Review of "Evaluation of global continental hydrology as simulated by the Land-surface Processes and eXchanges Dynamic Global Vegetation Model" by S. J. Murray, P. N. Foster, and I. C. Prentice.

General Comments:
It is well written manuscript that describes streamflow evaluation of Land-surface Processes and exchanges (LPX) Dynamic Global Vegetation Model. This manuscript contributes in evaluating the LPX model, which has a potential to describe changes in hydrologic cycle due to changes in vegetation composition and distribution. The topic of this manuscript is suitable for publication in the Hydrology and Earth System Sciences Discussions, and will be of interest to broader audience. I recommend the manuscript to be accepted for publishing under the condition that the following concerns are addressed in the revision:

Major Concerns:

1. In order for the model to be used as a hydrologic analysis tool, it needs calibration to estimate reasonable streamflow volume (and also streamflow timing to some extent) for at least few selected river basins. At present, Figure 4 is not convincing. I recommend the authors to perform calibration at selected basins and transfer the model parameters to other similar basins (see Nijssen et al., 2001, Journal of Climate, 14, 3307-3323).

2. It is recommended to illustrate the differences in model parameterization or set up that were different from the study by Gerten et al. (2004). Further, the authors indicate that few basins improved in streamflow estimation while few became worse. Then what is the additional contribution of this paper in improving streamflow simulations? For instance, page 4228, lines 6-9: "Of the catchments which are coincident with those validated by Gerten et al. (2004), albeit for different time periods, LPX better simulates interannual variability of runoff for the Amazon, Nile and Yenisei, but more greatly overestimates for the Congo and Mississippi."

3. Include a brief description how snow accumulation and ablation are represented in the LPX model (e.g. see Fassnacht and Soulis, 2002, ATMOSPHERE-OCEAN, 40 (4), 389-403).

4. Page 4224, lines 8-11: "Data are predominantly composed of observations, from GRDC, UNH and NCAR archives, but where gaps in the time-series or unmonitored regions exist, these are infilled by using nearby gauges and the Community Land Model, version 3 (CLM3; Oleson et al., 2004)."
   Where were the gaps in time-series of observations? How were the data filled? Trends are sensitive to the infilling method when gaps are large.

Minor comments:

1. Page 4223, lines 15-16: Spell out abbreviations (e.g. ISLSCP, UNH-GRDC) during their first usage throughout the paper.

2. Page 4224, Section 2.3: Mention in the section that the model was implemented at daily time step.

3. Page 4225, lines 8-9: "LPX simulated runoff shows a decrease of 0.12 mm/year through the period 1951–2000."
   Was the trend statistically significant?

4. Pages 4227-4228, last sentence: "The model tends to perform well in replicating trends and is often within, or close to the periphery of the runoff envelope shown by the composite and monthly summed converted river discharge data (Fig. 4)."
   This is not apparent from Figure 4.
   Suggestion: The authors may use rank correlations to examine goodness of fit and overall patterns in observed and simulated total runoff.

5. Page 4229, lines 6-10: "However, while LPX is more likely to overestimate the runoff from northern latitude catchments, Gerten et al. (2004) note a deficit in comparison to observations, which is attributed to systematic errors in high-latitude precipitation data, and the simplistic model representation of snow accumulation and melt."
   Was precipitation data different in this study than Gerten et al. (2004) study? Were snow accumulation and snow melt represented differently than the Gerten et al. (2004)? Please clarify.

6. Page 4229, lines 12-15: "At the individual cell scale, LPX has a slight propensity towards late prediction of global peak intra-annual flow (mean difference of 0.068 months compared to the composite dataset, in contrast to ~1 month early with the former version of LPJ: Gerten et al. (2004)), although there are notable regional differences (Fig. 6a, b and c). "
   Why LPX has a propensity to late prediction of peak intra-annual flow when the flow is available at the basin outlet within a given day?

7. Page 4231, lines 13-14: "Flow timings are generally well simulated, but sometimes disparate by a month."
   Page 4231, lines 16-18: "Overall however, peak and low flow timings are generally well captured ……."
   Figures 4, 5 and 8 indicate that flow timings are not well captured. Also quantify the streamflow comparisons wherever possible as opposed to using abstract terms such as "well simulated" and "well captured" throughout the manuscript.

8. Page 4233, lines 25-27: "The absence of wetlands and dams in the model affects both the peaks in flow rate and volume of discharge retained, and could be improved via inclusion of flow routing (e.g. Olivera et al., 2000; Oki et al., 2001)."
   Does flow routing improve peak flow rate and discharge volume by accounting for wetlands and dams? Be specific.

9. Color scale bars are not clearly visible in Figures 3 and 6.

10. Axes are not clear in Figure 5. Why summed monthly streamflow are higher than monthly summed precipitation (e.g. Amazon, Congo, Ob, and Lena) in Figure 5?