prescribed emission and averaged model output, it indicates that this VOC species is actively taking part in reactions, thus may have a larger potential influence. We quantify the potential influence by taking the absolute value of this difference in percentage, as shown in (Equ 3). x is the given VOC species, conc(x) and conc(ALL) is the spatial-temporal averaged model concentration for x and for all VOCs; emiss(x) and emiss(ALL) is the spatial-temporal averaged prescribed emissions for x and for all VOCs.

$$influence(x) = abs(\frac{conc(x)}{conc(ALL)} - \frac{emiss(x)}{emiss(ALL)}) \tag{3}$$

The VOC factor for a given VOC-observation pair can now be calculated simply by multiplying correlation and influence together, as shown in (Equ 4).

$$VOCfactor(x,y) = corr(x,y) * influence(x) \tag{4}$$

It should be noted that a VOC species with a higher factor value can not promise a better assimilation performance. But we think it is acceptable for this study, for we are not trying to select only one best VOC, but a group of VOCs to be assimilated. By selecting VOCs with factor values up to some given threshold, it can be expected that these selected VOCs contain most species with positive assimilation effects(correlated and active in reactions), and avoid negative effects brought by small-value VOCs(not correlated and/or not active in reactions). Thus, by doing this selection, a better performance than assimilating all VOCs could be expected.