# Review of the Validation Report for DayCent-CR version DCR 1.0.2

**(CAR1459_DayCentCR1.0.2_model_validation_20220426)**

**Proponent: Indigo Ag., Inc**

Reviewer: Dr. Michael Dietze, Boston University

Start date: 2022-04-27

End date: 2022-06-20

Revised to address CAR feedback: 2022-09-01

## Key Questions from Indigo for reviewer

- Does the report meet the Model Requirements? Enumerate the requirements that are and are not met.
- If it does not currently meet the Model Requirements, what changes are needed to bring it into compliance?

1. The report appears to be compliant. The DayCent-CR meets the CAR soil model validation requirements for both bias and uncertainty coverage.
2. If one were to read over the history of my dialog with Indigo, you will see that I was initially confused over the compliance of (ORG x corn x SOC and ORG x soy x SOC), but the original report stated that these were being evaluated under ORG x all x SOC. The revised report has made this more obvious.
3. Any restrictions on use are clearly articulated. The most notable change from the previous validation was the addition of a time-dependence (heteroskedasticity) on the model error, which is something I'd recommended in my previous review. While all classes are technically compliant with CAR's rules, I'd suggested that Indigo assess whether the coverage was consistent through time (Appendix E). The result of this was the conservative decision to inflate the variance of the ORG x all x SOC class to ensure 90% coverage (Appendix F).

## Recommendations the reviewer will be sending forward to CAR

*If your review finds places the Requirements are deficient (whether or not correcting them would require changes in the Validation Report), recommend changes here.*

1. I'm happy with the adjustments made around the PMU and excited to see the addition of a time-varying variance, which seems to be sufficient to remove some of the restrictions on the previous report. Overall, this is high quality work making use of sophisticated model-data fusion approaches – if this type of work was the norm for carbon crediting we'd all be in a better place.
2. In the process of discussing and assessing the time-varying variance implemented in this validation report, Indigo discovered that it is possible for a class to pass the validation requirements while generating a predicted coverage that was not conservative over the timescales that accreditation occurs in practice (~5 years). I believe Indigo's response to this situation, which involved assessing coverage over different time windows (Appendix E) and, in one instance, inflating the predicted coverage for a class that otherwise would not have met the coverage validation requirements over this shorter timescale (Appendix F). While I feel it is important to allow models to employ time-varying variances so as to provide a more honest assessment of model predictive uncertainties, it does suggest that the current

Requirements would benefit from (a) clarification on how such approaches should be validated so as to remain conservative without penalizing innovation and (b) additional reflection to ensure there are not instances where the default time-invariant variances do not also result in anti-conservative coverage or bias over the times-scales relevant to accreditation.

## Summary/General Comments

*Please include any overarching comments*

The current DAYCENT proposal, which updates one I previously reviewed and was approved by CAR, was reviewed in full and meets the requirements of the validation guidance document.

## Section by Section Comments

*Include comments as they relate to each report section. Add "none" as applicable. If comments are made directly in the PDF, there is no need to duplicate those here.*

### 1. Report Type

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document

### 2. Introduction

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document

### 3. Responsible Parties

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document

### 4. Model Calibration

1. **Time-varying heteroskedastic error**: In the first draft of the report the heteroskedastic error model structure, Var[resid] = sigma^2 exp(2tv) was first introduced without explanation for why this particular error model was selected, either theoretically or based on diagnostics from their previous calibration that suggested this shape. Resolution: Indigo clarified the basis for this choice ("it was straightforward to interpret and implement" and "it is one of the standard variance structures supported by the nlme package") and through their time-scale dependent coverage assessments (Appendix E) demonstrated that it performed well in practice. Suggestion for the future: Because the exponential variance is not expected theoretically [Dietze 2017 Ecological Applications https://doi.org/10.1002/eap.1589] it may be possible to further tighten the predictive intervals, while maintaining the required coverage, through the combination of including a nugget variance and a more slowly-growing error model (e.g. square root or asymptotic)
2. **Posterior draws**: Explanation for why fewer draws were saved for the posterior draws is valid. Nonetheless, it would be nice to have a quick check (e.g. for a subset of sites) that the ensemble variance in production runs is the same as that in the calibration. Resolution:

Figures and statistics were added demonstrating that the differences between the "production" parameter ensemble and the full MCMC posteriors are negligible.

3. **SOC interpolation**: Would have appreciated more info on how the NASIS SOC interpolations were performed. Resolution: This section, and the three associated validation sites, was dropped from the report.
4. **Other points**
    a. Cross-validation approach of using the full calibration in the end is the same as the previous report and is supported by model diagnostics
    b. Selection of parameters to calibrate conforms to best practices
    c. Construction of cross-validation folds conforms to CAR protocol, which allows different experiments within a site to be in different folds. I appreciate the checks for residual spatial autocorrelation.

## 5. Project Domain

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document.

## 6. Description of Data Requirements

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document.

## 7. Description of validation data collection process

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document.

## 8. Documentation of validation and calibration datasets, per CFG-PC-ES combination

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document.

## 9. Bias evaluation

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document. Specifically, the DAYCENT model meets the SOC validation requirements for model bias for the listed practices, crops, and LRRs.

1. Calculation of PMU differs slightly from SEP protocol to account for potential differences in sample size between the first and second measurement periods. While not explicitly called out as such, the report authors appear to have simply determined that nj from protocol eq 3.2 is (ni1+ni2)/2 (i.e. the mean sample size). When factored out the 2 in the denominator of the mean cancels out between the top and bottom of the sum but causes the 1 in eq 3.2 to become a 2 (i.e. 1 = 2/2 and the 2 canceled out). All of this seems completely appropriate to me and I cannot think of any reason this might introduce any sort of bias in the PMU. When ni1 and ni2 are the same, this estimate is identical to the existing SEP protocol.
2. In the first draft of the report all biases were less than PMU, except WATER x wheat x SOC where there was insufficient data to determine the PMU. Here the bias, -302.53, is less than the PMU across the entire data set (623.46) and the PMU for 12 out of 15 of the individual categories assessed. I couldn't find specifically where the protocol stated how such cases should be handled, but this case appears consistent with the spirit of the protocol. Furthermore, the bias is consistently negative, making any model errors conservative. That said, it is the largest individual bias reported, so in the future (i.e. not required for this validation) I'd encourage the Indigo team to take a closer look at what might be leading to this bias. Similarly, while completely validated I'd also recommend keeping an eye on DISTURB x cotton x SOC, ORG x All x SOC, ORG x soy x SOC, ORG x wheat x SOC, where the biases were likewise nontrivial, and positive in the case of ORG. Resolution: No action required on Indigo's part. Furthermore, the WATER x wheat x SOC class was dropped when the 3 sites that employed SOC interpolation were removed. This discussion was retained in the report to preserve the record of the discussion around how to handle cases where PMU cannot be determined confidently, since this this isn't fully addressed in the CAR protocol.

## 10. Model prediction error

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document. Specifically, the DAYCENT model meets the SOC validation requirements for model predictive error for the listed practices, crops, and LRRs.

1. Figure 18: given the large number of points in these figures it is difficult to determine what fraction do not overlap the observed value. Some summary statistics would be useful. Resolution: cross references were added to the figure caption pointing the the summary statistics table

2. Given the new time-varying error model, it would be useful to produce some plots/statistics that would help the reader assess how this error model is performing (e.g. how is the width of the interval changing through time? Does coverage vary with time?). I will say, visually it does not appear that the points that don't overlap with the 1:1 line are failing because they have unusually narrow interval estimates --- their interval estimates are comparable to the sites that "pass", but those individual runs were biased. Response: Appendix E was added to assess how coverage varied with time

## 11. Model validation outputs for use in SEP uncertainty calculations

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document.

## 12. Evaluation of final parameter set

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document.

## 13. Restriction on application of the model

This section was reviewed in full and found to have met the appropriate requirements of the validation guidance document. The listed restrictions are appropriate and clearly documented.

## A. Documentation of calibrated parameter sets

This section was reviewed in full and no errors were found.

1. Interesting that there's non-trivial variability in your error model statistics across the folds. I don't think this is an issue as (a) the final value is in all cases centrally distributed and (b) these parameters are more analogous to Empirical Bayes rather than fully Bayes, so it's not shocking the distributions are a bit tighter.
2. All of you P* and T* parameters are edge hitting. While not invalid, it does raise the question of how sensitive the fits are to the prior boundaries. In some cases, priors may be keeping parameters within a biologically plausible range ("optimal" parameters might not be plausible), but in others the priors may be keeping the model from a better possible fit. Would need to be assessed carefully on a parameter-by-parameter basis. To be clear, I'm not suggesting changes here, but pointing out things to think about in future rounds. Response: Indigo clarified that the ranges do correspond to what is biologically plausible

## B. Declaration of Practices

This section was reviewed in full and no errors were found

## C. Sampler diagnostics

This section was reviewed in full and no errors were found. The statistics provided were helpful for assessing the spatial independence of model errors. While there is an autocorrelation in initial SOC (which is a reflection purely of field measurements and is to be expected given climatic, biogeographic, and pedogenic processes), there was no spatial autocorrelation in the predicted change in SOC.

## D. Proposal for disambiguating pooled measurement uncertainty (PMU)

This section was reviewed in full and no errors were found.