Dear Editor,

As stated in the previous review, the manuscript entitled "Modeling the biogeochemical effects of rotation pattern and field management practices in a multi-crop (cotton, wheat, maize) rotation system: a case study in northern China" is within the scope of BG. To ensure reliability, models should be tested and improved as part of their development and application. The manuscript is important in that context but the important part of this manuscript lies with the optimization of different rotation patterns (of three cultivars: cotton, wheat, maize) and management practices which is very complex. The manuscript has been revised and improved. Nonetheless, there are important gaps (see below) which should be seriously considered. The current manuscript is not acceptable for publication in BG as it is poor in terms of validation as well as in novelty. For more details please see my comments below:

GAPS:

- it lacks for a 6 years validation that includes a rotation of all three commodity crops as well as all studied management practices in question

Yes, ideally the model should be validated with a 6-year rotation of all three commodity crops as well as all studied management practices in question. It is unfortunately that such a dataset is still lacking, even though the dataset involved in this study has been the most complete one so far available for the three-crop system. Due to the shortage of financial resources for the very expensive field observations of the trace gases emissions, the experimental period was limited to only three years, during which all the constraint and decision variables were observed for almost all the management practices in question at least in the wheat-maize fields (in practice, some observations failed to obtain expected data due to some technical failures). Three-year observations for the currently applied management practices were simultaneously conducted consecutively in the two neighbor lands cultivated with cotton and wheat-maize, respectively (see Table S3). Because the experimental cotton was in its 3rd to 5th consecutive year of monoculture following the transition from the previous wheat-maize cultivation while the experimental wheat-maize was also in its 3rd to 5th consecutive year following the transition from the previous cotton cultivation, both were assumed to be representative for the cotton or wheat-maize within a 6-year rotation cycle when their observations were used to validate the model. Nevertheless, we still emphasizes at the end of the paper that in the future study it is necessary for a 6-year model validation that includes a rotation of all three commodity crops as well as all studied management practices in question (see lines 690−692 in the revision).

- In the site simulation, NEE is still predicted with lower accuracy by the model; then how this impacted the optimization of mitigation options? Furthermore, decision variables such as NH3 volatilization is not validated by observations.

As a response to this question, in this revision the algorithm of NEE in DNDC was modified (see lines 147−166 and Eqs. 2−6 in subsection 2.2). The model modification significantly improved the simulation accuracy of daily NEE (see lines 346−351 in subsection 3.1 and Figures 1c−e in the revision). Regarding the cumulative NEE for two years of the cotton field, and three crop seasons of the wheat-maize field, the modified model showed very good performance compared to the original model (see lines 378−383 in subsection 3.2 and Figure

2b in the revision). For the three full-year cumulative NEE (two for the cotton and one for the wheat-maize field), the modified model simulations showed model relative biases (MRBs) of −13% to 8%, which was much smaller than the reported uncertainty (25% as one times standard deviation) of the eddy covariance observations. Relying on the five annual/seasonal cumulative NEE, which were derived from the eddy covariance measurements in the fields with the crop residues fully retained, and the corresponding observed crop yields, the five annual/seasonal changes in soil organic carbon stock (∆SOC) were estimated following the ecosystem carbon balance approach (or according to the mass conservation law), which showed statistically significant agreement with the modified model simulations (see lines 392−398 in subsection 3.2 and Figure 2d in the revision). Further, the corresponding five annual/seasonal net ecosystem aggregate greenhouse gas emissions (NEGEs) were also estimated based on the observation-derived estimates of ∆SOC and measured methane and nitrous oxide fluxes, which showed significantly consistence with the modified model simulations (see lines 384−391 in subsection 3.2 and Figure 2c in the revision). Regarding the NH_3 volatilizations, the model validation in this revision further involved two urea top-dressing events in addition to the previously used one with reported cumulative volatilization during 11 days following the fertilizer amendments. Using the observed cumulative NH_3 volatilization of the three urea application events in the wheat-maize fields, the validation showed small MRBs of −9% to 4%, which were much smaller than the double CVs of the spatially replicated measurements (see lines 419−426 in subsection 3.2 and Figure 2h in the revision). Relying on these model errors resulted from the validations using sub-year measurement-derived estimates or event-based observation with a marginally small sample size to screen the BMP alternatives would inevitably influence the accuracy of the results. We are aware of such an insufficiency of this study duo to the insufficient dataset available for the model validation. Hence, in the last section of the paper we emphasize the necessity of comprehensive observations in the future studies to cover all the constraint and decision variables and other factors as well as the crops and management practices in question (see lines 690−692 in the revision).

- There are some vias between modeled annual no3 leaching and assumed observations for no3 leaching. You stated "For the simulations of other nitrogen losses from the cotton field, the NO3−leaching accounted for 9−12% of the applied fertilizer nitrogen for model validation, which was comparable with the field measurements of 16−17%". This is around 10 Kg N year, rigth? Please, think about what that would means if you scale this value on a regional level (for the whole Northern China) up.

The validation of the modified model still showed large MRBs of −32% to −27%, which were less than the two times CVs of the spatially replicated field observations for the annually cumulative nitrate leaching. These MRBs represented the model-underestimations by respectively 3−4 and 13−21 kg N ha^{-1} yr^{-1} for the annual nitrate leaching rates in the cotton and wheat-maize fields subject to the currently applied field management practices (see lines 427−433 in subsection 3.2). Understandably, it would be problematic if these underestimations were directly up-scale to the entire northern China region. In fact, these simulation errors in nitrate leaching were also found to overwhelmingly dominant the simulation errors of the NIPs (see lines 626−631 in subsection 4.3 and Table 1 in the revision). However, in this study we did not attempt to make model modification to reduce these MRBs

in the nitrate leaching simulations. This is because we were hard to judge whether there were insufficiencies in the scientific structures or inappropriate parameters in the model to dominate these large MRBs due to the too large measurement errors (with two times CVs of 109−115%) for the observed annual cumulative quantities of the nitrate leaching with a too small sample size (*n* = 2). In the revision, we added discussion on this problem (see lines 631−637 in subsection 4.3).

How all this gaps affected your BMP and NIP calculations?

In order to improve the quality of this study (which is poor in terms of validation and novelty), I suggest you to do a Parameter-Induced Uncertainty Quantification for NEE + NH3 + NO + N2O and NO3 Leaching. The Bayesian framework using a Markov Chain Monte Carlo (MCMC) method, will help you to estimate the joint model parameter distribution and you can obtain and pick up the best parameter set combination (using a cost function based on model fitting parameters) that represent your measurements. After this, you can use Monte Carlo in order to derive the best management practice for the studied site.

Adding the MRB-based quantification of uncertainties for concerned variables (see lines 292−322 in subsection 2.6 and 613−626 in subsection 4.3), as our answers to the question "*how all this gaps affected your BMP and NIP calculations*", we used these uncertainties due to the simulation gaps to quantify the NIP uncertainties and thus to define the precision of BMP screening (see Table 1 and Figure 4 in the revision).

Using MCMC to pick up the best parameter set combination would be necessary if there are significant model biases for model outputs of interest relative to observations with sufficient precision while the key internal model parameters dominating the significant biases, as well as their priori distributions, are known. In this study, the key internal parameters dominating the model biases and their priori distributions were unknown and the observational precision were still low for the annual quantities of ∆SOC, NEGE, NH_3 volatilization and nitrate leaching due to too small sample sizes (*n* =2 or 3). This situation did not facilitate the MCMC based Bayesian method to do the internal parameter-induced uncertainty quantification for these variables. Fortunately, the validations showed statistically meaningful consistence between the simulations and observations of the constraint variables (crop yield, ∆SOC, NEGE) and the decision variables (NEGE, NH_3 volatilization, NO emission and nitrate leaching) used in screening the best management scenarios. In addition, most of the model-input parameters were obtained from field observations at the field site while minor input parameters on the crops were calibrated using observed yields at harvest. In these regards, the MCMC-based Bayesian method was not necessary to be used in this study. Instead, as a response to this reviewer comment, we used the model biases resulted from the validation of the individual variables to quantify the model error of NIP for each scenario and thus determine the BMP screening precision (see subsections 2.6 and 3.2, Table 1 and Figure 4 in the revision). In addition, we used the Monte Carlo test to quantify the uncertainties induced by the model input uncertainties of soil parameters (four key soil properties) for the concerned variables and the NIP of the individual management scenarios (see subsections 2.6, Table 1 and Figure 4 in the revision), which were involved in the discussion for the influences exerted by the model simulation errors for decision variables on the simulation errors of the NIPs (see lines 613−637 in subsection 4.3).

I still think that a regional inventory for Northern china should be included in the scope of your study (especially because of the importance of the crop rotation cotton, W-M in this region). For this you can pick up the best 50-100 parameters set combinations (derived with the Bayesian framework) together with the BMP (derived with Monte Carlo) and do a Parameter-Induced Uncertainty Quantification of Regional NEE + NH3 + NO + N2O and NO3 Leaching. This can help to policy decision makers to support farmers in Northern China.

As the reviewer suggested, to expand this case study to a regional inventory is very important for policy decision makers to support farmers in northern China. In fact, this is in our plan for the future study. To fulfill this very important task, however, there are still two big challenges very tough to be solved. One is the lacking of survey data on historically/currently applied field management practices at a spatial resolution higher than the sub-county or even county level. The other is the lacking of comprehensive observations covering all the constraint and decision variables obtained at spatially replicated field sites in northern China. If these two problems are still there, we do not think reliable BMPs could be resolved using the MCMC-based Bayesian approach. In the future, we will devote to the solution of these challenges to some extent while further improving the biogeochemical model, thus gradually approaching to the great goal.