International Tables for Crystallography (2012). Vol. F, Section 11.4.8.2, pp. 290–291.

thus $|\boldsymbol{\omega}|/|\mathbf{S} \times \boldsymbol{\omega}'|$ is the Lorentz factor); P is a polarization factor (Azaroff, 1955); T is the transmission of the beam [related to the absorbance defined as $A = -\ln T$, $T = \exp(-A)$]; v_u is the volume of a primitive crystal unit cell; V is the volume of the crystal exposed to the beam; $|\mathbf{F}(hkl)|^2$ is the square of the structure-factor amplitude for the given reflection hkl; D_A is the absorption of X-rays by the detector's active material; and D_C is the detector's response to a single absorbed X-ray photon.

SCALEPACK determines the components of the total scale factor, *i.e.* the product of all factors that multiply the structure-factor amplitudes squared in equation (11.4.8.1):

$$I(hkl) = K|\mathbf{F}(hkl)|^2, \qquad (11.4.8.2)$$

where I is the intensity and K is the total scale factor.

All components of the total scale factor K can be calculated from non-diffraction measurements and calibration of the data-collection system. However, the absolute calibration of the whole system is rarely available and part of the scale factor (the overall scale factor k_o) is determined by comparing the scaled data to the squared structure-factor amplitudes predicted from an atomic model obtained after the structure is solved:

$$K = k_o k_r = k_o(k_{\rm beam}\, k_{\rm polarization}\, k_{\rm detector} \ldots). \qquad (11.4.8.3)$$

The relative scale factor k_r is calculated, but in practice we assume that some of its components are known from detector calibration, beam monitoring and the diffraction geometry. The scaling procedure determines the remaining parts of the scale factor (often iteratively), based on knowledge from subsequent stages of crystallographic analysis. When iterative procedures are applied, the scaling model has to include information about experimental uncertainties (discussed in Section 11.4.10).

SCALEPACK uses an exponential modelling approach (Otwinowski *et al.*, 2003), which is flexible with regard to correlations among parameters and provides a uniform description of the parameter optimization process for various scaling models. In *SCALEPACK* scale factors s_i for each observation are calculated using a set of *a priori* unknown parameters p_i:

$$s_i = \exp\left[\sum_i p_i\, f_i(hkl)\right], \qquad (11.4.8.4)$$

where f_i are pre-defined modelling functions of experimental conditions and i is a hierarchical index referring both to the type of correction and to the indices of the functional parameters describing this correction.

The simplest scaling model has a separate scale factor for each group (or batch) of data, *e.g.* one scale factor per image. In such a case,

$$f_{i_s} = \delta_{ij}, \qquad (11.4.8.5)$$

where j is the batch index for a particular reflection. From equation (11.4.8.4) we obtain P_{i_s} as the logarithm of the scale factor of batch i.

To correct for a radiation-damage component represented as resolution-dependent decay, one temperature factor is used per batch of data,

$$f_{i_B} = [(\mathbf{S} \cdot \mathbf{S})/2]\delta_{ij}, \qquad (11.4.8.6)$$

where $\mathbf{S}$ is the scattering vector for each reflection. Using the same approach as for equation (11.4.8.5), p_{i_B} is obtained as the relative temperature factor.

Another, more complex, multiplicative correction addresses unknown crystal absorption using an average of absorbances in the incoming and diffracted beam directions (Kopfmann & Huber, 1968). The correction is parameterized by real spherical harmonics as a function of the direction of the incoming beam and the diffraction vector $\mathbf{S}$, expressed in the polar coordinate system of the rotating crystal (Katayama, 1986; Blessing, 1995):

$$\begin{aligned} f_{as,lm} &= \frac{1}{2}\left[\frac{(2l+1)(l-m)!}{4\pi(l+m)!}\right]^{1/2} [P_{lm}(\cos\theta_i)\sin(2\pi m\Phi_i) \\ &\quad + P_{lm}(\cos\theta_o)\sin(2\pi m\Phi_o)], \\ f_{ac,lm} &= \frac{1}{2}\left[\frac{(2l+1)(l-m)!}{4\pi(l+m)!}\right]^{1/2} [P_{lm}(\cos\theta_i)\cos(2\pi m\Phi_i) \\ &\quad + P_{lm}(\cos\theta_o)\cos(2\pi m\Phi_o)], \end{aligned} \qquad (11.4.8.7)$$

where as, lm and ac, lm are parts of the hierarchical index i from equation (11.4.8.4); l, m are indices of the spherical harmonics; P_{lm} is a Legendre polynomial; and θ, Φ are the polar coordinates of the incoming (index i) and the outgoing (index o) directions in the crystal coordinate system. The odd-order spherical harmonics should have zero coefficients when describing pure absorption of X-rays. However, due to correlation with other effects, the scaling may benefit from the inclusion of low-order odd-order harmonics.

It is beneficial to correct even for a small discrepancy between the actual and assumed directions of the crystal rotation axis. The inaccuracy results in an error in the calculated value of the Lorentz factor [equation (11.4.8.1)]. The scale factor to correct for this error can be described using the parameter p_l, the value of which represents a small angular error. The corresponding function is

$$f_l = \frac{\mathbf{S} \cdot \boldsymbol{\omega}'}{|\mathbf{S} \times \boldsymbol{\omega}'|}. \qquad (11.4.8.8)$$

There are additional effects that are parameterized using exponential modelling [equation (11.4.8.4)], for example, a correction for uneven crystal rotation and/or exposure (Otwinowski *et al.*, 2003), or a correction for uneven detector response, which is very important for multi-CCD detectors. This approach can be extended to other experimental factors if there is a need to correct them during scaling without changing the overall logic of scale-factor determination (Otwinowski *et al.*, 2003).

11.4.8.1. *Global and local scaling*

In principle, global scaling can be followed by local scaling (Matthews & Czerwinski, 1975). Local scaling is mostly applied to calculate differences of phasing signal, where it is assumed that a group of measurements, *e.g.* those close together in reciprocal space or detector space, should be on a similar scale. A flexible parameterization by the exponential modelling allows for a good description of all kinds of smooth corrections. Local scaling is much more limited in terms of what type of smooth variation is being corrected for, so it is unlikely to provide additional benefit to the general scaling method described here. In practice, if there is an improvement from such procedures at the stage of heavy-atom search, it implies that the scaling parameters for global scaling were not properly chosen.

11.4.8.2. *Stabilization of scaling parameters based on prior knowledge*

The unknown parameters in equation (11.4.8.4) are estimated with various level of uncertainty depending on the multiplicity of observations and how symmetry-equivalent reflections are related to each other. Potentially, this may result in unreasonable values of scaling parameters due to insufficient information to determine the values of parameters. In *SCALEPACK*, the method to stabilize such ill-conditioned calculations is closely related to Tikhonov stabilization (Tikhonov & Arsenin, 1977), where additional, *a priori* knowledge about the expected magnitude of the physical effect modelled is used to restrain the solutions, based on the same argument as in the case of restraints in the atomic refinement.

For example, logarithms of scale factors typically do not fluctuate by more than w_s between frames, where expectation about w_s is a function of the data-collection stability (beam stability, goniostat and/or crystal vibrations). This knowledge is described by adding a penalty term (*scale restrain*) to the functions being optimized:

$$(1/w_s^2)(p_{i_s} - p_{(i+1)_s})^2. \qquad (11.4.8.9)$$

A similar approach can be used in calculations of absorption coefficients. For smooth absorption with the expectation of decreasing magnitude of parameters for high orders of spherical harmonics [equation (11.4.8.7)], a reasonable restraint term parameterized by w_a results in

$$\frac{l^2(p_{as,lm}^2 + p_{ac,lm}^2)}{w_a^2}. \qquad (11.4.8.10)$$

If we do not want to penalize high-order terms more than the low-order ones, the following restraint can be used:

$$\frac{(p_{as,lm}^2 + p_{ac,lm}^2)}{w_a^2}. \qquad (11.4.8.11)$$

11.4.9. Global refinement or post refinement

The process of refining crystal parameters using the combined reflection intensity measurements is known as global refinement or post refinement (Rossmann, 1979; Evans, 1993). The implementation of this method in *SCALEPACK* allows for separate refinement of the orientation of each image, but with the same unit-cell value for the whole data set. In each batch of data (a batch is typically one image), different unit-cell parameters may be poorly determined. However, in a typical data set there are enough orientations to determine all unit-cell lengths and angles precisely. Global refinement is also more precise than the processing of a single image in the determination of crystal mosaicity and the orientation of each image.

11.4.10. Merging – assessment of the error model and signal magnitudes in the data

Proper error estimation requires the use of Bayesian reasoning and a multi-component error model (Schwarzenbach *et al.*, 1989; Evans, 1993). In principle, the error estimates may be derived solely from a theoretical understanding of the measurement process. However, the complexity of error propagation and correlations between various sources of effects have led crystallographers to rely on hybrid approaches also involving self-consistency analysis of symmetry-equivalent reflections.

11.4.10.1. Estimation of random errors

The random errors in *DENZO* are estimated by a heuristic procedure that also accounts for small components of systematic errors (Borek *et al.*, 2003). Initially, *DENZO* estimated errors of integrated diffraction peaks from X-ray film. After introducing detectors with larger dynamic range, the procedure was adjusted accordingly.

The initial estimates of errors are obtained by

$$\sigma_0 = \frac{1}{\sum_i P_i^2/(B_i + P_iI)} \times \left\{ e_d \left[\sum_i P_i^2(B_i + P_iI) + \frac{e_d}{n_b} \sum_i \frac{P_i^2 B_i}{(B_i + P_iI)^2} \right] \right\}^{1/2}, \qquad (11.4.10.1)$$

where n_b is the number of pixels used in background estimation and e_d is the error-density parameter defined for each instrument, which can also be overridden by the user (Gewirth, 2003) with other variables defined in equation (11.4.7.1). The sums are calculated over all the pixels in a reflection profile. The expression within the braces { } describes two components of uncertainty: the left sum accounts for contributions resulting from pixels in the peak area, whereas the right sum adds an adjustment resulting from uncertainty of the background estimate. The denominator in the front of the expression in braces is derived from error propagation for the profile-fitted intensity [equation (11.4.7.3)].

Next, the goodness-of-profile-fitting factor g is calculated:

$$g = \left[\frac{1}{(n_i - 1)} \sum_i \frac{(M_i - B_i - P_iI)^2}{e_d(B_i + P_iI)} \right]^{1/2}, \qquad (11.4.10.2)$$

where n_i is the number of pixels in a reflection profile. For weak reflections the parameter g should be relatively close to 1. If it is systematically off by a large factor, the error-density parameter e_d should be adjusted (Borek *et al.*, 2003). *SCALEPACK* applies an additional level of adjustment to the estimates produced by *DENZO* (Borek *et al.*, 2003):

$$\sigma_S = 1.2\sigma_o g^{1/2}, \qquad (11.4.10.3)$$

which is scaled either by the user or by an automatically adjustable factor E_S (called the error scale factor) to make disagreements among symmetry-related measurements consistent:

$$\sigma_I = E_S \sigma_S. \qquad (11.4.10.4)$$

Even this scaled estimate of random error σ_I does not account for all types of errors and additional adjustments for systematic effects are needed.

11.4.10.2. Estimation of multiplicative errors

The multiplicative scale factor has its own uncertainty independent of random errors with typical values in the range of a few per cent. However, even such small errors are important in calculations of the phase signal. Errors in the scale factors have a correlated component that equally affects measurements of intensities in phasing differences, so it does not impact on the differences themselves. The important part is estimating the magnitude of the remaining component of scaling errors, described by σ_K. Comparing symmetry-related reflections estimates only the relevant component of multiplicative errors. The total scaling error would have to be estimated differently, but typically it has little relevance to macromolecular crystallography and can be ignored.

The σ_I [equation (11.4.10.4)] can be combined with σ_K to obtain the final estimated error of the scaled measurement:

$$\sigma_E = (1/K)(\sigma_I^2 + I^2\sigma_K^2)^{1/2}. \qquad (11.4.10.5)$$

references