General assessment

The manuscript describes the implementation of MEGAN v3.2 in SURFEX v9.0 and quantifies through sensitivity tests how global isoprene emissions respond to the SURFEX upgrade, the MEGAN version upgrade, soil moisture stress, emission factor parametrization, land-cover database and meteorological forcing. This is useful and necessary work. Model update papers of this type are what make later papers of coupled chemistry studies traceable, and the code, inputs and outputs are openly archived.

The experimental design is strong. The Baseline simulation separating the SURFEX upgrade from the MEGAN upgrade is good practice and necessary, and running every configuration under both ERA5 and IFS gives a clear impact assessment of changes.

The paper is very much in scope of this journal and should be a good addition to this journal with some modifications. My major concerns are mainly about framing and interpretation rather than the implementation itself. In short:

- The study contains no comparison with observations. This is not a problem on its own, but phrases in the paper represent the new parametrization being more realistic which is not necessarily true since there is no validation. The study should be framed as a model sensitivity study, and statements about MEGAN v3.2 being more realistic should be removed.
- Several attributions are not isolated by the experiments, in particular the claim that the updated emission factor maps drive the v2.1 to v3.2 reduction. This is not a separated factor among things that change in MEGAN3.2, so it is not a claim that can be made in the current configuration.
- Some things are thinly represented or interpreted. The derivation of the PFT-dependent emission factors is insufficiently documented. Examples of these are given below in the major comments section.
- The scope (isoprene only) is narrower than the title and abstract suggest. Again, this is not a problem on its own, the language used in the paper however should be modified to reflect what is being done.

None of these require a new modelling system; most can be addressed through rephrasing, clarification and a small number of additional simulations. I am going to recommend major review for these reasons.

Major comments

- Simulations are not compared with observations (flux towers, satellite-derived top-down estimates, or independent inventories). This is acceptable if the paper is framed explicitly as a model sensitivity study which its design supports well and falls into the journal scope which is model development. However taking that into account, statements implying that the modifications make SURFEX more realistic should be removed or rephrased, because the paper does not and cannot show this. For example , in the abstract and conclusions it is said "robust basis" and "considerably improve the flexibility and realism" In the text itself, it is written “the lower totals are presented as consistent with v2.1 overestimating emissions.” The supporting reference (Ciarelli et al., 2024) is a regional boreal study and cannot support a global conclusion dominated by tropical emissions. It is fair to say the totals lie in the range of other MEGAN v3.2 studies; it is not supported to imply they are therefore more accurate. The reference provided is informative and correct, but the conclusion is overly broad. As a last example, the section title "Evaluation of the coupled SURFEX-MEGAN model" should reflect that this is an intercomparison of configurations. If the authors would like to keep claims about realism, a comparison against independent data and observations would be needed (e.g. eddy-covariance sites, or HCHO-based top-down estimates such as those of the Müller/Stavrakou/Opacka group already cited).
- The paper attributes the 25% reduction (Baseline to R1) primarily to the updated ε^map_ISOP dataset. However, R1 changes several things at once: the emission factor maps, the revised γ_TP canopy response, the reduced set of 6 canopy types, and the new stress functions (γ_HW, γ_LT, γ_HT). The contribution of the emission factor maps is therefore asserted rather than demonstrated. I would suggest either an additional simulation combining MEGAN v3.2 algorithms with the v2.1 ε^map (or the other way around), or rephrasing all attribution statements to reflect that the aforementioned effects are combined. Similarly, the new stress functions are implemented but never evaluated. It would be useful to show whether they have any effect on isoprene at 1 degree resolution, where grid-mean winds above 12 m/s and temperatures above 40 °C should be rare.
- Eq. 1 is linear in EF. If γ were unchanged between two runs, activity factors could not compensate for or buffer a reduction in emission potential. The discrepancies discussed (e.g. a 76.4% median EF reduction in North America but only a 3.3% emission reduction; 78% globally vs 25.2%) most likely arise from two other causes. Fig. 3 shows unweighted distributions over grid cells (medians), while Table 4 compares flux totals. A large median reduction combined with a small flux reduction means EF decreased least, or increased, where emissions are concentrated. Fig. S3 shows exactly this for the eastern United States. Emission-weighted (or γ-weighted) EF changes would make Fig. 3 comparable to Table 4. γ itself differs between the Baseline and R1 (mentioned above), so part of the regional differences reflects algorithm changes. The discussion of non-linearity, compensation by temperature and radiation, and the buffering capacity of activity factors should be therefore revised.
- Tree-species composition is one of the main strengths of the MEGAN v3+ emission factor framework, so reintroducing PFT-dependent factors needs careful justification and documentation. I understand that species-level information is difficult (and mostly unreliable) to carry

in a global model with dynamic vegetation, but the manuscript does not explain how ε_k,i were derived. The phrase "consistent with the 19 lumped compounds classes" does not describe a method. How the high-resolution EFP inputs were aggregated to each PFT (e.g. area-weighted mean of ε^map over cells dominated by each PFT, aggregation of the growth-form and ecotype database, or another approach), and over which regions? What "assigned to each 16 canopy types ... thereafter redistributed consistently to the 6 new PFTs" means in code and how this relates to Table 1 and the index range k = 4–19 in Eq. (4), given that shrubs are split into three types? Why shrubs are regionalised (polar, temperate, tropical) but trees are not. A single "temperate broadleaf" factor combines high-emitting oak-dominated forests and low-emitting maple- or beech-dominated forests, which seems a larger source of error. Why R3 is 24% lower than R1 if both are derived from the same underlying data? This suggests the aggregation loses the covariance between high-emitting species composition and favourable climate, particularly in the tropics. How grid cells containing the two ECOCLIMAP-SG NVTs without emission factors (irrigated grass and crops) are treated in Eq. 4? If they are excluded from both numerator and denominator, these cells inherit the emission factor of the remaining vegetation, which may inflate emissions in R4.

The authors could also consider (or discuss) a hybrid approach: retaining cell-specific emission factors per vegetation type from the EFP, so that species-composition information is preserved while vegetation fractions evolve dynamically.

- Only isoprene is analysed, for one year, in offline mode. The title ("global biogenic emission simulations") and the statement that the model estimates emissions of 201 speciated BVOCs (L122) overstate this. Plus, the MEGAN v3.2 pre-processor computes emissions for 19 compound classes; the 201-species output is obtained by a fixed speciation χ_j that, as stated in the paper itself, no longer depends on PFT composition. This should be stated clearly, together with the distinction between the MEGAN v3.0 and v3.2 pre-processor treatment of speciation. Either the scope should be broadened (e.g. monoterpenes and at least one oxygenated class) or the title and abstract narrowed to isoprene.

- Mentions are made about CO_2 inhibition and ozone stress in future work. Listing these responses as future work is reasonable, but there are problems related to these topics that need to be addressed. 1) The paper states that the low totals "may also reflect the absence" of γ_CO2 and γ_AQ. Both responses reduce emissions, so their absence would make R1 higher, not lower. This sentence has the wrong sign and should be removed or corrected. Activating them would logically move totals further below the 350–800 Tg yr^{-1} range. 2) The justification for disabling γ_CO2 (no recent, spatially consistent forcing data) is weak. CO_2 is well mixed, a global annual mean of about 411 ppm for 2019 would be adequate, and the SURFEX offline forcing already includes CO_2 as per the paper itself. The justification is reasonable for γ_AQ, which requires ozone exposure fields. 3) Sect. 2.3 describes these functions as implemented, but they are never exercised. The text should state that they are untested in this configuration. Also, why mention the implementation if they are never used or evaluated? 4) The description of γ_CO2 as acting "for concentrations superior to 400 ppm" should also be modified since the inhibition is defined relative to a reference concentration.

- The statement that the ERA5–IFS discrepancy increases from about 5% in MEGAN v2.1 to about 10% in R1, and that MEGAN v3.2 therefore amplifies forcing sensitivity (L401–405), compares against the SURFEX v8.1 configuration of Oumami et al. Within this paper, the Baseline (SURFEX v9.0, MEGAN v2.1) gives 425 vs 369 Tg yr^{-1} so -13%, while R1 gives 318 vs 289 Tg yr^{-1} so about -9%. The MEGAN upgrade therefore reduces the relative forcing spread. Please verify or explain if this claim is correct or not.

- Also, the conclusions state that the SURFEX upgrade has no significant impact, based on the ~-4% change under ERA5. Sect. 3.2.4 seems to report -12% under IFS, which is more than half the MEGAN v3.2 effect under IFS (369 to 289 Tg yr^{-1}, -22%). This should be reflected in Sect. 3.1 and the conclusions since it is not negligible. In addition, the argument that "differences over Australia below 1.5×10^{-10} kg m^{-2} s^{-1} are negligible" does not hold, since Australian fluxes in Fig. 1 are of the same order (a few $\times 10^{-10}$).

- The soil moisture comparison to Oumami et al. Should be revisited. The weaker relative suppression (about -25% vs about -38%) is attributed to lower baseline emissions. A relative reduction does not depend on absolute magnitude. A more likely explanation is spatial reweighting: v3.2 reduced emission factors most strongly in drought-prone Southern Hemisphere regions (Australia, southern Africa), which then contribute less to the global total. Please also justify retaining the v2.1 γ_SMT formulation rather than the revised drought response associated with MEGAN v3 (Jiang et al., 2018, already cited).

- There seems to be a reversal in R4 between IFS/ERA5. Under IFS, R4 is 24% higher than R3, compared with 9% under ERA5. The entire difference lies in the Southern Hemisphere, the ERA5/IFS ordering reverses, and Fig. 7e shows sharp negative spikes in July–September that do not appear in other panels. The proposed physical explanation (forcing-dependent amplification of surface energy balance differences) is speculative. Before interpreting this as a physical result, please rule out technical causes: land-sea mask or grid mismatches between IFS and ECOCLIMAP-SG, missing or corrupted forcing days, and the treatment of NVTs without emission factors (another point mentioned above). This result also qualifies the paper's main conclusion: in R4 the forcing effect (+11%) is comparable to the land-cover effect (+9% under ERA5).

Minor and technical comments

Abstract: please specify which simulations the 216–318 Tg yr^{-1} range comes from.

Seasonal values are labelled Tg yr^{-1} but appear to be seasonal totals (Tg per season). Please correct the units.

Table 4: regional totals sum to slightly above the global totals. Please clarify whether the regions overlap, and where China, India and the Middle East are assigned.

L372–373: clarify whether "East Africa (26.30%), Australia (20.70%) ..." are shares of the global increase or relative regional changes.

Fig. 3 should state that the violins are unweighted grid-cell distributions if its the case.

L184–186 and Table 1: describe the mapping of 19 NVTs to 6 canopy types in one consistent place; the text alternates between "6 canopy types", "16 canopy types" and "6 new PFTs".

The manuscript would benefit from careful language editing. Non exhaustive examples: "required to accurately simulating", "prevents them to fully constraining" (should be from), "stand highly uncertain" (remain?), "MEGAN v3.2 now relied" "relied only" (relies for both?), "may have alter" (altered?), "dynamics vegetation" (dynamic), etc.