mathsf{final}}|\mathcal{S}, \mathbf{Z}_K) \propto [1 - P_D(\mathbf{x}_{\mathsf{final}})]p(\mathbf{x}_{\mathsf{final}}|\mathbf{Z}_K), \tag{11.1}$$

where the constant of proportionality is determined to ensure that the posterior integrates to unity. If a particular area $\mathcal{A}$ is searched with a constant probability of detection P_D, the probability of finding the aircraft is

$$P(\mathsf{find\ during\ search\ of}\ \mathcal{A}) = P_D \int_{\mathcal{A}} p(\mathbf{x}_{\mathsf{final}}|\mathbf{Z}_K)\, \mathrm{d}\mathbf{x}_{\mathsf{final}}. \tag{11.2}$$

In the analysis used for prioritising the search for flight AF447, the probability of detection for areas searched using side-scan sonar was modelled as 0.9 [40]. This type of analysis could be used, for example, to prioritise whether to revisit a more likely area or to search a less likely area that has not yet been searched.

Based on the quality assurance process which has been implemented in the MH370 search, which includes revisiting items that have been assessed as potential debris, it is considered highly unlikely that the search would fail to detect the aircraft if the correct location is searched. Due to sensor drop-out and terrain masking, there will inevitably be small pockets which are not covered in a first pass of the search, and (11.1) can be used to determine the priority of returning to ensure that these are examined.

11.2 Reunion Island Debris

The apparent lack of debris from the aircraft was a mystery which was resolved in part when a flaperon was discovered washed up on Reunion Island on 29 July 2015, 508 days after the accident, and later confirmed to be from MH370. Two questions arise from this find:

1. What information does the discovery of the flaperon on Reunion Island provide about the final location of the aircraft?
2. What information does the lack of any other debris to date provide about the final location of the aircraft?

Debris information was used to inform the search for AF447 [40], and was proposed for MH370 in the white paper [19]. The most directly relevant source of information on the likely drift of the flaperon is from the Global Drifter Program [32]. The program provides 30 years of data on buoys referred to as drifters, which are regularly deployed in oceans worldwide. The primary goal of the program is to measure ocean currents rather than surface effects. To ensure that the drifter motion is dominated by the ocean current, the drifters are deployed with drogues, which are sea anchors that sit around fifteen metres beneath the ocean surface. If the drogue is detached then there is an observable change in the motion of the drifter [28]. A drifter without a drogue is referred to as an undrogued drifter and is more buoyant than the drifters with an attached drogue. The buoyancy of the flaperon has not been characterised at this time but it is expected that the undrogued drifters better describe its likely motion because the drogues by design cause drifters to move according to deeper currents.

In this section, we derive an updated pdf of the final location of the aircraft based on an analysis of data from the Global Drifter Program. It should be noted that the characteristics of the drifters are not accurately matched to the flaperon, thus there is some uncertainty in the applicability of the results, but it is believed to be the most relevant data source available. The authors gratefully acknowledge Dr. David Griffin (CSIRO, Australia) for providing specialist analysis and advice in support of

this effort, as well as the results illustrated in Fig. 11.1. Further information may be found at [17].

The first question is answered by modifying the pdf provided in the previous chapter. This is achieved by developing a likelihood function of the source location of the flaperon using the drifter data. The likelihood is used to calculate the posterior pdf of the final aircraft location given both the information provided by the Inmarsat satellite data and the discovery of the debris.

The second question relates to the potential information provided by the lack of other debris. We consider a statistical framework for quantifying this information but the framework requires parameters that cannot be reliably determined. For this reason we limit ourselves to qualitative conclusions about the lack of other debris.

11.2.1 Update of Final Location Distribution

We start with the form which results from the particle filter: (sampling to incorporate the effect of the descent kernel of Sect. 10.5)

$$p\left(\mathbf{x}_{\text{final}}|\mathbf{Z}_K\right) \approx \sum_{p=1}^{P} w^p \delta\left(\mathbf{x}_{\text{final}} - \mathbf{x}^p\right), \tag{11.3}$$

For a single item of debris, we seek a model which permits us to update (11.3) using the knowledge that the item arrived at a given location and time. For this, we require a transition distribution for how the debris would move over 508 days. Denoting the location after 508 days as $\mathbf{y}$, this transition distribution is $p(\mathbf{y}|\mathbf{x}_{\text{final}})$. Thus, assuming a single item of debris, the kinematic distribution updated by the knowledge of the debris find at location $\mathbf{y} = y$ is:

$$p(\mathbf{x}_{\text{final}}|\mathbf{Z}_K, \mathbf{y} = y) \propto p(\mathbf{y} = y|\mathbf{x}_{\text{final}})p(\mathbf{x}_{\text{final}}|\mathbf{Z}_K)$$

$$\approx \sum_{p=1}^{P} w^p p\left(\mathbf{y} = y|\mathbf{x}^p\right) \delta\left(\mathbf{x}_{\text{final}} - \mathbf{x}^p\right), \tag{11.4}$$

Equation (11.4) can be interpreted as re-weighting each particle $\mathbf{x}^p$ by the likelihood that an item of debris at that location would end up at location $\mathbf{y} = y$.

A field of debris can be modelled as a Poisson point process (e.g., [42]). Suppose that the expected number of debris items is λ, that an item that is washed up on the shore is identified with probability P_I, and that no items are found while they remain in the ocean. Then the likelihood of the first item of debris being found at $\mathbf{y} = y$ given that the end-of-flight location was $\mathbf{x}_{\text{final}}$ is:

$$l(\text{debris}|\mathbf{x}_{\text{final}}) = \exp\left\{-P_I\lambda \int_{\mathcal{C}} p(\mathbf{y}|\mathbf{x}_{\text{final}})d\mathbf{y}\right\}$$

$$\times \exp\left\{-\lambda p(\mathbf{y} = y|\mathbf{x}_{\text{final}})\right\} \lambda p(\mathbf{y} = y|\mathbf{x}_{\text{final}}) \tag{11.5}$$

where the integral is over the coastal region $\mathcal{C}$, i.e., the region where debris would be likely to have been found. This expression could be used to update the model as:

$$p(\mathbf{x}_{\text{final}}|\mathbf{Z}_K, \text{debris}) \propto l(\text{debris}|\mathbf{x}_{\text{final}})\, p(\mathbf{x}_{\text{final}}|\mathbf{Z}_K)$$

$$\approx \sum_{p=1}^{P} w^p l\left(\text{debris}|\mathbf{x}^p\right) \delta\left(\mathbf{x}_{\text{final}} - \mathbf{x}^p\right), \tag{11.6}$$

Similar to the case of a single debris item, this model re-weights each particle $\mathbf{x}^p$ with the likelihood that a debris field starting from that location would not yield any identified coastal debris within the time other than a single item at $\mathbf{y} = y$. The model could be easily refined to incorporate thinning, i.e., the likelihood that over time, items of debris might sink, as well as the temporal aspect that the debris did not arrive until some 508 days later, by splitting the transition distribution $p(\mathbf{y}|\mathbf{x})$ into a number of intermediate, smaller duration transition steps. However, the required model parameters, are very difficult to quantify with any degree of confidence. Specifically, the expected number of debris items, and the probability that an item washed up would be identified, are assumed to be known. Since these parameters cannot be adequately specified, we rely on a qualitative approach for incorporating information provided by the lack of debris, and use the quantitative approach of (11.4) for incorporating the information on the debris item that was found.

11.2.2   Data from Global Drifter Program

The data available for undrogued drifters are summarised in Fig. 11.1. Each diagram shows drifter trajectories which pass through a region of interest in months between February and April (i.e., the time of year of the accident). The middle figure centres the region of interest on the search zone, while the top figure examines the region to the North, and the bottom figure shows the region to the South. Coloured dots are drawn along each drifter trajectory every 100 days to illustrate time along the trajectory. Trajectory segments coloured in blue indicate water colder than $18\,^\circ$C, the temperature above which barnacle nauplii settlement and growth is accelerated; evidence of accelerated growth was present on the flaperon [17].

The following observations can be made from the figures:

1. In the top diagram (the region to the North of the search zone), many trajectories head West at quite a fast rate, reaching Madagascar and Tanzania in 300 days or less. The absence of debris being identified in the Western Indian Ocean many months earlier than July 2015 would tend to indicate this region as being less likely.
2. A significant proportion of the trajectories in the middle diagram (the search zone) arrive in the general vicinity of Reunion Island at around 500 days.
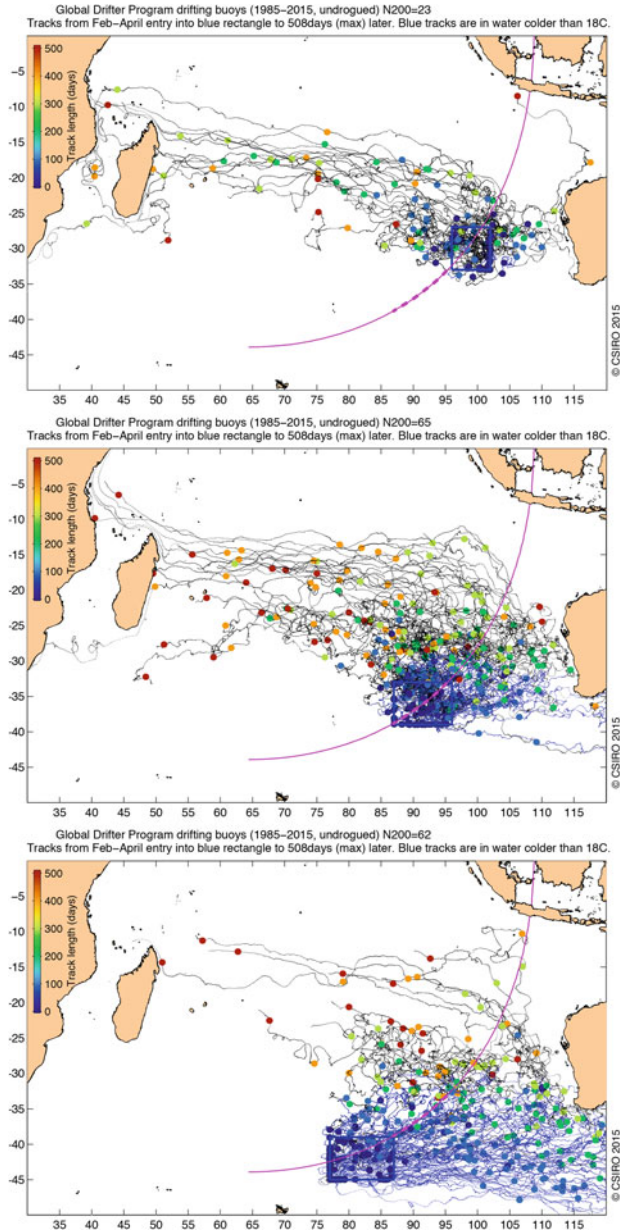
Fig. 11.1 Trajectories from global drifter program. *Top* diagram shows trajectories which pass through the box with latitudes from 27°S to 33°S and longitudes from 96°E to 106°E. *Middle* diagram shows trajectories which pass through the box from 33°S to 39°S and longitudes from 87°E to 96°E. *Bottom* diagram shows trajectories which pass through the box from 39°S to 45°S and longitudes from 77°E to 87°E. *Coloured dots* show time, marking every 100 days. Blue trajectories are in water colder than 18°C, the temperature above which barnacle nauplii settlement and growth is accelerated. Figure courtesy of Dr. David Griffin, CSIRO; see also [17]

3. The majority of trajectories in the bottom diagram (the region to the South of the search zone) head to the East, towards Australia and New Zealand. Very few head toward the vicinity of Reunion Island.

Thus the discovery of the single flaperon on Reunion Island after 508 days would seem to be more consistent with the existing search region, and less consistent with regions significantly to the North or South.

11.2.3 Posterior Distribution Using Debris Data

Figure 11.1 represents the total undrogued drifter data available passing through each box at the relevant time of the year over the 30 year history of the Global Drifter Program. It is insufficient to construct a transition distribution, hence additional processing was performed to enhance the data set. Specifically, pairs of trajectories that passed close together (possibly in different years) were "joined," i.e., used to create additional synthetic trajectories comprised of the head of one with the tail of the other. A synthetic data set was generated by joining four trajectory segments. In each joining process, each previous trajectory creates around 30 new trajectories.

The joined trajectories provide a sampled representation of the distribution:

$$p_{\text{drifter}}(\mathbf{x}, \mathbf{y}) \approx \frac{1}{P} \sum_{p=1}^{P} \delta(\mathbf{x} - \mathbf{x}^p)\delta(\mathbf{y} - \mathbf{y}^p) \tag{11.7}$$

where $\mathbf{x}$ is the location of the drifter on 8 March 2014, and $\mathbf{y}$ is the location on 29 July 2015. As in (11.4), the distribution of interest is $p(\mathbf{y}|\mathbf{x}_{\text{final}})$ evaluated for $\mathbf{y}$ at Reunion Island, or, pragmatically, integrating over a small region $\mathcal{R}$ surrounding Reunion Island:

$$l(\mathbf{x}_{\text{final}}) = \int_{\mathcal{R}} p(\mathbf{y}|\mathbf{x}_{\text{final}})\mathrm{d}\mathbf{y} \tag{11.8}$$

The conditional distribution $p(\mathbf{y}|\mathbf{x})$ requires normalisation by the prior distribution $p_{\text{drifter}}(\mathbf{x})$; this represents the overall mean density of drifters in the ocean, which is neither spatially uniform nor time invariant. Applying a small kernel $K(\mathbf{x})$ around each sample, the likelihood is evaluated as:

$$l(\mathbf{x}_{\text{final}}) = \frac{\int_{\mathcal{R}} p_{\text{drifter}}(\mathbf{x}_{\text{final}}, \mathbf{y})\mathrm{d}\mathbf{y}}{p_{\text{drifter}}(\mathbf{x}_{\text{final}})} \approx \frac{\frac{1}{P} \sum_{p=1}^{P} K(\mathbf{x}_{\text{final}} - \mathbf{x}^p) + \varepsilon}{\frac{1}{\tilde{P}} \sum_{p=1}^{\tilde{P}} K(\mathbf{x}_{\text{final}} - \tilde{\mathbf{x}}^p) + \varepsilon} \tag{11.9}$$

where $\{\mathbf{x}^1, \ldots, \mathbf{x}^P\}$ is the set of locations for trajectories which pass though the region $\mathcal{R}$ in the right time window, including those synthetically augmented using the aforementioned joining process. $\{\mathbf{x}^1, \ldots, \tilde{\mathbf{x}}^{\tilde{P}}\}$ is the overall set of drifter locations for the time interval from February to April. The ε terms are added for regularisation,
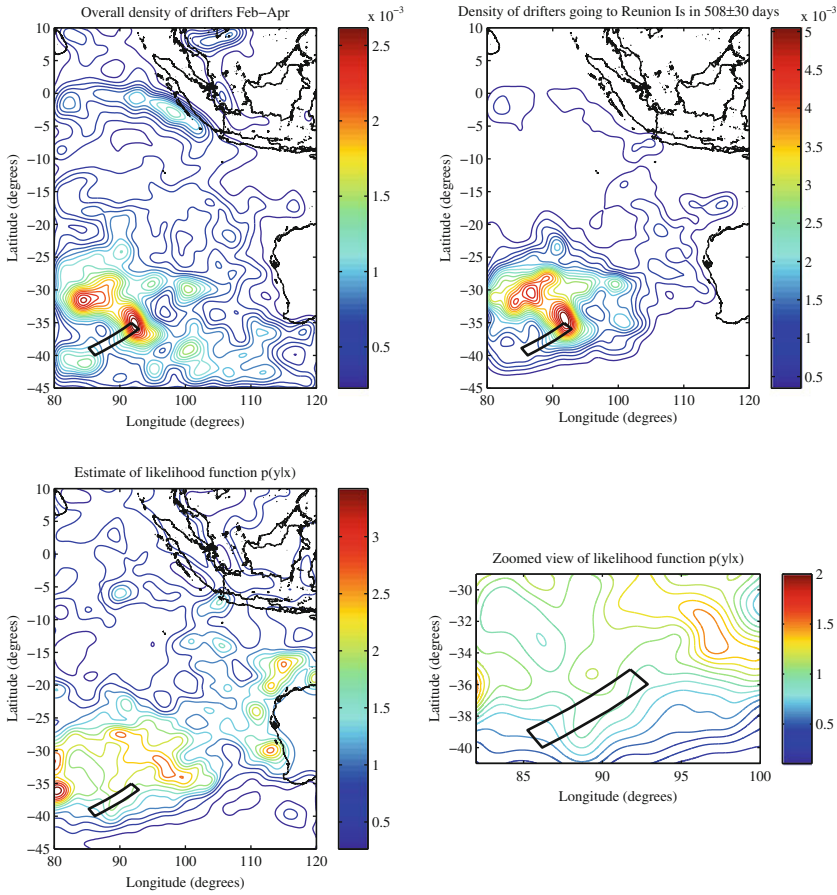
Fig. 11.2 Distributions obtained using undrogued drifter data from Global Drifter Program. *Top-left* diagram shows the overall density of drifters in the months of interest (over all years). *Top-right* shows distribution of synthetically augmented trajectories which pass though the Reunion Island region after 508 days. *Lower-left* diagram shows the quotient of these two distributions, which is the likelihood $p(\mathbf{y}|\mathbf{x}_{\text{final}})$. *Lower-right* diagram shows this same likelihood, zoomed in to the region of interest for the posterior distribution. An indicative search area is marked as a *black rectangle*. Data courtesy of Dr. David Griffin, CSIRO; see also [17]

i.e., preventing very large values in the conditional resulting from the finite sample support of the denominator (we set $\varepsilon = 10^{-4}$).

The result of this analysis is shown in Fig. 11.2, using a fixed kernel of standard deviation $1°$ in latitude and longitude, and evaluating the density estimates using the KDE toolbox [22]. The top left diagram shows the overall spatial distribution of drifters; the indicative search area is again shown in black. The overall distribution clearly has significant spatial variation, particularly in the search area. The density of drifters that pass though the Reunion Island region after the right duration is shown
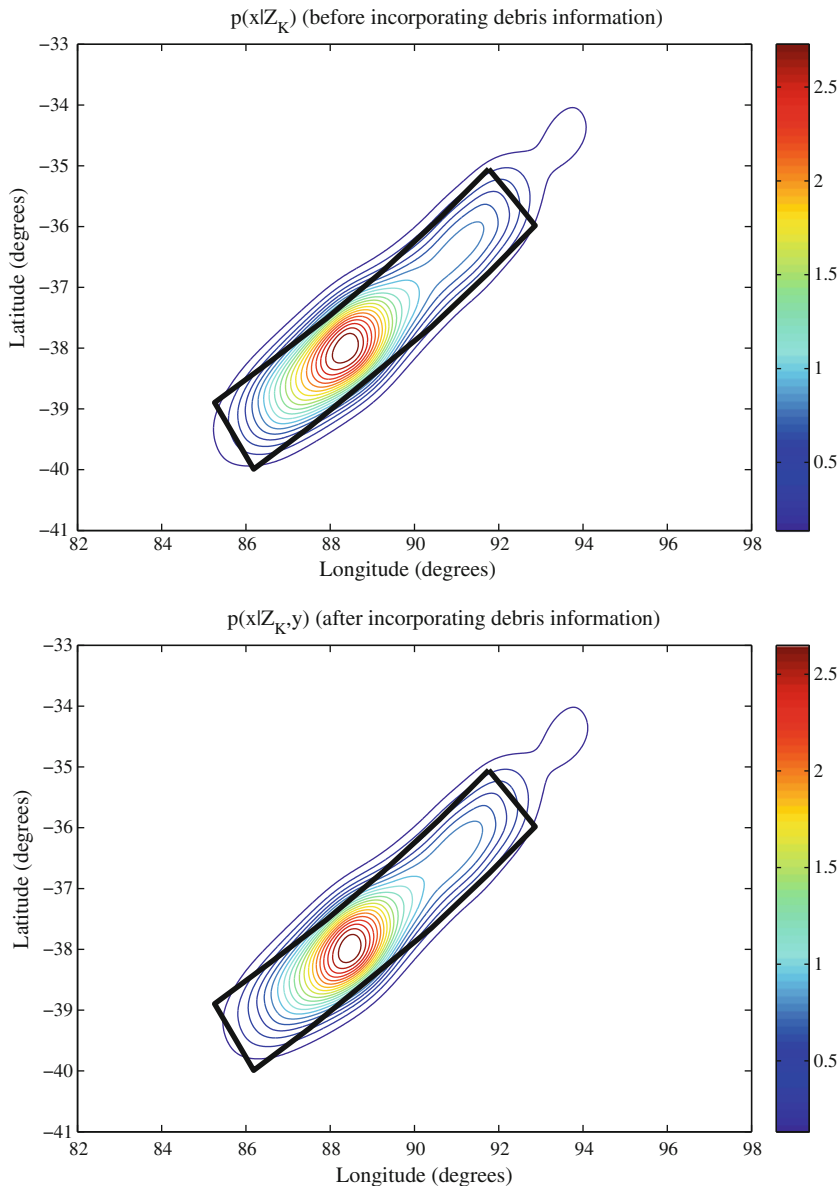
p(x|Z_K) (before incorporating debris information)

p(x|Z_K,y) (after incorporating debris information)

Fig. 11.3 Posterior distribution based on Inmarsat data (*top*), and incorporating debris discovery at Reunion Island (*bottom*)

in the top right. It exhibits spatial variations very similar to the overall variations in the top left. The likelihood function $p(\mathbf{y}|\mathbf{x}_{final})$ is the ratio of these two densities and is shown in the lower left. Note that large values can occur in areas of limited

sample support due to noise. The variation in the distribution of drifters that arrived at Reunion Island is almost completely explained by the variation in the overall starting distribution, therefore the likelihood shows much less spatial variation. The lower right diagram shows a zoomed in region around the search area.

Figure 11.3 shows the posterior distribution based on the Inmarsat data from Fig. 10.10, alongside the updated distribution, which incorporates the information from the debris item located at Reunion Island. The updated distribution is shifted very slightly to the North, but the effect is negligible. This result is not unexpected given the long duration between the accident and the debris discovery.