Anonymous Referee #2

The manuscript presents a method consisting in a Stochastic Dynamic Programming (SDP) management model for a system including one reservoir and one aquifer. The aquifer is represented as a box model. The problem is solved with a combination of Genetic Algorithms and Linear Programming (GA-LP) to tackle the non-linearities and non-convexities caused by the head-dependent pumping costs. The framework is applied to the Ziya River system (North China), where groundwater overdraft has led to a significant decrease in the aquifer levels. The results of the SDP are provided in the form of water value tables used as prices in a forward-moving simulation run. The estimated costs given by the model when the aquifer levels reach equilibrium, in comparison with business-as-usual values not considering groundwater overexploitation (previous paper), serve as estimation of the cost associated to a recovery in the aquifer level.

The provided manuscript refers to a critical problem in many arid and semiarid areas: persistent groundwater overexploitation, which has caused considerable damage in both water quantity and quality across the world. The methodology is well-presented and exposed in the case study. Coupling stochastic programming and groundwater simulation is cumbersome, and new approaches to alleviate its complexity and transform those results into management policies could support the application of those tools in water resources management. For that, this paper has a considerable potential interest for publication in HESS. In addition, it is well-written and well-structured. However, there are some important points that the authors should address in order to enhance the manuscript.

General comments

The method strongly simplify the hydrology (just a Budyko model for assessing runoffs, and fixed % of groundwater recharge no justified), as well as the spatial representation of the system (all surface reservoirs lumped into a single one) and the groundwater simulation (a lumped box model with unclear if not missing representation of stream-aquifer interaction). Despite the presentation as a hydroeconomic model, the economics is also highly simplified (constant water demands, constants curtailment cost). These simplifications need to be justified, including an analysis of how realistic these assumptions are. This can be done along the text when the assumptions are presented. Overall, the limitations of the modelling approach should be clearly stated either in the Discussion or the Conclusions.
Answer from authors*: There are two main reasons for the high level of simplification: Limited data availability and the limitations of the SDP method (curse of dimensionality). Despite the highly simplified system representation, we believe that the modeling framework provides interesting and non-trivial insights which are extremely valuable for water resources management on the NCP. All assumptions, simplifications and their implications will be carefully discussed in the revised manuscript.*

The paper constantly refers to the previous analysis done by the authors, published in another paper, whose results represent the business as usual situation, not shown in this one with the exception of the total annual cost (Discussion). Thus, the presented paper looks like a second part of the one previously referred, since which it is quite hard to fully understand it without the other one. Maybe the authors could briefly include more description of the method and results for the business as usual situation, or update those at the light of the findings of this paper, in order to facilitate the comparison between both alternatives in this paper.
Answer from authors*: The previous study was a traditional implementation of SDP on a single-reservoir system and shows optimal management disregarding dynamic groundwater storage and head-dependent groundwater pumping costs. We will add a brief summary of the previous study to facilitate comparison between both alternatives within this paper.*

Comment 2 – Introduction. While being successful in presenting the problem, the Introduction seems a little confusing. At first, one would expect some comments about why is important to jointly manage surface and groundwater prior to enumerating the state of-the-art on conjunctive use optimization. While the division between deterministic and stochastic programming is adequate, the state-of-the-art presented consists in describing several references rather than explaining briefly both approaches supporting both explanations with references. It is said in the paper that "has been addressed widely in the literature" (which is true) but then only 4 references for deterministic and 2 for stochastic are shown. I would prefer to not explain what has been done in a little number of papers, but to discuss the different approaches employed and then enumerate the references. Besides, the review seems to not have moved prior to the 90's, when the topic appeared in the 60's and 70's.

Answer from authors: *We will expand the introduction as suggested by the reviewer.*

Comments 3 (Case study) p. 5935. It is assumed that the full storage capacity can be managed flexibly without consideration of storage reserved for flood protection or existing management rules. Why ?

Answer from authors: *Reservoir rule curves and flood control volumes were not available as such information is classified in China.*

So how flood protection pools are taken into account? Are you using a realistic useful storage?

Answer from authors: *Flood protection is not taken into account in this study. It will, however, be easy to implement a volume reserved for flood storage within the proposed framework. This will reduce the available storage and increase water scarcity in the long dry season. In the present model setup, we find the lower limit on water scarcity costs, assuming that the entire storage capacity is available for storing water. Reservoir spills will cause an economic loss and the model tends to avoid spills by entering the rainy season with a low reservoir storage level.*

p. 5935. . . . analysis of dynamic interactions between the groundwater and surface water resources. It seems that the box model that you use for groundwater does not account for any dynamic interaction between groundwater and surface water. Is this correct? If that is the case, groundwater discharges (outflow) and stream-aquifer interaction are not considered . . . Please show that it is correct to neglect this groundwater outflow components. Otherwise, we have an incomplete groundwater balance.

Answer from authors: *The groundwater model is a simple box model (Infiltration + Storage = Pumping + Overflow). The groundwater overflow is only used in extreme cases where the total demands + empty storage < infiltration. The spills will go to the spill variable and leave the system, practically as baseflow to the rivers (unavailable to allocation). The aquifer is so heavily over-exploited that no significant baseflow is being created or will be created in any foreseeable future. We will clarify this in the manuscript.*

A rainfall-runoff model previously used in the paper of the business-as-usual run. It is unclear if you simply took the resulting inflow values of that study or if you update that model. If it is an update, then the calibration results should be presented.

Answer from authors: *The exact same hydrological model results were used in both studies. No new calibration was performed and space was therefore not used on repeating the details. Note that the hydrological model does not represent the actual modified discharge in the rivers today, but is an estimate of the natural water availability. We will clarify this in the manuscript.*

In addition, I do not see the point of developing a daily model and then aggregate the results. It would have been easier to directly develop a monthly model.

Answer from authors: *We need an estimate of the natural water availability and chose to reuse the estimate from our previous peer-reviewed study. In this study, we had access to daily weather data from the Chinese Meteorological Services.*

Besides, it is said that the recharge is estimated upon the precipitation, using the average precipitation value corresponding to the inflow class as characteristic value. That assumes a perfect correlation between precipitation and inflows, which is uncommon. Would have then possible to be included in the Markov chain? . . . although it would suppose an increase in the curse of dimensionality phenomenon . . .

Answer from authors: *It would be possible to include another Markov Chain describing the groundwater recharge transition probabilities. With 3 flow classes for both runoff and recharge the number of inflow* *scenarios would increase to 3x3 = 9. However, we do not have any observations of groundwater recharge to develop these statistics. In the absence of such data, we decided to assume perfect correlation.*

Comment 4 (2.2. optimization model formulation) There is a variable named "groundwater spill". Does it refer to "groundwater discharge". Where does physically go this discharge? Please give an explanation about what means this spill, and how this is modeled.

Answer from authors*: The groundwater spill is only used in rare extreme cases where the total demands + empty storage < infiltration. These spills will go to the spill variable and leave the system, practically as base flow to the rivers (unavailable to allocation). As we are discretizing the entire groundwater storage (empty to full), we experience this situation occasionally in the backward iteration. The resulting lower water values and the large discrete storage intervals will prevent the forward simulation to get near these spills. We will clarify this in the manuscript.*

Comments 6 (2.4 Solving non-linear and non-convex sub-problems) The non-linearities tackled by your GA-LP algorithm are the decision variables regarding final storages. In an alternative SDP approach, these variables are kept discrete. If you keep them discrete, the problem becomes linear again and there is no need to maintain the timeconsuming GA procedure. In fact, that ability to work out non-linearities is one of the main advantages of Dynamic Programming (DP). Why have you not taken the ending groundwater table Vgw,t+1 discrete? It would have saved you a huge amount of time, although with less quality in the results, as you point out. I would think it would have been worth it, specially regarding at the steady water values found in the aquifer.

Answer from authors*: This was also our initial idea. First problem is the discretization. We would need a very fine discretization of the groundwater aquifer to allow discrete storage levels and decisions. If not, the discrete volumes of the large aquifer become much larger than the combined monthly demands. Storing all recharge will therefore not be sufficient to recharge to a higher discrete storage level. Similarly, the demands will be smaller than the discrete volumes and pumping the remaining water to reach a lower discrete level would also be infeasible. For this reason, we decided to allow free end storage. Free end storage requires interpolation between the discrete storage levels.*

With free surface water and groundwater end storages, the future cost function has three dimensions (surface water storage, groundwater storage and expected future costs). With our head-dependent pumping costs and increasing electricity price, we observed that the future cost function changes from strictly convex (very low electricity price) to strictly concave (very high electricity price). At realistic electricity prices, we observed a mix of concave and convex shape. As we use Benders decomposition (require strict convexity), this caused a problem.

Instead, we developed the hybrid LP-GA model which was applied successfully. This model can deal with any electricity price and (= any groundwater pumping costs) at any storage level.

This is an important message from the study and we will focus on communicating this better in the manuscript.

Comment 7 (2.4 Solving non-linear and non-convex sub-problems) A misunderstanding regarding piecewise linear interpolation is found in this section. You said that, according to Pereira and Pinto, piecewise linear interpolation requires strict convexity. However, Pereira and Pinto used a Benders decomposition, which employs piecewise linear approximations and requires convexity, but it is different from the regular procedure, which does not need the cost-to-go function to be convex. You can fit a linear function between your point and the neighboring ones, as you did when interpolating the future costs with cubic functions. Please correct that.

Answer from authors*: In our case we used a Benders decomposition instead of piecewise linear interpolation. We will update this in the manuscript.*

Comment 8 (3 Results) In the first paragraph of page 5946, it can be read that, at the equilibrium groundwater storage level, the willingness to pay is equal to 2.3 CNY m-3. In Figure 6 user's price for groundwater is always below that threshold if initial groundwater storage is at equilibrium. If the user's price for groundwater is always below the curtailment cost, why is the model curtailing the wheat agriculture? One would expect that pumping would fluctuate according to surface water availability, but without any curtailment, since it is more profitable to pump. Is there any constraint forcing that curtailment? Please elaborate.
***Answer from authors**: The 2.3 CNY m-3 is a mistake. The downstream wheat user has a curtailment cost at 2.12 CNY m-3 (rounded to 2.1 CNY m3 in table 1). The user's price for groundwater reported in Figure 6 is ~2.15 CNY m-3 (groundwater value at ~2.06 CNY m-3 and a pumping cost at 0.09 CNY m-3). This exceeds the curtailment cost of wheat agriculture (2.12 CNY m-3) and this user is therefore curtailed.*

Comment 9 (3 Results) Why a reservoir storage evolution plot does not appear in the manuscript? It would be important to see the surface and the groundwater storage in order to identify possible conjunctive use patterns. Please include the surface reservoir storage evolution or explain why it is not necessary.
***Answer from authors**: The reservoir storage plot was not included in an attempt to reduce the length of the manuscript. We will prepare a figure with comparison of groundwater and surface water storage and include it in the manuscript.*

Comment 10 (4 Discussion) In the first paragraph of page 5948, you say that SDDP only samples around the optimal decisions and, consequently, you will not be able to get the complete set of shadow prices for all state combinations. However, the SDDP sampling procedure actually employs samples that are not subjected to a pre-defined grid and, therefore, the samples are not evenly distributed across space, concentrating in the region located near the optimal decisions. The extrapolation process applied in SDDP covers the whole space but with different levels of accuracy depending in which region you look at. The difference between SDP and SDDP regards to the fact that the SDP results have the same accuracy for the whole space, while the SDDP results' accuracy varies across the space, focusing near the optimal decisions while usually decreasing when moving far from them. With SDDP you will get a complete set of shadow prices as well, but with different accuracy levels: some of them better than SDP and some of them worse. Choosing between them does not regard to having or not shadow prices, but to the degree of accuracy that you can accept on them. Please re-elaborate the comparison between SDP and SDDP.
***Answer from authors**: Thanks for clarifying this. We will review SDDP and improve the comparison in the manuscript.*

Comment 11 (3 Results and 4 Discussion) Although a sensitivity analysis was made with regard to the water demands, the curtailment costs and the transmissivity; there are other sources of uncertainty that must be taken into account. Factors like inflow and storage discretization, assumption of perfect correlation between rainfall and in- flow, pumping costs estimation, usage of a lumped model for the aquifer and so on, add a considerable amount of uncertainty to the problem. An explanation about the implications of those sources of uncertainty in the results should be added to the manuscript.
***Answer from authors**: We will expand the section on uncertainty and elaborate on the factors that are presently not mentioned.*

Comment 12 (5 Conclusion) As presented, the conclusions would not attract the reader. They seem to appear as part of the discussion rather than a separate section. It should be re-organized in order to clearly highlight what are the novelties of the study and what conclusions can be extracted from the methodology applied and the results obtained in the case study.
***Answer from authors**: We will reorganize and put focus on a brief presentation of the clear conclusions.*

DETAIL COMMENTS

COMMENT 1 (page 5934, line 11) One would expect here references about the water value method, not about the SDP one. In addition, Pereira and Pinto (1991) did not used SDP, but SDDP.
Answer from authors*: Yes, this is indeed confusing. We will remove Pereira and Pinto (1991) and leave the reader with Stage and Larsson (1961) (water value method) and Stedinger et al. (1984) (SDP in reservoir operation).*

COMMENTS 2 (page 5935) Line 11: upper storage capacity ?. This is storage capacity, what it is represented through a upper bound constraint, but the combination of terms here is unclear. I suggest to remove "upper". Please correct it in all the times this appears in the text.
Answer from authors*: We will remove "upper" as suggested.*
Line 24: Why only the upstream users have a pumping limit?
Answer from authors*: The river basin has two aquifers (upstream and downstream) which are only connected by the river. Ideally, each aquifer should be modelled as a box model, but this extra state variable would be computationally challenging within the SDP framework. We therefore set up the box model for the downstream and most important aquifer. The upstream aquifer is only bound by an upper pumping limit corresponding to the average monthly recharge. We will clarify this in the manuscript.*

COMMENT 3 (page 5940, line 21) Replace "the thickness of the aquifer" by "groundwater pumping"
Answer from authors*: Yes, we will replace this.*

COMMENT 4 (page 5941, line 1) Is it realistic to assume an even distribution of total pumping across all the wells?
Answer from authors*: The agricultural management practice is very homogeneous on the NCP. Given that 1) the majority of the groundwater wells are for irrigation, 2) the timing of irrigation + crop + climate is homogeneous and 3) the groundwater wells visited at our field trip had comparable capacity, we think that this is a fair assumption.*

COMMENT 6 (page 5943, line 18) Replace "program" by "programming".
Answer from authors*: Yes.*

COMMENT 7 (page 5944, line 24) I think that, besides the larger storage, one important reason beyond the stability shown by the groundwater values is the fact that the interaction between surface water and groundwater is not reprsented. If some sort of stream-aquifer interaction had been found, the groundwater values would have been affected by surface waters and vice versa.
Answer from authors*: Yes, for large permanent rivers this would probably be an important factor. However, the rivers/canals are very small most of the year, and most areas are quite far from a river (>10km). We therefore think that the interaction in this case study area is of less importance. We will clarify this assumption in the manuscript.*

COMMENT 8 (page 5945, line 1) Rather than decision rules, the water values tables act as pricing policies. In fact, you do that in the Discussion and the Conclusions sections.
Answer from authors*: The water value tables are the main drivers behind the release decisions and, if fully implemented in the decision process, should be referred to as decision rules. For consistency, we will use "pricing policy" throughout the manuscript.*

COMMENT 9 (page 5947, line 17) You should add "with SDP" after "feasible today". Other alternatives are able to handle large water resources systems.
Answer from authors*: Yes, indeed.*

COMMENT 10 (page 5947, line 24) Has a simulation model with higher spatial resolution been used? If not, please clearly indicate in the results section (page 5945, line 1) that the forward-moving simulation uses the same system scheme.
***Answer from authors**: No, we have only used a simulation model with the same system scheme. We will clarify this.*

COMMENT 11 (page 5949, line 24) I think that the reason beyond the small differences between SDP and DP regard to the inclusion of the aquifer rather than a very good performance of the SDP algorithm (although it is good). If you consider groundwaters in the analysis, their buffer value gives a high robustness to the surface system. This is reflected in the fact that the SDP empties the reservoir almost every year while not doing that if groundwater was not considered: it can always pump so it hedges the reservoir in an aggressive way.
***Answer from authors**: Yes, we will add this point more clearly in the manuscript.*

COMMENT 12 (page 5950, line 15) The groundwater results are independent in the recharge as well. It should be added to the list.
***Answer from authors**: Yes, we will add this point more clearly in the manuscript.*

COMMENT 13 (page 5951, line 4) I do not understand how the opportunity costs are reduced if electricity prices grow. This would apply exclusively if all the demands could freely pump and all of them had the same pumping head, which is not the case (you have demands that are subjected to pumping quotas while other cannot pump). However, the fact that electricity prices can be used to internalize the groundwater prices is valuable regardless of that.
***Answer from authors**: This is true. We will remove the electricity price statement and divert focus to the internalization of the groundwater price.*

COMMENT 14 (page 5951, line 7) Rather than opportunity cost pricing (OCP), the name should be marginal cost pricing (MCP). Please replace this definition hear and in the rest of the document.
***Answer from authors**: Yes, we will update this through the manuscript.*

COMMENT 15 (page 5951, line 10) The title of the section should be “Conclusions”.
***Answer from authors**: Yes.*

COMMENT 15 (page 5951, line 20) The non-convexity is caused by the headdependent pumping costs rather than the inclusion of the groundwater reservoir.
***Answer from authors**: Yes, we will clarify this in the conclusions.*

COMMENT 16 (page 5958, Table 2) This table has not been cited in the text. Remove it or cite it.
***Answer from authors**: An error has happened in the layout version. The reference is wrongly listed as “Table 1” on page 5945 in line 15 and 26. We will make sure that this has been corrected in the final version.*

COMMENT 17 (page 5963, Figure 4) In the surface water values part of the Figure, Vgw must be 50% rather than 80%.
***Answer from authors**: We have plotted for 80% (SW) and 50% (GW) to better represent the changes. The surface water values are changing mostly at higher storage levels while the groundwater values are not depending on the SW values. However, the figure caption wrongly states 50% and this will be updated.*

COMMENT 18 (page 5965, Figure 6) Do you mean Davidsen et al (2015) rather than Davidsen et al (2014)? If not, please add Davidsen et al (2014) to the reference list.
***Answer from authors**: Yes, Davidsen et al (2015) is the correct citation. The paper was only published online (2014) when this manuscript was submitted. We will update and clarify.*