This paper presents a comprehensive approach to implement the hydrostratigraphic reconstruction of multi-aquifers and corresponding climate change dependent groundwater flow modeling in Milan Po Plain area of Northern Italy. In my opinion, the topics of this paper might be of interest to the readers of this journal, but it cannot be considered acceptable for publication in its current state.

1. A major concern is that the paper does not appear to be significantly innovative, but consists of a complex exercise that applies varieties of approaches well established in the literature. The novelty of this paper should be reinforced to illustrate the scientific and academic findings in the study.

AC: We believe that the paper novelty does not rely in a single innovative idea, but on the formalization and testing of a rigorous hydrostratigraphic approach for the characterisation and modelling of aquifers in mixed urban and agricultural area. This approach is demonstrated for a complex case study and may be followed and replicated by other researchers in this and in similar areas. Thus, we believe that the paper may have significant scientific and practical impacts.

Here, we will discuss the main strengths of the paper and their novelty.

(1) The adoption of a hierarchical classification of hydrological units is not the first attempt in the literature. In this paper, we demonstrate how a sedimentological analysis of the stratigraphic data is required to avoid simplified and unreliable approaches and to help grouping lithological layers in significant hydrostratigraphic units (. In fact, the attribution of hydraulic parameters following the qualitative description of single layers as from the drilling phases is often affected by the subjectivity in the description. Recognition of specific sedimentological and stratigraphic sequences adds meaning and robustness to this work.

(2) For the hydraulic parametrization, new Transmissivity-Specific Capacity relationships (equations 7 and 8) are presented and compared to those proposed in the literature. Such relationships may be adopted for similar aquifers where such type of relationships are still required because of the incomplete investigation techniques adopted in the past for aquifer parametrization. This empirical approach, which uses regression analysis of specific capacity and transmissivity values from pumping tests from the same aquifer is a viable alternative to the analytical approach. However, a limitation of the method for estimating transmissivity from specific capacity is that drawdown in a well is function of well efficiency and losses, which for high pumping rates may be substantial. In fact, this method generally tends to under-predict transmissivity (Razack and Huntley, 1991). Thus, the error needs to be considered in the context of few orders of magnitude of transmissivity. For this reason, the parametric bootstrapping technique has been used to generate bootstrap samples of T-Sc, both from well and step-drawdown tests data to estimate uncertainty of the proposed equations. In particular, the results (i.e. A_1 and D parameter of equation 5) obtained for the step-drawdown tests are affected by larger uncertainties (Fig. 8) and larger confidence intervals (10th-90th percentiles) with respect to the ones obtained for the well tests. As demonstrated by Meier et al. (1999), this larger uncertainty can be related on the analysed time test duration. In fact, slopes and regression coefficients are smaller and the intercept is larger for late time than for early time data. This is also visible in Fig. 8a, where the T-S_c data points for step-drawdown tests (Fig. 8) show a higher degree of scatter with respect to T-S_c

data points for well-tests. In any case, the proposed empirical relations are similar to those available for heterogeneous sandy and gravelly sediments.

(3) Finally, the investigation of future climate scenarios allows to quantify the potential effects of global change on the groundwater levels and the water budget. As shown by several studies, this effect is very site-specific and climate-model-specific. In our paper, we demonstrate that, for the case study of Milano, the impact of climatic change is secondary with respect to anthropogenic stresses, at least for short/medium-time scenarios (20 years), and that it should not affect water resources in the near future. This is an important and novel finding that has significant practical effects for a densely populated area such as Milan. At the same time, we suggest that more complex interactions can arise from combinations of scenarios (e.g. groundwater use for irrigation, change in crop type).

In the revised version of the paper, we modified the introduction and discussion to strengthen scientific and academic findings and to keep the focus on discussing the novel elements, as discussed above.

2. Another important concern relates to the groundwater flow modeling. Although the simulation model is generally well calibrated and technically sound, the uncertainty associated with recharge should be considered in simulating future scenarios. As shown in Section 4.2, recharge for different subdomains was obtained by different methods/models and has not been calibrated in the groundwater flow model. Actually, groundwater flow and its level should be mostly dominated by recharge. Then the recharge should be evaluated to proof the value of the modeling work for improving the reliability of groundwater level under changing climatic conditions. To some extent, the change of direct recharge may be more sensitive to the modeling results than the value of indirect precipitation due to climate change described in the paper. So, I strongly recommend authors investigate the sensitivity analysis regarding the fitting parameters of the model and the input values to the model.

AC: We agree with the reviewer. Also following the comments of the first reviewer, we performed a sensitivity analysis for the hydraulic conductivity and input variables (i.e. rainfall and recharge) to establish the effect of uncertainty on the calibrated groundwater flow model, and we include this analysis in the revised version of the manuscript. Here, we briefly anticipate method and results.

The sensitivity analysis was carried out on the calibrated model. Accordingly, the horizontal and the vertical hydraulic conductivity, and the recharge (i.e. rainfall infiltration and irrigation) values of each subunit were modified by specific percentage for the input variables (±25%, ±50%, and ±100% of the initial values) and by a multiplicative factor for the hydraulic conductivity (0.1, 0.2, 0.5, 2, 5, and 10, thus allowing to test three orders of magnitude of K). The model outputs of the different parameters sets (i.e. hydraulic heads at observation points) were compared to those resulting from calibrated model to understand the sensitivity of the model to such parameters (Figs. 1, 2).

This analysis is introduced in the revised manuscript in section 4.4 and in the related result/discussion sections. As shown in figures 1, the modelled groundwater system is mostly sensitive to groundwater abstractions and rainfall recharge. Another important result is that

the model is not strongly sensitive to irrigation recharge, which is the most uncertain input in the study area.

The results of the sensitivity analysis on input parameters (i.e. recharge, irrigation, and withdrawals) can be analysed from a climate change point of view. For example, an increase of groundwater recharge may result both from "wet" and "cold" future scenarios (Allen et al., 2003) due to additional rainfall or a reduction of evapotranspiration triggering an increase of recharge, respectively. These scenarios, may lead to a further increase of groundwater level in the unconfined aquifer of about 2 m (considering an increase of 25%).

On the other hand, a dry scenario may lead to a decrease of recharge amount because most of the rainfall water is evaporated. Considering only changes in recharge rates, this may lead to a decrease of shallow groundwater levels (c.a. -2 m for a 25% reduction). Obviously, the irrigation requirements might need to be reconsidered to counteract the effect of dryness. This could lead to more groundwater withdrawals, which could lower the groundwater levels (c.a. -3 m up to -6 m, for a 25% increase), or to an increase of irrigation water volume (considering the fully gravity-driven irrigation system) which, in its turn could lead to a groundwater level increase (c.a. +3 m up to 6 m, for a 25% increase). Furthermore, according to modeling results a change in river discharge as a consequence of climate change can not be excluded (e.g. higher for rainy year and lower for droughty years such as for 2003 and 2014, respectively).

Overall, the sensitivity of the groundwater system depends on how precipitation, temperature, water demand for crop production, and withdrawals are related. Their combined effects might be cancelled out leading to moderate changes of the water balance components (*Woldeamlak et al., 2006*). This support the idea that both climatic and socio-economic changes are fundamental for simulating future scenarios. This is particularly significant because, even if the impact of climate on simulated recharge is expected to be more important, the socio-economic factors may produce highly significant regional changes especially in densely populated areas and where major changes to the temporal distribution of groundwater abstraction may occur (*Holman, 2006*; *Holman et al., 2011*). For this reason, we considered well established scenarios (i.e. *IPCC* climate projections, decrease of irrigation, and decrease of groundwater withdrawals according to demographic scenarios) for simulating medium-term scenarios (Section 4.5). For the adopted scenarios, in terms of duration, the changes in the recharge in the upper part of the catchments are not so relevant. In fact, isotopic data suggest a relative young groundwater age (i.e. between 20-30 years, *Gorla et al., 2016*).

Regarding the hydraulic conductivity, we observe that the groundwater model is particularly sensitive to changes in hydraulic conductivity of unconfined aquifer sub-units. Among these, proximal sectors of alluvial fans/megafans are the most sensitive zones. For these sectors, we also observe that extremely low values (i.e. 1 order of magnitude lower) lead to numerical instabilities. On the other side, the hydraulic heads are quite insensitive to changes in hydraulic conductivity values of the semi-confined aquifer and of the aquitard between unconfined and semi-confined aquifers. This is particularly relevant and it supports the idea that the aquitard does not affect much the behaviour of the hydrogeological system. The behaviour of the proximal unconfined aquifer (where most of the fontanili are located) is

apparently anomalous. In fact, a reduction of hydraulic conductivity results in a reduction of the hydraulic head in the whole model.

This may be the consequence of the imposed boundary conditions (i.e. Dirichlet condition based on hydraulic head surveys) and of an unreliable increase of discharge through the fontanili springs, which have been simulated as flux-constrained Dirichlet condition.

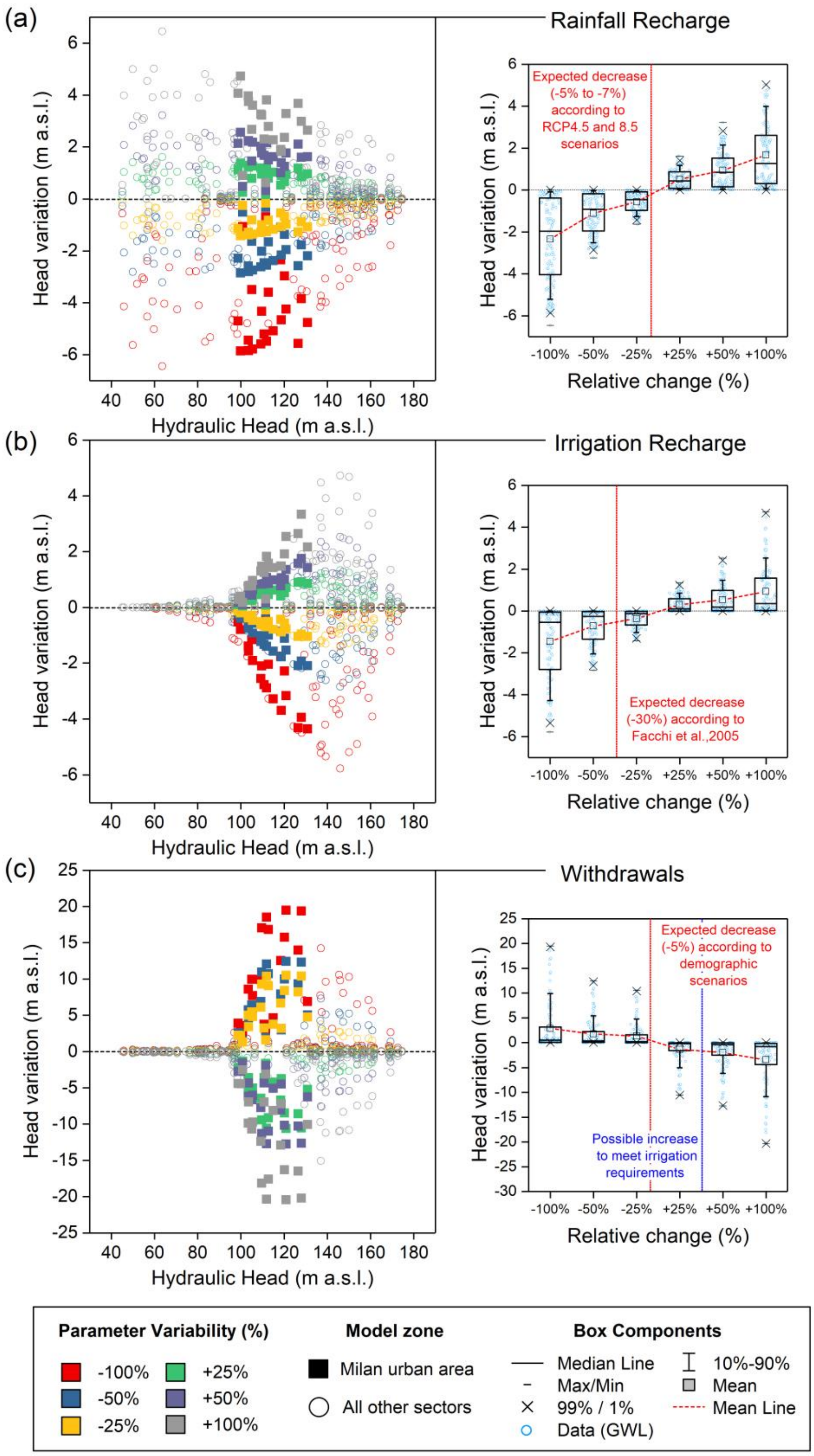

Fig. 1 – Sensitivity of groundwater levels to model inputs: (a) Rainfall recharge, (b) Irrigation recharge, and (c) groundwater withdrawals. Results for the Milan city area has been distinguished (i.e. solid squares). The scatter plots summarize the differences (m) between calibrated hydraulic heads (i.e. at observation points) and hydraulic heads of sensitivity scenarios. Inputs variability fields, according to simulated climate, irrigation, and pumping scenarios (section 4.5) are reported (red lines).

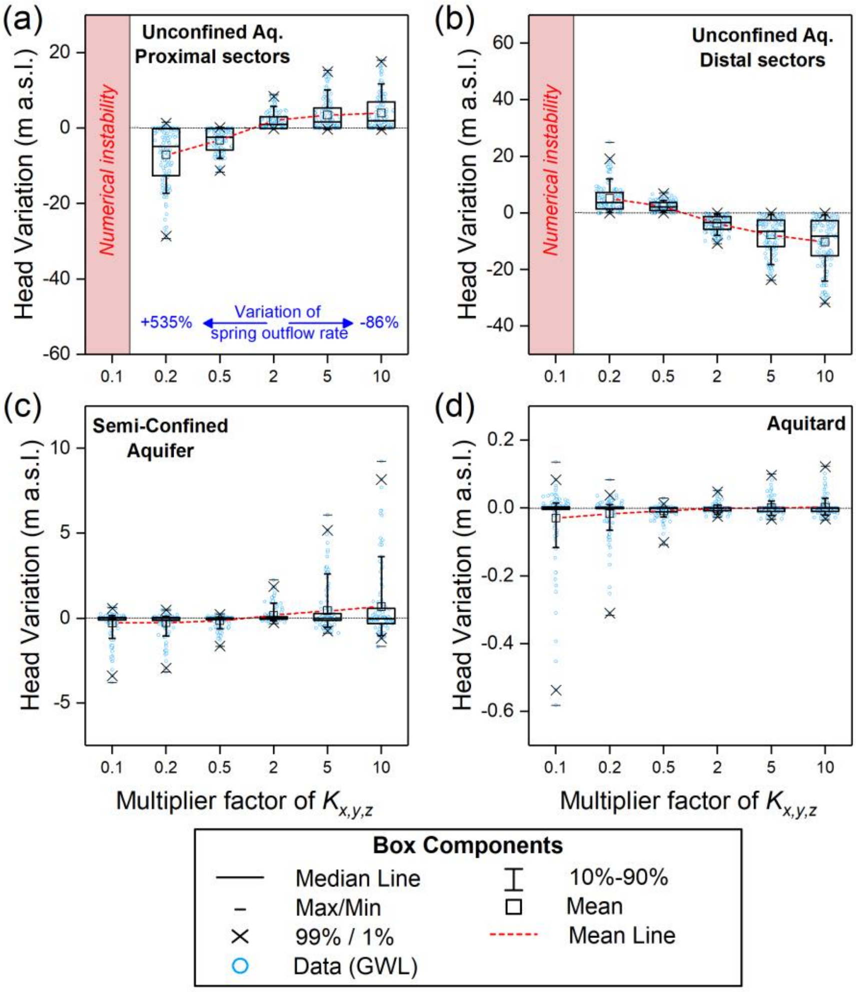

Fig. 2 – Box plots showing the sensitivity of groundwater levels to hydraulic conductivity for (a) proximal sectors (i.e. proximal fringes of fan deposits) of unconfined aquifer, (b) distal sectors of unconfined aquifer (i.e. distal fringes of fan deposits), (c) semi-confined aquifers, and (d) aquitard between semi-confined and unconfined aquifers.