In the early experiments, Richardson (1922) and Charney *et al.* (1950) performed hand interpolations of the available observations to grid points, and these fields of initial conditions were manually digitized, which was a very time consuming procedure. The need for an automatic "objective analysis" quickly became apparent (Charney, 1951), and interpolation methods fitting data to grids were developed (e.g., Panofsky, 1949, Gilchrist and Cressman, 1954, Barnes, 1964, 1978). However, there is an even more important problem than spatial interpolation of observations to gridded fields: the data available are not enough to initialize current models. Modern primitive equations models have a number of degrees of freedom of the order of 10^7. For example, a latitude–longitude model with a typical resolution of 1° and 20 vertical levels would have $360 \times 180 \times 20 = 1.3 \times 10^6$ grid points. At each grid point we have to carry the values of at least four prognostic variables (two horizontal wind components, temperature, moisture), and the surface pressure for each column, giving over 5 million variables that need to be given an initial value. For any given time window of ± 3 hours, there are typically 10–100 thousand observations of the atmosphere, two orders of magnitude less than the number of degrees of freedom of the model. Moreover, their distribution in space and time is very nonuniform (Fig. 1.4.1), with regions like North America and Eurasia which are relatively data-rich, while others much more poorly observed.

For this reason, it became obvious rather early that it was necessary to use additional information (denoted *background*, *first guess* or *prior information*) to prepare initial conditions for the forecasts (Bergthorsson and Döös, 1955). Initially climatology was used as a first guess (e.g., Gandin, 1963), but as the forecasts became better, a short-range forecast was chosen as the first guess in the operational data assimilation systems or "analysis cycles". The intermittent data assimilation cycle shown schematically in Fig. 1.4.2 is continued in present-day operational systems, which typically use a 6-h cycle performed four times a day.

In the 6-h data assimilation cycle for a global model, the background field is a model 6-h forecast x^b (a three-dimensional array). To obtain the background or first guess "observations", the model forecast is interpolated to the observation location, and if they are different, converted from model variables to observed variables y^o (such as satellite radiances or radar reflectivities). The first guess of the observations is therefore $H(x^b)$, where H is the observation operator that performs the necessary interpolation and transformation from model variables to observation space. The difference between the observations and the model first guess $y^o - H(x^b)$ is denoted "observational increments" or "innovations". The analysis x^a is obtained by

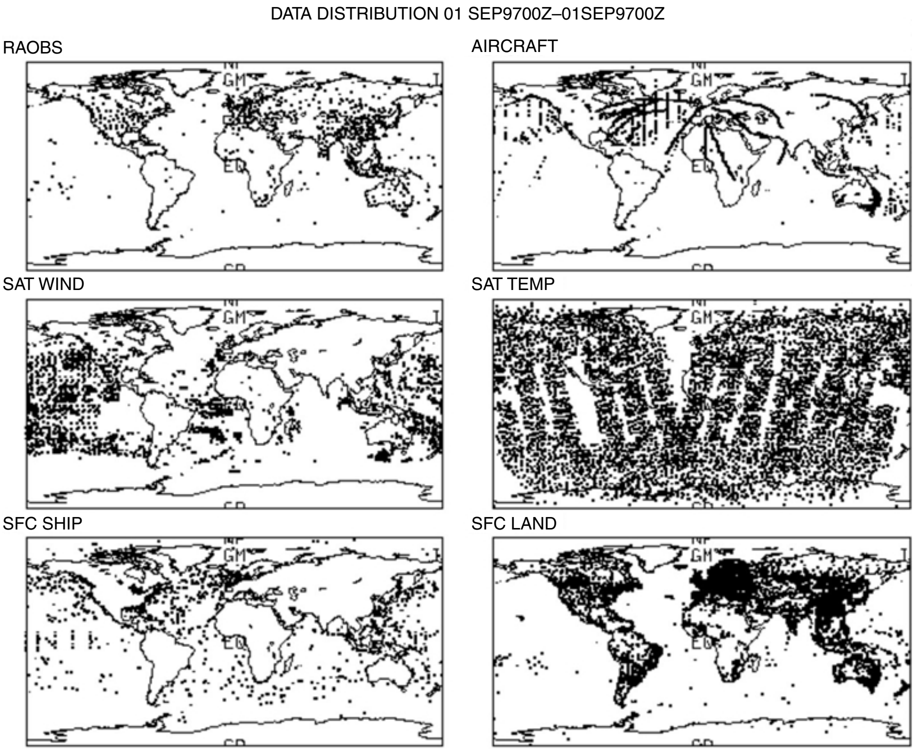

Figure 1.4.1: Typical distribution of observations in a ±3-h window.

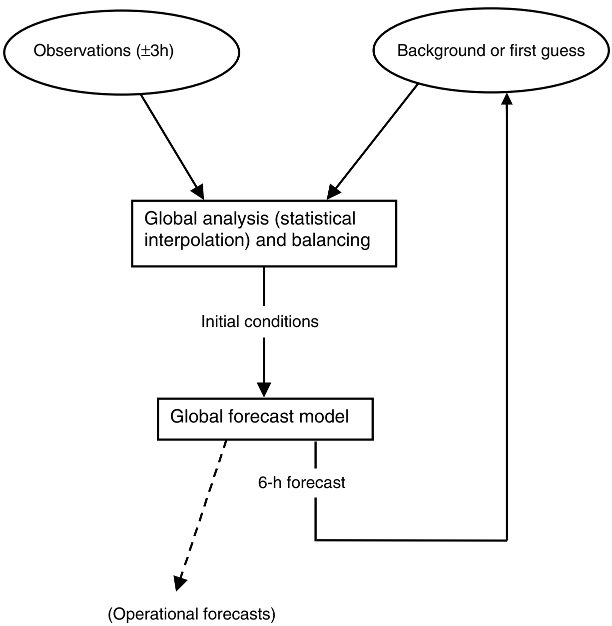

Figure 1.4.2: Flow diagram of a typical intermittent (6-h) data assimilation cycle.

adding the innovations to the model forecast (first guess) with weights W that are determined based on the estimated statistical error covariances of the forecast and the observations:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{W}[\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)] \tag{1.4.1}$$

Different analysis schemes (SCM, OI, 3D-Var, and KF) are based on (1.4.1) but differ by the approach taken to combine the background and the observations to produce the analysis. Earlier methods such as the SCM (Bergthorsson and Döös, 1955, Cressman, 1959, Barnes, 1964) were of a form similar to (1.4.1), with weights determined empirically. The weights are a function of the distance between the observation and the grid point, and the analysis is iterated several times. In OI (Gandin, 1963) the matrix of weights W is determined from the minimization of the analysis errors at each grid point. In the 3D-Var approach one defines a cost function proportional to the square of the distance between the analysis and both the background and the observations (Sasaki, 1970). The cost function is minimized directly to obtain the analysis. Lorenc (1986) showed that OI and the 3D-Var approach are equivalent if the cost function is defined as:

$$J = \frac{1}{2}\{[\mathbf{y}^o - H(\mathbf{x})]^T R^{-1}[\mathbf{y}^o - H(\mathbf{x})] + (\mathbf{x} - \mathbf{x}^b)^T B^{-1}(\mathbf{x} - \mathbf{x}^b)\} \tag{1.4.2}$$

The cost function J in (1.4.2) measures the distance of a field x to the observations (the first term in the cost function) and the distance to the first guess or background x^b (the second term in the cost function). The distances are scaled by the observation error covariance R and by the background error covariance B respectively. The minimum of the cost function is obtained for $x = x^a$, which is defined as the "analysis". The analysis obtained in (1.4.1) and (1.4.2) is the same if the weight matrix in (1.4.1) is given by

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}^{-1})^{-1} \tag{1.4.3}$$

The difference between OI (1.4.1) and the 3D-Var approach (1.3) is in the method of solution: in OI, the weights W are obtained for each grid point or grid volume, using suitable simplifications. In 3D-Var, the minimization of (1.4.2) is performed directly, allowing for additional flexibility and a simultaneous global use of the data (Chapter 5).