The discussion phase of bg-2009-193 is going to end soon (11 December) and we have three reviews available, thus I am contributing an editorial comment at this point to guide the revision of the paper. The three reviews are generally very positive - the reviewers consistently felt that the paper represents a novel and important contribution to the field. They, however, raised a number of concerns - two reviewers were recommending major, one minor revisions are necessary to make the paper acceptable for publication. I think the issues raised by the reviewers will indeed (further) improve the paper and I am thus looking forward to a revision of the paper which takes these recommendations (and mine, which follow below) into account.

The revised manuscript should be line-numbered and accompanied by a point-by-point reply to the reviewer (and my) comments.

General comments: Similar to one of the reviewers I had the feeling that at times the paper was extremely condensed. In particular when discussing linear vs. non-linear models I sometimes lost track of the differences between the models, in particular which processes and how these are (not) represented - maybe the authors can invest a bit more into getting down to the processes when discussing differences between the models.

Clarifications and further details have been provided in the presentation of the two alternative models in section 2.3. In addition, we have spelled out why both models, especially the nonlinear model, differ from purely statistical models (see comments to Referee #1).

I also view the lack of other validation data for the model a drawback - not only regarding soil water content as one reviewer pointed out, but also regarding Eq.(1), i.e. the basic relationship between the green vegetation fraction and NDVI.

Regretably, there are currently no available continuous recordings of soil moisture in these grasslands. A few measurements have been made in the 1970s on two sites, but these data are much too scarce to represent a robust calibration/validation dataset for our purpose and they do not span the MODIS period.

To match the scale and the spirit of our study, remotely-sensed estimates of soil moisture would be the most appropriate. The three available products (RADAR, estimates from NDVI and surface temperature, gravimetry) all have problems which prevent them from being used in the short term. These problems include data processing and validation, non-independence from NDVI, too coarse spatial resolution. Our models have generated hypotheses on soil water dynamics and ecophysiological parameters that can (and will) be further tested as data improve over time.

There have been a large number of studies showing that the re-normalized NDVI is a good proxy of fractional vegetation cover in grasslands (Choudhury *et al.* 1994; Carlson *et al.* 1995; Carlson & Ripley 1997).

The main focus of our paper is to capture regional-scale leaf dynamics as seen by satellites. We showed that low-dimensional ecohydrological model can do a good job

in this respect. This improves on existing literature because most studies on NDVI-rainfall relationships in water-limited ecosystems have used statistical models. In our view, showing that remotely sensed data can be used to calibrate and test process-oriented models of phenology is novel.

We have added extra lines to include Log Likelihood and AIC in Table 2, and have addressed the implications of different numbers of parameters.

Both models attempt to incorporate the processes and their feedbacks that govern the mass conservation balance of soil water (W) and leaf carbon store (C).
They may be derived from the following mass balance equations :
$$dW/dt = F_P - F_E - F_T - F_R$$
$$dC/dt = F_G - F_M$$
where F are fluxes in metres of water or grams of carbon per unit of time. (F_P : precipitation, F_E : soil evaporation ; F_T : transpiration ; F_R : run-off ; F_G : carbon gain ; F_M : carbon loss).

Hence, both models are biophysical in origin, based on conservation principles, rather than being statistical models based on purely empirical curve fitting where introduced parameters are biophysically meaningless. The differences between our two models lie in the different phenomenological equations used to determine the fluxes.

Statistical methods were used to estimate model parameters and evaluate the performance of both models.

These comments call for clarifications in section 2.3, and introduction, as already foreshadowed in comments to the Editor.

Second, I understand that there are many different ways to define the objective function. But the current objective function (Eq. 4) might not be an effective one, as the median function is less sensitive compared to other common used functions. While the authors argued that MAE or CVMAE is a better statistic than MSE in parameter estimation, however, MSE (or CVMSE), r2 and other statistics were calculated and shown as goodness-of-fit indicators (Table 2). I would suggest that the authors use the commonly used objective function for parameter estimation.

Though MSE or RMSE are commonly used as a goodness-of-fit statistics, recent studies have highlighted the need to change this practice. '*RMSE is an inappropriate measure of average error because it is a function of three characteristics of a set of errors, rather than of one (the average error). Our findings indicate that MAE and MBE are natural measures of average error and that (unlike RMSE) they are unambiguous.*'(Willmott & Matsuura 2006)

We have tested alternative cost functions and these do not change the conclusions of our study. This has been mentioned in the text. However, we propose retaining the objective fuction as currently defined for the main discussion.

Third, it's a pity that no measured soil water content data were available. Otherwise, these data could provide an independent validation of the model.

We agree but see our comments in the reply to Editor.

I'm also not clear how the evapotransition data were used in the models. It seems to me that the potential evapotranspiration data (Et) were derived (P8667, line 4) and Et consisted evaporation plus plant transpiration (p8668, L5), as used in model M1. However, model M2 separated Et into bare soil evaporation and plant transpiration. Was Et in model M2 the same as Et in model M1? If not, how was evapotranspiration in model M2 calculated?

It is correct that an extra term for soil evaporation is included in Eq 3 compared to Eq 2. In both models, Et refers to a Priestley-Taylor estimate of Potential Evapotranspiration given by $1.26\,s\,R_n/(s+\gamma)$, where R_n is the net radiation absorbed by vegetation and soil, s is the slope of vater vapour saturation versus temperature curve and γ is the psychrometric coefficient. This has been clarified in Section 2.3 and elsewhere.

The authors may need to clarify these issues before the manuscript can be accepted for publication.

The followings are some minor comments and suggestions.
Specific comments:

The models are based on fundamental physical principles but statistical methods are necessary to estimate parameters and model performance. See above.

There is a lot of evidence that LAI and NDVI are linearly and strongly positively related in grasslands (Choudhury *et al.* 1994; Carlson *et al.* 1995; Carlson & Ripley 1997). This has been further emphasised.

K is the maximum value of V for the time series. It has been renamed V_{max} to improve clarity. Uncertainties in V_{min} and V_{max} refer to estimates of uncertainty in input NDVI data. Regretably, we (as others) do not have any firm uncertainty to provide here, except one from subjective expert judgment (Raupach *et al.* 2005) which has been cited.

This has been clarified in rewritten section 2.3.

There was an error here as this constraint only holds for β_4, the fraction of existing V that disappears at each iteration.

Standard errors of model parameters can be calculated whatever the chosen cost function. Mean and standard errors are estimated from the 30 different calibration runs (each giving a single value for the parameters).

We deliberately chose not to include these purely statisical models of NDVI-rainfall relationships because our paper focused on comparison between different process-oriented models of phenology. Data not shown indicate that statistical models give very similar results to M1 but with a larger systematic error and strong deterioration in driest part of the gradient. This point has been added in the discussion.

Yes. These dates may be easily retrieved from modelled time series. However, our validation test does not favour any particular date but takes into account the full time series of NDVI.

This has been corrected.

Precipitation data from these and other weather stations have been used to generate spatial interpolations maps of climate.

We have strengthened this particular point in the Introduction.

As highlighted by referee #3, the first paragraph of the Discussion summarizes why model 2 is 'better' than model 1.

Uncertainty in climate drivers. Our best effort to tackle this issue was to examine model residual and to test for an effect of the distance from calibration site to the nearest weather station on this residual (see Figure 5).

Uncertainty in soil information. The presence of cracking clay soil is a remarkable and very consistent feature of the all the investigated sites. There was no impact of soil type on model residual. We also noticed that changes in W_{cap} (between 350 mm and 500 mm) did not significantly affect the results. Regretably, there is not much we can do to better quantify uncertainty.

We are in no position to assess uncertainties in the reflectances measured by the satellites and their effects on the NDVI values reported in the MODIS dataset. We have used the reported quality control information to screen the data prior to our analysis.

It is beyond the scope of this paper to describe the data processing steps used by NASA scientists in poducing the MOD09A2 Collection 5 NDVI product. We did not

test all the noise reduction techniques that have been used with NDVI time series. Savitsky-Golay filtering is a common one. The number of missing data is less than 2% of total points, so it is very unlikely that the gap-filling method could impact significantly the final results.

Section 2.3 needs to be re-written. It is a very dense section, and very quickly becomes tedious to read. What is a ramp function, exponential decay parameter (what is it that decays?), logistic growth term– this kind of thing. Why not include two flowcharts that show how each of the models work? Provide a key for the variables.

See comments to Editor. We believe it is unnecessary to add an extrafigure with this revised section 2.3. However, we will leave the final decision to the Editor.

p. 8863 line 11 – 'temperature' is not a resource, whereas heat can be considered a resource in this context

This has been corrected.

p. 8863 line 13 – what is 'Lotka-Volterra type' – please explain briefly or remove

We have added an explantion.

p. 8863 line 24 – what is meant by 'linear modelling' – in fact, early on the manuscript, it may be a good idea to introduce 'linear' and 'non-linear' with succinct clarifications

See reply to referee #1.

p. 8864 lines 1-10 – these type of extended explanation feels misplaced – it reads more like it belongs in the methods

We think the paragraph is fine and should be left where it is. We will follow Editor's decision on that issue.

p. 8865 line 22 – explain what is meant by 'anomalous'

Because overgrazed patches are dominated by annuals (and not perennials), the NDVI peaks tend to occur more rapidly after the first rainfalls and to be narrower than for a perennial-dominated response. We have added a sentence to explain this.

p. 8666 line 26 - Seaquist et al. (2003) reports 0.5, not 0.75. Furthermore, the context is different as the authors use MODIS data, not the Pinatuboaffected Pathfinder.
p. 8670 lines 10-12 – please explain
p. 8672 lines 16-20 – this seems more appropriate in a previous section e.g. study area
p. 8673 line 23 – time series of what?
p. 8674 line 13 – remove 'tease out,' it is a colloquialism

Corrections have been made on all these points.

Referee #3

This is an interesting study that uses a nonlinear approach to model phenology in water-controlled ecosystems, coupling the dynamics of plant cover and bucket-type modeled water balance. The study is comprehensive and incorporates long-term climatic data (8 yr) and MODIS-derived estimates of vegetation cover across a large area (400000 km2) and a large number of sites (400). The paper is well written and the study is presented with sound logic. The nonlinear approach (Equation 3) is really interesting and very puzzling from a mechanistic perspective. I personally am an advocate of this type of approach, thus I believe this paper, provided the authors address my comments below, would make a nice contribution to the audience of Biogeosciences.

My main (and only) concern about the paper begins precisely with the application of the nonlinear model and the suitability of the term 'ecohydrological' as a descriptor of the model. It is highlighted throughout the Abstract and Introduction that this model captures "coupled plant - soil moisture dynamics." Yet in my concept this assertion is misleading, because the model only captures plant dynamics (as shown in all figures) but not soil moisture dynamics per se.

While the concept of feedbacks between V and W (Equation 3) is a novel one and important to present to the scientific community, I believe it is equally important to demonstrate that the model does indeed capture a realistic V/W feedback, or the hydrological part of the story.

Page 8675 (Line 8) states: "its [this model's] parameters are more meaningful because the model aims to capture the fundamental feedbacks between soil and plant growth through a more process-based approach." While this statement is likely true (and I believe the processes described are true) this study does not specifically test the validity of this feedback. I express this serious concern because there is no assessment as to whether equifinality in W is an artifact in their model (my guess is that equifinality in W might be an artifact, given the spatial coverage that the model is applied to).

There is no direct test because we do not have soil moisture data. There is an indirect test through the better performance of M2 compared to M1 across the rainfall gradient, as shown by the smaller systematic error of M2.

Equifinality. We assume that the reviewer is using the term "equifinality" in the usual hydrological sense (e.g. Franks *et al.* 1997), where equifinality refers to the situation where multiple values of a parameter yield essentially the same solution in state space. For the critical feedback parameter β_3 in the nonlinear model M2 (eq. 3) which describes the modification of plant growth by water, we do not observe equifinality: see Figure 3D, which demonstrates a significant sensitivity of state output (vegetation) to β_3. This is one respect in which the linear and nonlinear models are different, and we thank Referee 3 for raising the issue. The final version has included comments to this effect in the discussion of Fig. 3 (P 8673).

I understand the authors acknowledge that they had no soil water content data, but could not the model be applied to additional sites where there is soil moisture data and prove its efficacy to capture these dynamics? I do not intend to sound obstinate on this

Both models are ecohydrological models because they include equations for soil water dynamics and leaf dynamics, but they differ in the way processes are represented and coupled. Regrettably, there are no available data for soil moisture dynamics in the investigated area and so calibration and validation of these models were only based on remotely-sensed estimates of NDVI. We agree that, at this stage, the predicted soil moisture dynamics only represent testable hypotheses that require soil moisture data for validation. We have added a sentence to state this. See also the reply to Editor.

P 8667 Line 27. It appears as if "In the linear model M1: : :" required a group of citations – whose model is M1?

P 8672 Line 21, Fig 2a. Please report r2 and p values of this positive relationship.

P 8675, Nice first paragraph of the Discussion.
References: Teuling and Troch, 2005. Improved Understanding of Soil Moisture Variability and Dynamics. Geophysical Research Letters 32 (paper L05404) doi:10.1029/2004GL021935

Corrections have been made on all these points and the references have been added.

Cited references

Carlson T.N., Capehart W.J. & Gillies R.R. (1995) A New Look at the Simplified Method for Remote-Sensing of Daily Evapotranspiration. *Remote Sensing of Environment*, 54, 161-167

Carlson T.N. & Ripley D.A. (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62, 241-252

Choudhury B.J., Ahmed N.U., Idso S.B., Reginato R.J. & Daughtry C.S.T. (1994) Relations between Evaporation Coefficients and Vegetation Indexes Studied by Model Simulations. *Remote Sensing of Environment*, 50, 1-17

Franks S.W., Beven K.J., Quinn P.F. & Wright I.R. (1997) On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. *Agriculture and Forest Meteorology*, 86, 63-75

Raupach M.R., Rayner P.J., Barrett D.J., DeFries R.S., Heimann M., Ojima D.S., Quegan S. & Schmullius C.C. (2005) Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biology*, 11, 378-397

Willmott C.J. & Matsuura K. (2006) On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science*, 20, 89-102