**Authors' Reply to Referee #2**

**Major issues:**

1. The main concern I have with the manuscript is that based on relatively simple statistical analysis mostly only weak correlation could be achieved between land use and selected water quality variables. Nearly all presented Spearman rank correlation coefficients were relatively low (approximately below 0.7 or even much lower). Choosing this correlation coefficient means that outcomes are descriptive and no quantitative functional dependences can be derived.

We used relatively simple statistical analyses for the spatial comparisons because the patterns were obvious. Also, we believe the simplest explanations are more convincing to a broader audience (Occam's razor). For the temporal statistical analyses, we used the seasonal Mann-Kendall test, which is an appropriate test of temporal trends of monthly data. Many of the temporal trends we found were highly significant, with many also being 'meaningful' (i.e. greater than 10% change per decade). As for the relatively low correlation, these are large catchments with a considerable amount of physiographic variability, interacting factors, and feedbacks. When relationships are developed for these 77 large, diverse catchments, it is inevitable that correlations tend to be weaker than for small homogenous catchments or longitudinally along the same catchment.

2. Furthermore the study primarily focusses on suspended sediment driven water quality constituents like suspended sediment concentration, total nitrogen and total phosphorus but the analysis is restricted to monthly data, hence the most important short term events with high concentrations of the abovementioned compounds are not considered in the study.

Our study addresses both particulate and dissolved constituents. For each site, a wide range of flows were sampled over the 312 visits, including high flow events (range for all sites: 0 – 4,230 $m^3/s$). Similarly, TN ranged 5 – 7,000 $mg/m^3$ and TP ranged 0 – 8,000 $mg/m^3$ among all sites. If desired, we can include a supplementary table with the ranges of discharge and water quality variables for each site. It should be noted, however, that our spatial analyses among the 77 catchments assessed the median condition of water quality variables. Furthermore, although targeted sampling of high flow events is very relevant for load estimation of particle-related contaminants, it is not appropriate for state-of-environment monitoring like NRWQN (e.g. Davies-Colley et al. 2011; cited), covering both dissolved and particulate constituents, for which random or pseudo-random (e.g. regular monthly as in the NRWQN) sampling is most appropriate.

3. In general the manuscript is highly descriptive, clear hypotheses are missing, a sound reasoning for the approach based on former studies is not given, and some conclusions are made without clear evidence. Therefore it is not clear what is new and what has already been investigated in former studies.

We realize that the novelty and contribution of our work was not made clear in the original Introduction. Accordingly, we have revised the Introduction considerably and now address the objections of this reviewer as regards: descriptive statistics, hypotheses, justification of the approach, novelty of our contribution and conclusions. We have also changed the title of the paper to: **River water quality changes in New Zealand over 26 years (1989 – 2014):**

**Response to land use intensity**. We have pasted the last four paragraphs of the new Introduction below, which lays out what has been done in terms of multi-catchment land-water relationships, our current lack of understanding of land use intensity, and our novel contribution:

Many studies have used theoretical or numerical models to examine relationships between land use and water quality because of the lack of consistent water quality monitoring over long periods (bracketing land use change). While modelling approaches can be useful for small catchments where much is known about its landscape, modelling may not work well for larger catchments because land-water relationships are complex with interdependencies, feedbacks, and legacy effects. Empirical studies can shed light on some of these complexities, but they are only useful for their particular catchments and may have limited generality or transferability. Comparisons of many diverse catchments is probably most useful to advance understanding of broad-scale land-water relationships (Zobrist and Reichert, 2006).

One of the most comprehensive empirical multi-catchment studies to date on land use-water quality relationships has been Varanka and Luoto's (2012) study of 32 boreal rivers in Finland. They analyzed five water quality variables over 10 years as a function of a suite of physiographic, climate, and land use variables. A similar study was conducted on many of the same rivers in Finland, but with a more sophisticated temporal analysis (Ekholm et al., 2015). And several other studies have used this same river water quality dataset to investigate environmental drivers. In a study of 11 Swiss watersheds, Zobrist and Reichert (2006) analyzed export coefficients of six water quality variables from biweekly, flow proportional, composite samples over a 24-year period within the context of land use.

All of these studies, and most catchment land use studies, assessed land use (or land use change) as areal coverage. However, land use *intensity* – the inputs (e.g. fertilizer, livestock) and activities (e.g. vegetation removal) of land use – could be a better predictor of environmental impact for being a more direct measure of impact than areal coverage (Blüthgen et al., 2012; Ramankutty et al., 2006). Unfortunately, our understanding of the patterns, processes, and impacts of land use intensity is inadequate because of (1) its complex, multidimensional interactions with other landscape variables, and (2) the lack of appropriate datasets across broad spatiotemporal scales (Kuemmerle et al., 2013; Erb et al., 2016). New Zealand (NZ) provides a valuable test-bed for the patterns, processes, and impacts of land use intensity because over the past three decades pasture area has decreased but livestock densities and fertilizer inputs have increased (MacLeod and Moller, 2006; StatsNZ, 2015). Like Finland and Switzerland, NZ has an extensive long-term river water quality monitoring network, which has allowed many studies on river water quality state and trends (Smith et al., 1996, 1997; Scarsbrook et al., 2003; Scarsbrook, 2006; Ballantine and Davies-Colley, 2014) and effects of land use areal coverage (Davies-Colley, 2013; Larned et al., 2004, 2016). However, this dataset has not been assessed as regards changes in land use intensity that have occurred over the same period.

Here, we investigate long-term relationships among land use intensity, geomorphic processes, and river water quality in NZ – which provides a particularly valuable case study because: (1) it has had one of the highest rates of agricultural land intensification over recent decades and thus serves as a potential indicator for countries that are also increasing agricultural intensity; (2) it has a long, consistent, and comprehensive national water quality dataset; and (3) it is physiographically-diverse. We examined monthly data for a suite of water quality variables over a 26-year period for 77 very diverse catchments. We then compared these states and trends of river water quality to landscape data that characterized the catchments' geomorphology, soil properties, and hydro-climatology; as well as temporal changes in land use areal coverage and land use intensity, specifically livestock density and land disturbance, defined here as bare soil resulting from vegetation loss. Altogether, these analyses reveal coincident spatiotemporal patterns in land use intensity and water quality over a quarter of a century. Most of our analyses were performed at the catchment scale which integrates the spatiotemporal changes that are reflected in water quality measurements, is the appropriate scale to analyze diffuse pollution, and is the most appropriate spatial management unit (Howard-Williams et al., 2010).

4. Furthermore the manuscript is very long (41 pages text only) and not very specific including repetitions.
The manuscript is long because of our comprehensive coverage of both spatial and temporal effects of land use on a wide range of river water quality variables in complex large catchments. Arguably, the paper could be split into two manuscripts, but we feel it will have a greater impact as one paper. Further, an understanding of temporal effects is necessary in order to explain some of the spatial effects, and vice versa. We do not understand the comment 'not very specific.' We did a lot of investigation on land use practices and processes that were responsible for the patterns and relationships we observed. Maybe the reviewer is referring to our scale of analysis: catchment-scale. On line 95, we state: "Most of our analyses were performed at the catchment scale because it integrates the spatiotemporal changes that are reflected in our water quality measurements, it is the appropriate scale to analyze diffuse pollution, and it is the most appropriate spatial management unit (Howard-Williams et al., 2010)."

**Specific comments:**
Abstract: Not very specific, no specific numerical results are given on land use impacts on SSC and nutrient concentrations
As our ms shows, land use does not uniquely (predictably) mobilize SSC and nutrients. It is the *interaction* of land use with geology/climate/physiography that is important, and that is why our paper lacks specific numerical results for these contaminants. We have rewritten the abstract (below) to highlight the novelty and contribution of our work, and to address several comments among the reviewers.

Land use-water quality relationships are complex with interdependencies, feedbacks, and legacy effects. Most river water quality studies have assessed catchment land use as areal coverage, but here, we hypothesize and test whether land use *intensity* – the inputs (e.g. fertilizer, livestock) and activities (e.g. vegetation removal) of land use – is a better predictor of environmental impact. We use New Zealand as a case study because it has had one of the highest rates of agricultural land intensification globally over recent decades and it has a long, consistent, and comprehensive national water quality dataset. We interpreted water quality state and trends for the 26 years from 1989 to 2014 in the National Rivers Water Quality Network (NRWQN) – consisting of 77 sites on 35 mostly large river systems with an aggregate catchment amounting to half of NZ's land area. To characterize land use intensity, we analyzed spatial and temporal changes in livestock density and land disturbance (i.e. bare soil resulting from vegetation loss by either grazing or forest harvesting) at the catchment-scale, as well as fertilizer inputs at the national scale. Using simple multivariate statistical analyses across the 77 catchments, we found that visual water clarity was best predicted by areal coverage of high-producing pastures. The primary predictor for all four nutrient variables, however, was cattle density, with plantation forest coverage as the secondary predictor variable. While land disturbance was not itself a strong predictor of water quality, it did help explain outliers of land use-water quality relationships. From 1990 to 2014, visual clarity significantly improved in 34/77 catchments, which we attribute mainly to increased dairy cattle exclusion from rivers (despite dairy expansion) and the considerable decrease in sheep numbers across the NZ landscape, from 58 million sheep in 1990 to 31 million in 2012. Nutrient concentrations increased in many of NZ's rivers with dissolved oxidized nitrogen significantly increasing in 27/77 catchments, which we largely attribute to increased cattle density and legacy nutrients that have built up on high-producing grasslands and plantation forests since the 1950s and are slowly leaking to the rivers. Despite recent improvements in water quality for some NZ rivers, these legacy nutrients and continued agricultural intensification are expected to pose broad-scale environmental problems for decades to come.

Introduction: The first section is too general and not enough focused on the study presented. Review of international literature on dependencies between land use and water quality is weak. In Europe and the US much effort has been undertaken e.g. in connection with the SPARROW model (regression based water quality model) or time series analyses on land use induced N and C losses ( e.g. in the UK; Worrall et al.). A focus could be given on grassland impacts. Some former studies from New Zealand are mentioned, e.g. time series analysis of nearly the same data set (see References), but the manuscript could benefit much more from these studies if they would be used to clearly identify the present knowledge and use them to define research questions which are still open. The objectives are not quite clear. The study wants to illustrate long-term relationships among land management, geomorphic processes and river water quality. It is not clear why temporal changes in land use should be compared to temporal changes in water quality variables (line 92) when the investigation later on shows that land use changes are usually minor (line 313) or mostly negligible.
As we addressed in your 3rd Major Issue, we have revised the Introduction considerably and added several references for more international context. We did not talk about SPARROW studies because they address longitudinal/downstream changes in water quality, whereas we focus on differences among diverse catchments. The land use change we refer to on line 313 is areal coverage. One of the key and novel findings of our study is that while areal coverage has not changed, land use intensity (in the form of livestock densities and land disturbance) has changed; which explains some of the changes in water quality we observed. We have better articulated the difference between land use areal coverage and land use intensity in the Introduction and throughout the revised manuscript.

Line 281: These are very general and obvious findings, e.g. that discharge increases with catchment area, I would take this out
We agree that this is a general finding, but we need to include the correlation coefficient here in order to justify our selection of minimally correlated variables at the end of the paragraph.

Line 289-312: Land use distribution and patterns should be regarded as site description and not presented as part of the results (the objective of the study is to investigate relationships between Water quality and land use)
These distributions and patterns (and comparisons with physiography) resulted from our analyses of land cover and land use within and among the 77 NRWQN catchments. We would like to note

that our study is the most comprehensive analysis of land cover and land use change for the 77 NRWQN catchments. While previous studies have attributed changes in water quality from the NRWQN dataset to land use change, their comparisons (e.g. Ballantine and Davies-Colley, 2014) were not as rigorous as ours.

Line 313: The same is true for the land use change description, especially because land use change was mostly negligible.

The land use change we refer to on line 313 is areal coverage. One of the key and novel findings of our study is that while areal coverage has not changed much, land use *intensity* in the form of livestock densities and land disturbance has changed; which explains some of the changes in water quality we observed. We believe that is a novel and important contribution of our ms, and indeed we now highlight the focus on land use intensity in the revised title.

Line 324: livestock densities expressed as SU?

We report livestock densities in stock units per hectare (SU/ha), which is defined in section 3.3 (line 200).

Line 349: Relationships between disturbance (by the way a clear definition of disturbances would be helpful) and catchment characteristics should be restricted to those which are a) significant and pronounced and b) meaningful, for example for me it is unclear why disturbances has been related to mean annual sunshine duration, what does a rs of -0.25 tell us?

Previously, land disturbance was defined on line 210 as bare soil. In our rewritten abstract and throughout the revised manuscript, we provide clear definitions of land disturbance: "bare soil resulting from vegetation loss by either grazing or forest harvesting." As for the relatively low correlation, again, these are large catchments with a considerable amount of physiographic variability, interacting factors, and feedbacks, such that relationships are complex. We discuss the importance of sunshine duration on line 608: "Year-round and intense grazing is best supported by warm and sunny climates where pasture grasses are highly productive and recover quickly following intense grazing such as strip/rotational grazing which is common in NZ dairy farms." On lines 336, we also report that the highest cattle densities were found in sunny regions. Therefore for high-producing pastures with high sunshine hours, grasses recover quicker and thus have less disturbance over time.

Line 368ff: Very general outcomes are presented which we would expect in any catchment, e.g. that suspended sediment concentration decrease with flow, furthermore most of the findings have already been reported elsewhere (20 years ago, see references of the manuscript) for the study region. Because of the suggested statistical analysis no detailed functional quantitative relationships can be defined.

We agree that some of outcomes are general and intuitive; however, where we differ from previous studies on water quality in NZ is that we emphasize the *exceptions* including (but not limited to) effect of reservoirs on discharge-water quality relationships. By pointing out the exceptions, future studies can investigate their causes. Also, we believe it is a valuable contribution to provide empirical data over an entire country over 26 years to support these relationships. Further, some *HESS* readers may not be familiar with discharge-water quality relationships, and thus it is important to provide this foundation before moving on the land use effects on water quality. Finally, as already mentioned, in the revised ms we emphasize effects on water quality of change in *intensity* of land use rather than areal coverage, and this is a novel and valuable contribution.

Line 404: This section is very descriptive and simply repeats what is shown in Table 5 and table 6. A restriction on significant trends would be useful.
This section reports the results from the temporal trend analyses, which is a key element of our paper. While Tables 5 & 6 are summaries of the data, section 4.3.2 provides details on the processes and spatial patterns of these trends, another key element of our paper. Throughout the paper, we only report significant relationships (alpha < 0.05).

Line 439: The statement that total nitrogen was high if the concentrations are above 0.25 mgN/l is at least questionable. I would asses these levels still as pristine. The same is true for the assessment of TP with concentrations of 0.03 mg/l and DRP of 0.009 mg/l assumed as high levels. In Europe the eutrophication level for DRP is 0.05 for streams. Please revise.
We have revised this section, using ANZECC (2000) guidelines for New Zealand to define *relatively* high vs. acceptable values - to recognize that, in NZ, nutrient levels in near-pristine catchments are very low. We added the word 'relatively' in front of the word high for this section. For TN, we changed the guideline to 455 $mg/m^3$, which is the midway point between trigger values for upland and lowland rivers. These same guidelines for TP and DRP are 30 and 9.5 $mg/m^3$, respectively.

Line 459: It is boring to be informed again that discharge increases with catchment area. Please take this out, a number in a table should be sufficient
We removed this sentence.

Line 469: This correlation is quite small and should not be over interpreted, is there a table available on these correlations?
No, currently there is no table for all the correlations. Such a table is very large given the 25 independent variables and 11 dependent variables, and we feel it would clutter an already long ms. We only report the 'significant' (alpha < 0.05) relationships because the many weaker ('insignificant') correlations are mostly of minor interest.

Line 491ff: In the result section wordings like surprisingly , interestingly should be obmitted, especially if an additional discussion section follows. By the way why is it interesting that TP is increasing only in two catchments?
We use these terms to note when there are novel or counterintuitive relationships, as is clear from the context. We then consider why these 'unexpected' relationships may have occurred. It is interesting that TP increased in only two of the highly disturbed catchments because we usually connect intensive land use and land disturbance (i.e. bare soil) with sediment runoff, which has phosphorus associated with it.

Line 542: What does this mean, only single values or 90% percentiles?
These are ANZECC guidelines for lowland/upland, respectively. We added 'ANZECC' in front of guidelines for clarity.

Line 545: Please clearly state in the introduction which work has already been done regarding to state and trend analysis (e.g. Ballantine and Davies-Colley Water quality trends in New Zealand rivers: 1989-2009 Environmental Monitoring and Assessment 186, 3, 1939-1950)
We did state this on lines 79-82: "Like Finland, New Zealand (NZ) has an extensive river water quality monitoring network, which has allowed many studies on river water quality state and trends (Smith et al., 1996, 1997; Scarsbrook et al., 2003; Scarsbrook, 2006; Ballantine and Davies-Colley, 2014) and effects of land use (Davies-Colley, 2013; Larned et al., 2004, 2016)." However, these earlier studies did not relate water quality to catchment land use intensity and land disturbance as we do in this ms. We now highlight the novel contributions of our paper.

Line 617: What does this tell us, recovery is quick because of high Pret? But high Pret should decrease the release of P from soil to soil pore water?
In this sentence, we state that these pastures with high $P_{ret}$ are managed intensively. In the three preceding sentences, we lay out the reasoning that they recover quickly because they receive large amounts of P-fertilizer, respond favorably to P-fertilizer, and are managed intensively.

Line 667: 0.45 is definitely not a high correlation, please correct
We changed it to 'relatively high correlation.'. Obviously correlation strength is context-dependent, and as previously-mentioned, high correlations cannot be expected for large, complex (e.g. multiple land-use) catchments.

Line 672: already discussed before
The first mention of riparian fencing (line 579) was related to general improvements in water clarity among all catchments, with references to previous studies that also attributed improvements to riparian fencing. In this second instance, we relate these improvements specifically to dairy cattle and the 2003 Dairying and Clean Streams Accord (requiring exclusion of dairy cattle from channels), in view of the known degrading effects of cattle on streambanks (Trimble and Mendel 1995). We have revised this 2nd mention of riparian fencing as follows: "Second, we found that *CLAR* has actually *improved* in catchments where $SUD_{cattle}$ is high and/or has increased (Fig. 5), which we noted earlier could be a result of increased riparian fencing. In 2003, NZ implemented the *Dairying and Clean Streams Accord*, which has led to the exclusion of dairy cattle from 87% (as of 2012) of perennial rivers greater than 1 m in width (Bewsell et al., 2007; Howard-Williams et al., 2010; Gunn and Rutherford, 2013). By excluding (dairy) cattle from channels and riparian zones, the contribution of riverbank and bed erosion to degraded *CLAR* has likely been mitigated and reduced over time (Trimble and Mendel, 1995; Hughes and Quinn, 2014)."

Line 896: it is assumed that there is no doubt that fencing has improved the clarity of NZ rivers, but this is only a possible explanation. No evidence is given in the manuscript. In the presented study no specific quantitative analysis on fencing and sediment concentrations has been conducted.
We have toned down the language here and on line 672. Line 896 now reads: "While riparian fencing has plausibly improved the clarity of NZ rivers, …" Note that we have also beefed up the section on line 672 with references (below) and data.

*Gunn, A., and Rutherford, C.: The Dairying and Clean Streams Accord: Snaphot of Progress 2011/2012, Ministry for Primary Industries, Wellington, 12, 2013.*

*Hughes, A. O., and Quinn, J. M.: Before and After Integrated Catchment Management in a Headwater Catchment: Changes in Water Quality, Environmental Management, 54, 1288-1305, 10.1007/s00267-014-0369-9, 2014.*

*Trimble, S. W., and Mendel, A. C.: The cow as a geomorphic agent—A critical review, Geomorphology, 13, 233–253, 1995.*