Authors' Reply to Editor

Major issues:

Novelty of the results: It is difficult to identify what is novel with this manuscript as compared to other published studies on water quality trends in New Zealand rivers (and in general). In the Introduction for example, the reader cannot find clear statements about findings from the NZ monitoring activities and the open question that need to be addressed (e.g., L. 79 – 82: what are the reported trends, what are open questions?). In which sense does this manuscript add to previously published work?

We realize that the novelty and contribution of our work was not made clear in the original Introduction. Thus, we have revised the Introduction considerably to emphasize water quality response to change in land use *intensity*. We 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).

<u>At the same time, there is a lack of international context that illustrates which scientific questions should be tackled and how this study contributes to generating new insight.</u>

We have added numerous references (below) that add international context, mostly relevant to the comment above (all but one cited in the new introductory material given above). The exception is Erb et al (2013), which we use in the Methods to cover the various metrics of land use intensity.

Blüthgen, N., Dormann, C. F., Prati, D., Klaus, V. H., Kleinebecker, T., Hölzel, N., Alt, F., Boch, S., Gockel, S., Hemp, A., Müller, J., Nieschulze, J., Renner, S. C., Schöning, I., Schumacher, U., Socher, S. A., Wells, K., Birkhofer, K., Buscot, F., Oelmann, Y., Rothenwöhrer, C., Scherber, C., Tscharntke, T., Weiner, C. N., Fischer, M., Kalko, E. K. V., Linsenmair, K. E., Schulze, E.-D., and Weisser, W. W.: A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization, Basic and Applied Ecology, 13, 207-220, http://dx.doi.org/10.1016/j.baae.2012.04.001, 2012.

Boesch, D. F.: Challenges and opportunities for science in reducing nutrient over-enrichment of coastal ecosystems, Estuaries, 25, 886-900, 10.1007/bf02804914, 2002.

Erb, K.-H., Haberl, H., Jepsen, M. R., Kuemmerle, T., Lindner, M., Müller, D., Verburg, P. H., and Reenberg, A.: A conceptual framework for analysing and measuring land-use intensity, Current Opinion in Environmental Sustainability, 5, 464-470, http://dx.doi.org/10.1016/j.cosust.2013.07.010, 2013.

Erb, K.-H., Fetzel, T., Haberl, H., Kastner, T., Kroisleitner, C., Lauk, C., Niedertscheider, M., and Plutzar, C.: Beyond Inputs and Outputs: Opening the Black-Box of Land-Use Intensity, in: Social Ecology: Society-Nature Relations across Time and Space, edited by:

Haberl, H., Fischer-Kowalski, M., Krausmann, F., and Winiwarter, V., Springer International Publishing, Cham, 93-124, 2016.
Kronvang, B., Andersen, H. E., Børgesen, C., Dalgaard, T., Larsen, S. E., Bøgestrand, J., and Blicher-Mathiasen, G.: Effects of policy measures implemented in Denmark on nitrogen pollution of the aquatic environment, Environmental Science & Policy, 11, 144-152, http://dx.doi.org/10.1016/j.envsci.2007.10.007, 2008.
Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P. H., Estel, S., Haberl, H., Hostert, P., Jepsen, M. R., Kastner, T., Levers, C., Lindner, M., Plutzar, C., Verkerk, P. J., van der Zanden, E. H., and Reenberg, A.: Challenges and opportunities in mapping land use intensity globally, Current Opinion in Environmental Sustainability, 5, 484-493, http://dx.doi.org/10.1016/j.cosust.2013.06.002, 2013.
Ramankutty, N., Graumlich, L., Achard, F., Alves, D., Chhabra, A., DeFries, R. S., Foley, J. A., Geist, H., Houghton, R. A., Goldewijk, K. K., Lambin, E. F., Millington, A., Rasmussen, K., Reid, R. S., and Turner, B. L.: Global Land-Cover Change: Recent Progress, Remaining Challenges, in: Land-Use and Land-Cover Change: Local Processes and Global Impacts, edited by: Lambin, E. F., and Geist, H., Springer Berlin Heidelberg, Berlin, Heidelberg, 9-39, 2006.
Zobrist, J., and Reichert, P.: Bayesian estimation of export coefficients from diffuse and point sources in Swiss watersheds, Journal of Hydrology, 329, 207-223, http://dx.doi.org/10.1016/j.jhydrol.2006.02.014, 2006.

Specific comments:

L. 20: The term "disturbance" is not clear for a reader at that stage of reading. Only when going through all your later explanations it gets obvious what is meant by this term.
L. 34 – 37: This conclusion is not really based on results in this manuscript but refers to an interdisciplinary approach that is quite common in studies on land use and water quality.
We have rewritten the abstract (below) to highlight the novelty and contribution of our work, and to address several comments among the reviewers.

New ABSTRACT

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.

L. 40: This statement does hold in its generality – many important human activities (e.g., arts, science etc.) are not reflected in water quality.
Sentence revised to say: "River water quality reflects multiple activities and processes within its catchment, including geomorphic processes, vegetation characteristics, climate, and anthropogenic land uses."

L. 55 – 64: These sentences evoke the impression that mitigation has only focused on point-source pollution and that causal understanding about diffuse pollution is lacking. However, abating diffuse pollution has been pursued in many countries for several decades and many essentials are known about the drivers of diffuse pollution. Corresponding monitoring activities have been implemented and yield insight about success and short-comings of such programs (see (Kronvang *et al.* 2008) as just one arbitrary example). Your wording should better reflect the international state-of-the-art.
We added a sentence before the last sentence that reads: "Although considerable effort has been directed at monitoring and reducing diffuse pollution with some success, the legacy of pollutants from various land uses remains (Boesch, 2002; Kronvang et al., 2008)."
Boesch, D. F.: Challenges and opportunities for science in reducing nutrient over-enrichment of coastal ecosystems, Estuaries, 25, 886-900, 10.1007/bf02804914, 2002.

L. 65 – 67: I think this statement does not hold true (see comment above).
Based on the literature, we do believe that *most* studies on land use effects on water quality have been modeling studies, but we have toned down the language here and replaced 'most' with 'many'. We have also added the condition of "bracketing land use change" (see revised Introduction above).

L. 73 – 74; 87 – 88; 124 - 125: I doubt whether these statement hold true: There are many countries that have well-designed monitoring programs that run for several decades already. Some of them may also have more advanced sampling strategies than what you describe in this manuscript. In Switzerland for example, the national monitoring program starting in the 1970s for some parameters provides decade-long time-series of flow-proportional samples at weekly or bi-weekly resolution instead of monthly grab samples (Zobrist & Reichert 2006; Zobrist 2010). In the UK, some time-series date back in time over 100 yrs and make clear links between water

quality and land use (Howden *et al.* 2011). Your statements should better reflect what others have been doing for quite some time.
Thank you for bringing the Zobrist & Reichert (2006) reference to our attention. It fits well with this paragraph and we have added it. We have also added this reference to the last sentence of the previous paragraph because that was one of their concluding recommendations: "Comparisons of many diverse catchments is probably most useful to advance understanding of broad-scale land-water relationships (Zobrist and Reichert, 2006)." (We would note though, that although flow-proportional sampling such as that by these authors is valuable for load estimation, it is not appropriate for state-of-environment monitoring which needs to be random or pseudo-random (e.g. regular) with time.) The other references you suggest do not fit because this paragraph only covers studies that have compared water quality to land use for multiple catchments. To make this clear on line 73, we replaced the word 'riverine' with 'multi-catchment.' Between now and the final revision, we will keep looking for more of these types of studies. In the statements on lines 87-88 and 124-15, we say 'one of …', which is true. NZ's dataset of consistently monitored (same agency using the same protocols) monthly water quality for 15 variables for 77 catchments over 26 years is rivaled by only a few other datasets. We do not claim that NZ has *the* longest or *the* most comprehensive water quality dataset, and have modified the wording in the ms to emphasize the NRWQN's longevity, consistency and comprehensiveness, while not making any global claims for first ranking.

L. 161: P retention: please provide a reference for this method.
The Webb and Wilson (1995) reference at the beginning of the paragraph describes the methods for all soil properties used in our study, including P-retention. But we have added another reference (Saunders, 1965) that better describes the concept, process, and patterns across New Zealand.
Saunders, W. M. H.: Phosphate retention by New Zealand soils and its relationship to free sesquioxides, organic matter, and other soil properties, N. Z. J. Agric. Res., 8, 30-57, DOI: 10.1080/00288233.1965.10420021, 1965.

L. 223 – 228: Do you have ground truth data?
We have added the following sentence to the end of the paragraph: MODIS disturbance data were visually validated against 7500 random pixels from Landsat imagery and corresponding 15 high resolution Orbview-3 and Ikonos images. The overall accuracy of the disturbance index based on Landsat data was 98%.

Table 1: Give the temporal resolution of the data sets.
In the caption, and in text describing the NRWQN, we state that these are monthly data. We have added that the protocol was to take 'grab' samples to characterize a site; there was no compositing over time or space.

Table 8: For judging the quality of the regressions one should see scatter plots of the actual data.
We do show scatterplots for the primary and secondary explanatory variables in Figure 4. We now mention this in the caption to Table 8.

Fig. 4: The plots reveal pronounced heteroscedasticity. How did you deal with this issue?

We mention this on line 507. We did not correct for this in any way, but we did identify some of the outliers on Fig. 4 and explain why these occur in the Discussion section. For example, RO2 has high TP and DRP despite low SUD_{cattle} because it has high plantation forestry coverage (line 836).