correlation is weaker, which is expressed in the weighted-difference (WD) vector (Eq. 2). The larger the number of proxy maps used for downscaling emissions from a particular sector, the stronger the correlation generally is. This is due to the damping effect on outliers. This results in the following set of equations:

$$\mathrm{STD}_{c,m,g} = f_{c,m} \cdot E_{m,g} \cdot \sigma_m, \tag{1}$$

$$\mathrm{WD}_{c,m} = \left| \frac{\mathrm{STD}_{c,m,\mathrm{CO_2}}}{\sum(W_{c,m,\mathrm{CO_2}})/n} - \frac{\mathrm{STD}_{c,m,\mathrm{CO}}}{\sum(W_{c,m,\mathrm{CO}})/n} \right|, \tag{2}$$

$$\mathrm{PC}_c = \frac{\mathrm{stdev}(\mathrm{WD}_{c,m})}{n} \text{ and } \mathrm{PF}_c = \frac{\max(\mathrm{WD}_{c,m})}{n}, \tag{3}$$

where g is one of the two trace gases (CO or CO_2), f is the fraction of a proxy map in a grid cell, and n is the number of proxy maps contributing to a grid cell. Note that the predictor is slightly different for GNFR C and GNFR F, being based on the standard deviation and maximum value, respectively, of the WD values of the proxy maps contributing to each grid cell. We define a relationship between the predictor and the Monte Carlo-based correlation coefficient to calculate the $CO : CO_2$ error correlation per grid cell based on the predictor. For the Monte Carlo method, an ensemble of gridded emissions was produced by randomly perturbing the grid cell emissions of CO_2 and CO for each proxy map, following the defined uncertainty ranges. The perturbations are applied equally to CO_2 and CO emissions, assuming a full error correlation for each proxy map, which is a valid assumption at the grid cell level. The correlation coefficient results from a linear regression of the total CO_2 and CO emissions per grid cell for a given GNFR sector. The results are shown in Sect. 3.1 and the Supplement.

For the other sectors, we estimate a fixed value for all grid cells. For the public-power and shipping sectors, the correlation is likely to be very strong since there is little variation in sub-sector activities. Therefore, we set the error correlation to 0.95. For industry, the correlation is much smaller due to different sub-processes taking place, and we set the error correlation to 0.5.

2.2.3 Uncertainty propagation

We have now gathered all relevant information on the uncertainties, which needs to be propagated to match the level of detail pertaining to the dataset on prior emissions.

The country-level uncertainties represent a 95 % CI (normalized to be unitless), which is given either as one value or as lower and upper values. For the latter, when the lower and upper values show less than a 5 % difference, we use a Gaussian uncertainty distribution; otherwise, we use a log-normal uncertainty distribution. For CO, the uncertainty distribution is often log-normal. When the reported standard deviation exceeds 30 %, we also use a log-normal uncertainty distribution to avoid obtaining negative values. We use uncertainty propagation to estimate the uncertainty in emissions from the standard deviations (σ) in AD and EFs:

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{\sigma_{\mathrm{AD}}}{\mathrm{AD}}\right)^2 + \left(\frac{\sigma_{\mathrm{EF}}}{\mathrm{EF}}\right)^2}. \tag{4}$$

To examine the importance of error correlations in AD and EFs, we performed a sensitivity analysis on the European emissions (see the Supplement). We found that including error correlations in AD and EFs has limited importance, and, henceforth, we ignore these correlations.

Since these uncertainty error propagation equations assume Gaussian errors, we need to translate log-normal error distributions into equivalent Gaussian distributions. We approximate the Gaussian standard deviation of a log-normal distribution using

$$\frac{\sigma_X}{X} = \frac{(\ln(\mathrm{lim}_{\mathrm{upper}}) - \ln(\mathrm{lim}_{\mathrm{lower}}))}{4}, \tag{5}$$

where $\mathrm{lim}_{\mathrm{upper}}$ is the 97.5 percentile and $\mathrm{lim}_{\mathrm{lower}}$ is the 2.5 percentile of the log-normal distribution. Note that the combination of Gaussian and log-normal functions does not result in a log-normal function because the result can be negative. However, here we assume that the combined distribution is log-normal because the Gaussian uncertainty is often relatively small compared to the log-normal uncertainty.

The sub-sector-level emission uncertainty estimates are propagated to obtain an uncertainty estimate at the GNFR level:

$$\sigma_{E,\mathrm{agg}} = \sqrt{\sum_{i=s}^{n} \sigma^2_{E,\mathrm{sub},s}}, \tag{6}$$

where the subscript "agg" refers to the aggregated emissions and uncertainties and the subscript "sub" refers to the sub-sectors that are part of the aggregated sector. To use Eq. (6), we need the emission budgets because this equation uses actual standard deviations instead of normalized ones.

These simple uncertainty propagation functions work well under specific circumstances. When uncertainties follow a non-Gaussian distribution or are correlated, a Monte Carlo simulation can provide a more reliable estimate of the final uncertainty. However, a Monte Carlo approach is also computationally demanding when using such an extensive dataset. We tested and compared both approaches for selected countries and sectors. Detailed information can be found in the Supplement, but the main conclusion is that we can mimic the results from the Monte Carlo simulation well with the uncertainty propagation functions. The methods show a similar order of magnitude and variability between countries and trace gases. Although there is no perfect match between the two methods, we argue that this source of uncertainty is negligible compared to the uncertainty in the prior uncertainty data.

For the spatial proxies, the same set of equations is applied, but to calculate the standard deviations, weighted