Addtion to the Metrics section (2.5)

The probability estimate of the ensemble analysis is investigated using the brier score

$$BS=\frac{1}{N_y}\sum_{j=1}^{r}\sum_{i=1}^{N_y}\left(p_{j,i}-E_{j,i}\right)^2,$$

where r is the number of verification classes, $p_{j,i}$ is the forecast probability of the ensemble for class j to predict event i, and $E_{j,i}$ is the respective observed probability. The probability of the analysis ensemble to model volcanic ash concentrations within eight verification classes is analyzed. These classes are [10 μgm^{-3} , 50 μgm^{-3}], [50 μgm^{-3} , 100 μgm^{-3}], [100 μgm^{-3} , 250 μgm^{-3}], [250 μgm^{-3} , 500 μgm^{-3}], [500 μgm^{-3} , 1000 μgm^{-3}], [1000 μgm^{-3} , 1500 μgm^{-3}], [1500 μgm^{-3}, 2000 μgm^{-3}], and [2000 μgm^{-3} , ∞). The observed probability is $E_{j,i}$ = 1 if the nature run volcanic ash concentration is within a certain class and $E_{j,i}$ = 0 otherwise. A perfect probabilistic forecast result in a brier score close to 0. The small threshold values are chosen to see the performance of the ensemble for analyzing the full volcanic ash cloud. Further, the number of grid cells with large volcanic ash concentrations is limited, which renders the brier score inapplicable. The forecast probability of the analysis ensemble is the relative number of ensemble members predicting the event (i. e. the number of ensemble members forecasting volcanic ash concentrations within a certain class).

Addition to the probability analysis

The proper analysis of high volcanic ash concentrations in the atmosphere as well as their forecast accuracy are of great importance for air safety advisory services. Yet, only the ability of ESIAS-chem to provide reasonable estimates of vertically resolved volcanic ash forecasts and analysis is shown. Thus, in this section the probability estimate of the analysis ensemble for the volcanic ash emissions and the resulting concentrations are discussed. Fig. 11 shows the histogram of the relative emission factor for different assimilation window lengths for the test case on 15 April 2010 as given by the analysis ensemble. The relative emission factor is calculated for each time-height combination of the emission profile by dividing the emission rate of each member of the analysis ensemble by the respective nature run emission rate. Thus, emissions in the analysis ensemble that are temporally or vertically outside the nature run emission profile are not considered. The histogram in Fig. 11 is given for all emissions and for different emission strengths (the strongest 50%, 25%, and 10% emissions). The relative emission factor for the 12 hour assimilation window test case tend to underestimate the emissions of the nature run (Fig. 11a and Fig. 11d). By increasing the assimilation window length, the histograms peaks around 1, while the occurrences of underpredicting the nature run emission rates diminish. A relative emission factor of 1 indicates a good matching of the analyzed and nature run emission rates. This improvement by increasing the assimilation window length is especially true for the top 10% emission rates in Fig. 11d.

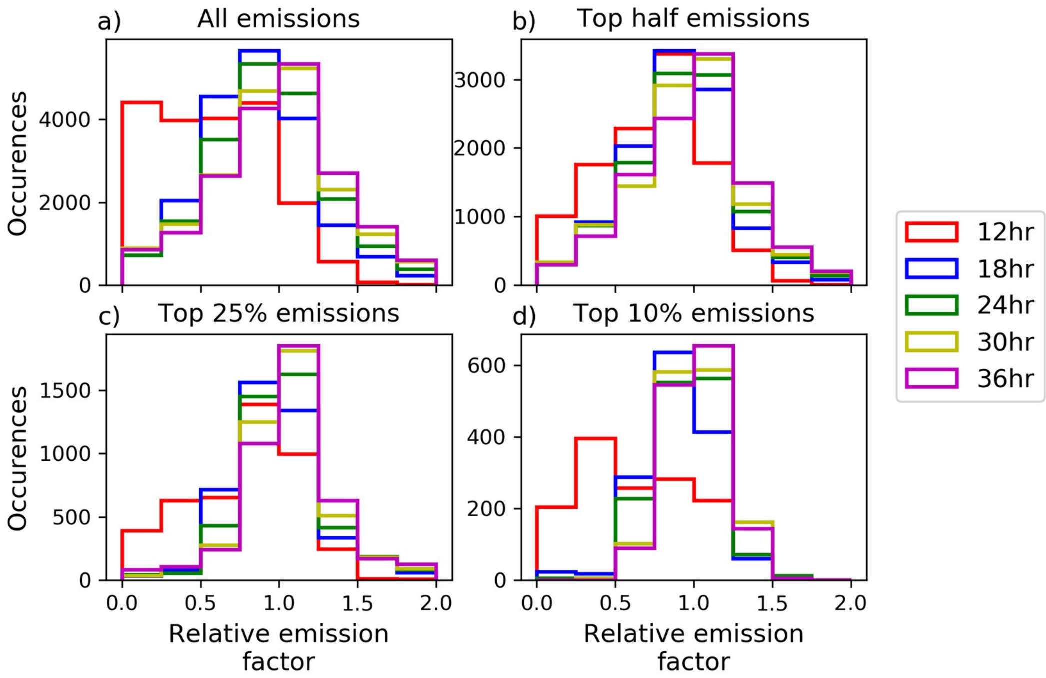

Figure 11: Histogram of the relative emission factors for different assimilation window lengths indicated by different colors according to Fig. 3 for the test case on 15 April, 2010. The relative emission factor is calculated for each grid cell of the emission profile by dividing the emission rate of each ensemble member by the respective nature runs emission rate. The histograms are shown for (a) all emission rates, (b) top half, (c) top 25%, and (d) top 10% emission rate.

Fig. 12 shows the histograms of the relative emission factors for the analysis on 29 April 2010. In general, the analysis tends to underestimate the emission rates as was previously discussed in Sect. 3.2.2. This results in a negative bias in the histograms. However, by increasing the assimilation window length, the underestimation of the emission rates by the analysis ensemble reduces. For the strongest 25% of the emission rates assimilation windows longer than 18 hours show a second maximum at a relative emission factor of 1 (Fig. 12c). These test cases also show a lower rate of underprediction for the top 10% emission rates (Fig. 12d). Thus, the results suggest that the ensembles confidence to analyze the strong emission rates in the upper emission plumes increases with increasing assimilation window length for both meteorological conditions.

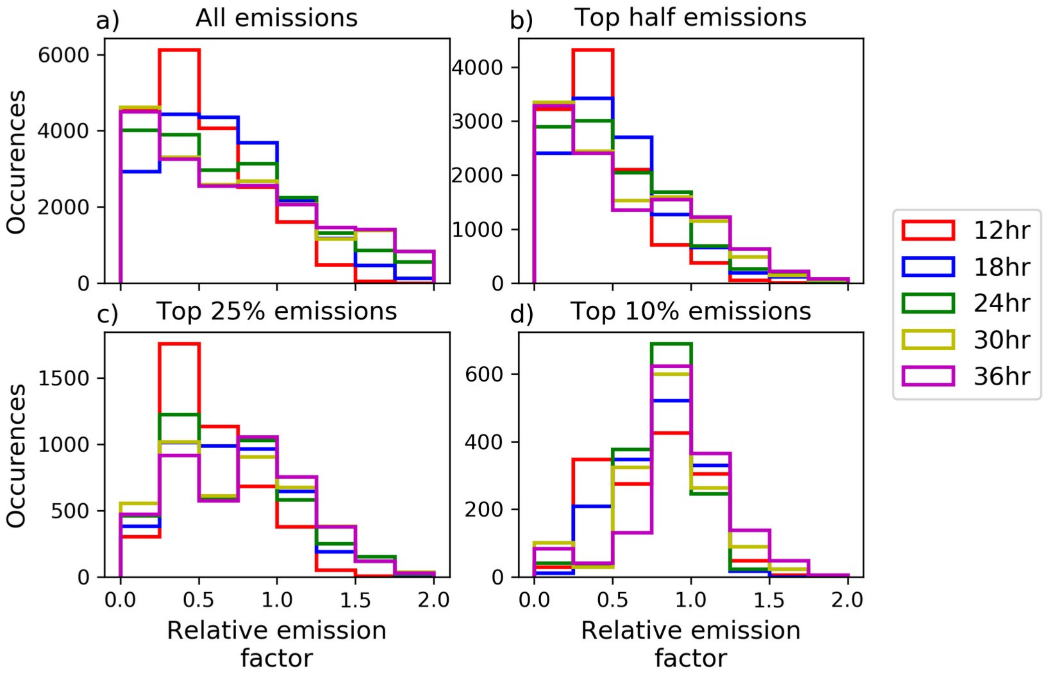

Figure 12: Histogram of the relative emission factors for different assimilation window lengths indicated by different colors according to Fig. 3 for the test case on 29 April, 2010. The relative emission factor is calculated for each grid cell of the emission profile by dividing the emission rate of each ensemble member by the respective nature runs emission rate. The histograms are shown for (a) all emission rates, (b) top half, (c) top 25%, and (d) top 10% emission rate.

The probability to which the ensemble predicts certain volcanic ash concentrations is analyzed using the brier score (cf. Sect. 2.5). The brier score is shown in in Fig. 13 for each hour and for all assimilation window length. The brier score for assimilation windows greater equal 18 hours shows a low value around 0.15 on both analysis days, which is constant over time. Shorter assimilation windows have larger brier score values that increase during the analysis time. This increase of the brier score for short assimilation windows is caused by insufficient estimates of the volcanic ash emissions, which leads to errors in the resulting volcanic ash concentrations compared to the nature run. Thus, with increasing forecast time the volcanic ash concentrations are attributed more and more to different classes used for the calculation of the brier score. This reduces the underlying probability and increases the brier score. With increasing time after the volcanic eruption the volcanic ash concentrations reduces due to dispersion and deposition. Lower volcanic ash concentrations have larger errors (not shown) meaning that ESIAS-chem is less able to predict these low concentrations with high confidence. Especially for shorter assimilation window length, ESIAS-chem is not able to estimate the emission profile properly. Thus, the corresponding volcanic ash is emitted into false layers or at false times leading to larger errors in the forecast probability.

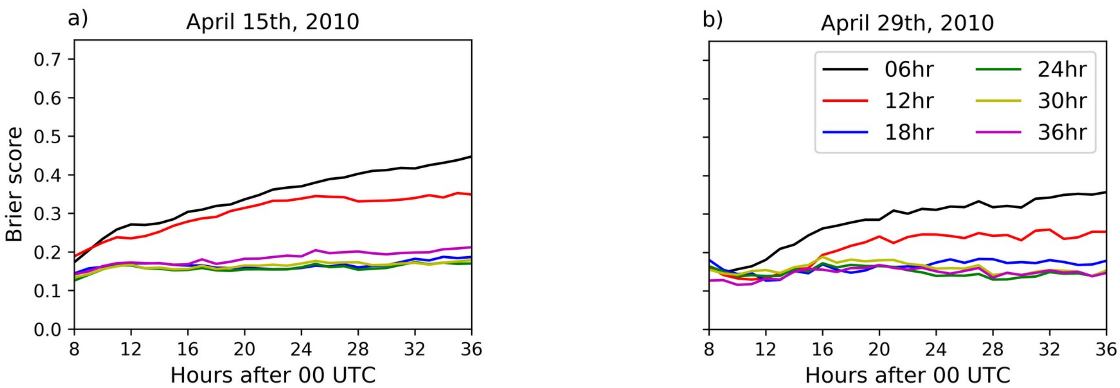

Figure 13: Brier score as calculated by (11) for each hour and all assimilation window length for (a) 15 April and (b) 29 April, 2010.