Technical Note: Demonstrating a 24/7 solution for monitoring water quality loads in rivers

P. Jordan and R. Cassidy

<u>Response (bullet points) to the short comment by C. Duvert & T. Grangeon (LTHE, Grenoble, France)</u>

We thank C. Duvert and T. Grangeon for their comments on our manuscript. We address these as follows:

“The authors state that the study catchment has a flashy nature during high flows, which is clearly visible in Fig. 1a (flood hydrograph with abrupt rise followed by fast recession). They calculated the “true loads” from high frequency data obtained every 20 min. What makes them assume that a 20-min frequency survey can be assimilated to continuous records? Given the flashiness of floods, are they certain that all phosphorus temporal variations can be accurately captured using such sampling frequency? Please elaborate on this. Also, can the authors quantify the error on real loads associated with this sampling design (for instance by using short-range records obtained at higher frequency)?”

- The comment is on the interpolation of the 20 minute sampling frequency to a continuous record and whether given the flashiness of the catchment this frequency is capable of covering the full range of temporal variability. In response we have not been able to quantify the error on true loads as higher frequency data were not collected – the system is at the limit of its sampling frequency at 20 minutes. The equipment is designed to alternate a digest and no-digest step to measure two P fractions. It is the TP fraction that is reported here. We understand the point concerning use of the term ‘true load’ and the small differences that might be accrued between 20 min and sub-20 min data. We use the term to make broad comparisons with national monitoring data resolution (once per month) and also use of the term for daily samples by Johnes 2007 (Journal of Hydrology, v332, p241-258). However, we realise that there is still a level of interpolation and that ‘true load’ may still be an unreliable (and possibly completely unquantifiable) term. We therefore propose modifying the term to ’20 min interpolated load’.

“Regarding the tested sampling frequencies, we believe that a few more simulations are missing between daily sampling (Sampling strategy S9) and sampling every 7 hours (S11). For instance, the authors could have tested the error on flux estimate associated with a sample taken every 10, or 12, or 14 hours, etc. Such strategies could result sufficiently accurate in terms of annual flux estimate, and they would be even more parsimonious and cost-effective than the 24/7 survey.”

- The reviewers request further simulations to cover the possible intervals between daily sampling and the 24/7 sampling strategy presented. In response we have completed sampling simulations at 12 h, 6 h and 3 h intervals and have included them in the revised figures 2a and 2b, below. Sampling twice daily does improve the load estimate compared with daily sampling strategies, but has a larger inter-quartile range and higher mean estimate than the 24/7 approach. As load estimates generally improve with increasing sample frequency, as is expected, the load estimates are more accurate and with a smaller inter-quartile range at less than 7hr frequency. The technical note is concerned with three issues; representation of total flux, representation of important point source (diurnal) signals and parsimony in terms of field and laboratory commitment. It is contended that anything less or more than the 7 hour sampling solution compromises at least one of these concerns for inter/intra-annual monitoring of pollutant delivery patterns from specific sources.

"The authors state that the 7-hour frequency sampling strategy provides accurate results because of "the increased probability of capturing short term fluctuations in concentration". Considering the high phosphorus peaks observed during floods (i.e. diffuse source transfers), it might be interesting to estimate the contribution of such peaks to the annual phosphorus load, as compared to the contribution of point source transfers to the annual load."

- "It might be interesting to estimate the contribution of storm peaks to annual loads". Yes, we agree and this is being investigated separately. We anticipate that these will be a high percentage of total flux in keeping with previous studies (Douglas et al., 2007 for example, Nutrient Cycling in Agroecosystems, 77, 199-212). While non-storm transfers constitute a smaller flux, their impact on river ecology may be significant due to loss of dilution in summer flows. There is, however, a difficulty in isolating large scale storm events, smaller river bed flushes and diurnal signals in point source inputs – for information, the investigation will be on a critical assessment of load apportionment models using the high resolution data. This is, we feel, is beyond the scope of this technical note on sample resolution issues.

"Instead of "load values", consider using "Errors on true load (%)" for the ordinates of Fig 2. This might add to the readability of the figure. What would be the acceptable range of error on load estimates?"

- We acknowledge the suggestion to change the ordinates of Figure 2 but we would prefer to maintain these as absolute values. As a compromise, however, we have calculated % errors and have included these on a 2nd y axis as amended in Figures 2a and 2b, below.

  The acceptability of % errors may be subjective other than with the objective of reducing them according to question, catchment and resources to sample. The smallest possible error on load is important although it is possible that this will not be the only parameter required for river water quality monitoring

  Here, for example, both annual flux due to diffuse storm influences and non-storm point source patterns are important - giving flux magnitude and duration metrics. What is perhaps more important for water resource managers and researchers/modellers is the comparative extent of error propagation using the PARCOM algorithm with certain sampling regimes – (1.34% with the uniform 24/7 approach) knowing this a priori will be important for expectations of long term monitored trajectories and model outputs. (There should also be careful consideration of the algorithm used for load estimation given the high variability among different approaches (including PARCOM) shown in Cassidy and Jordan, 2011, Journal of Hydrology, 405, 182-193).

“S9 (daily sampling between 8 am-5 pm) seems to underestimate the “true load” most of the time. Fig. 1a suggests that a significant part of the phosphorus load is exported during the nocturnal period (i.e. 6 pm-8 am). As a counterweight to diurnal S9, have the authors tried to test a nocturnal sampling strategy? Would they expect to obtain an overestimation that would be somewhat symmetrical to the underestimation given by S9?”

- Sampling confined to the normal working day does, as the reviewers point out, underestimate load most of the time. We overlooked an explanation for this in the manuscript and have now rectified this in the text and through inclusion of a nocturnal (18h – 8h) sampling regime (Fig 2) which demonstrates the corresponding overestimation of load when sampling is undertaken in the evenings and through the night. The time series shown in figure 1b shows a clear diurnal variation with a peak during the night. This may be attributed to a decline in biological activity at night and reduced uptake of P but is more likely to correspond to the period when household activity and hence point source pressure, peaks. One of the advantages of the 24/7 sampling regime is that the 7 hour offsets in sampling times ensures coverage of the daily cycle, including each hour of the day on a weekly basis (e.g. Halliday et al., Science of the Total Environment – in press)

“It is surprising how S8 (360 random) provides accurate results, especially when looking at the inter-quartile range. Have the authors tried to go further into the analysis of this sampling strategy?”

- Regarding the accuracy of S8 (360 random samples) we consider that the accuracy of the approach is due in part to the clustering inherent in randomly distributed points (sample times) providing a good likelihood of investigating the full spectrum of TP variability in the annual data set without bias. As with all strategies the accuracy of the estimate increases with sample numbers and the randomness ensures that any time over each 24 h period may be sampled so avoiding the issues with daily and nocturnal sampling regimes. In practice, however, such a sampling regime would be very difficult to implement.

“The authors might want to enlarge the scope of their work by discussing about the importance of their findings for flux calculation in small catchments, and not only for phosphorus-related issues. The works published by F. Moatar and colleagues, as well as by A.J. Horowitz and colleagues, might also be of interest to strengthen the discussion and widen it to the issues faced in larger catchments.”

- We acknowledge this and can widen the discussion to issues in wider catchments. However, we still contend that the 24/7 sampling approach represents a practical strategy that is within reach of many research and government agencies seeking to monitor river pollutant fluxes in the solid phase. At larger scales, the data stream is likely to offer less interpolation uncertainty and this is beneficial not only for inter and intra-annual flux monitoring, but also for increasing knowledge of chemical and sediment fluxes in rivers, still a challenge for basic understanding and modelling efforts (e.g. Neal et al., Science of the Total Environment – in press)

“Specific comment: “in Fig. 2, the antepenult strategy is named “daily 8-18H”, whereas in Table 2 S9 is named “Daily sampling (8 am-5 pm)””.

- Table 2 S9 is incorrect – sampling was from 8-18 h. We have corrected this.

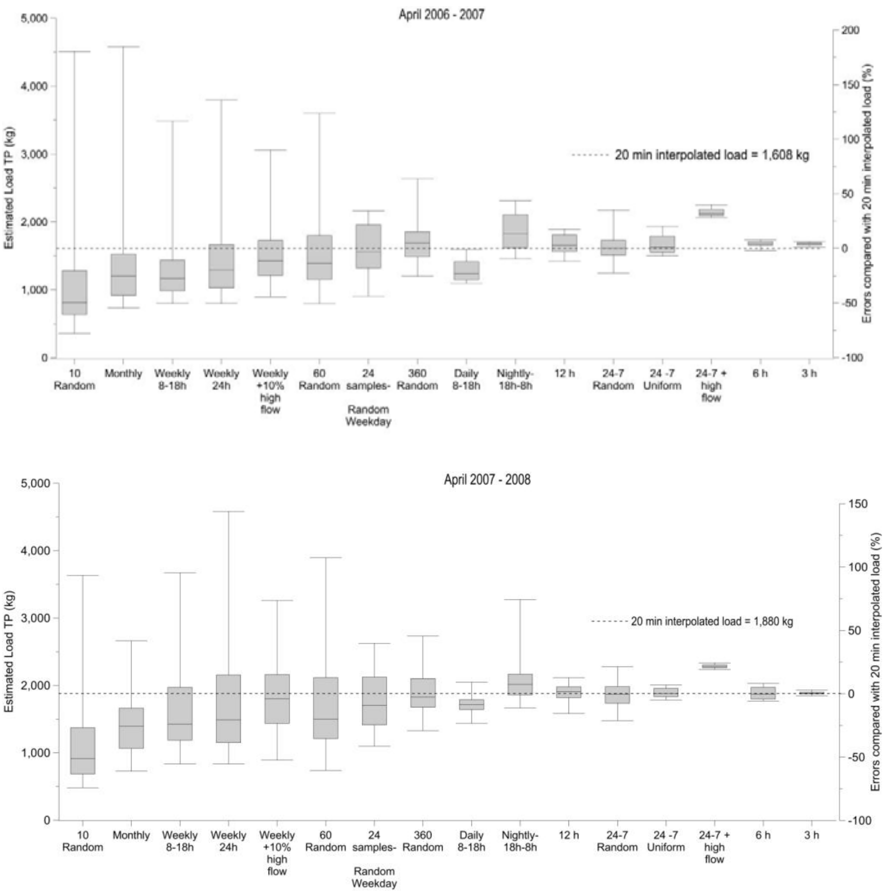

Fig. 2: Box whisker plots showing development of estimated loads (25th, 50th, 75th percentiles, maximum and minimum) using systematic and random sample sets from a decimated subhourly dataset of P concentration and discharge in 2006–2007 (a) and 2007–2008 (b).

Responses (bullet points) to the comment by Anonymous #2 Reviewer.

We thank **Anonymous Reviewer #2** for the review of our manuscript. We respond to these as follows:

"I agree that the presented sampling design is easily implemented and cost effective. Nevertheless, I wonder how the results would change in catchments with different size, storm hydrology characteristics, land use, soil characteristics and/or geology.

The presented '24/7 solution for monitoring water quality' seems to be catchment specific, since the different sampling strategies have only been tested in a 5-km2 catchment. As the authors discussed, all metrics (annual load...) are likely to improve as catchment increases. But, this might not be true in catchments with contrasting physiographic characteristics or smaller catchments."

- The reviewer queries the applicability of this sampling strategy to other catchments which differ in size, storm hydrology characteristics, land use, soil characteristics or geology.

  In terms of scale, and as we have discussed in the manuscript, the load estimates are likely to improve as the scale of the catchment increases owing to hydrological buffering which will attenuate flashy storm peaks by decreasing the rate of discharge change (Soil type and slope variability across a catchment influences the surface runoff-infiltration ratio and largely determines the flashiness of the system. Higher infiltration rates buffer the impact of storm events and slow the release of rainfall to surface waters). The catchment investigated here is on undifferentiated gley soils of poor permeability and with a high runoff to rainfall ratio also due to low evapotranspiration. As the soils are developed on drumlin tills, there is also a very high drainage density that is augmented with field drainage systems from field and river schemes in the 1980s. The catchment therefore represents a good benchmark in terms of flashiness. The Blackwater River system as a whole has been noted as being the flashiest of all rivers draining from the 5,000km^2 Lough Neagh basin (Wilcock, 1997: in J.G. Cruickshank (ed), Soils and Environment – Northern Ireland, Belfast). Our contention in this technical note is that for pollutant flux monitoring at this scale and above, and for this hydrological regime and those which are more buffered, due to soil type or geology, the 24/7 solution can be quantifiably justified as reducing annual load uncertainty and also capturing important inter-storm patterns. For information, Halliday et al. (Science of the Total Environment – in press) report the 24/7 method in the probably flashier upper Plynlimon catchment at approximately 1km^2. While these authors comment on the possibility of missing important rising limb features of chemical fluxes, in our note, we at least quantify this at 5km^2 by comparison with a unique 20 min dataset over 24 months.

"In a previous work, Cassidy et al. (2011) found that only sampling equivalent to hourly or sub-hourly frequencies were sufficient to accurately capture the scaling of the TP concentration and runoff-related discharge time series. Did you estimate the error associated with flux estimates using samples collected every 6 or 5 hours or even less? This sampling strategy might then be more expensive to implement, but in any case I would find it interesting to know how the errors decrease when planning a sampling design."

- Regarding simulations at higher frequencies, we have now included sampling at 6 h and 3 h frequencies for comparison. As discussed in Cassidy and Jordan (2011 Journal of Hydrology, 405, 182-193) higher frequencies provide better estimates as small events with high P concentration have an increased probability of inclusion within any sample set. In this paper we correctly conclude that only hourly to sub-hourly sampling was sufficient to accurately capture TP scaling. This is unchanged. In this technical note, we acknowledge that fully automated continuous monitoring is not likely to be widely available as a river monitoring solution outside focussed research projects. Rather, we attempt to quantify the uncertainty associated with decreasing the sample resolution to something that might be both manageable and acceptable.

"The authors discuss about the fact that extreme event sampling can be integrated into the 24/7 datasets (line 25, page 5042). Why did you not include the error estimation for this sampling strategy?"

- Inclusion of a single sample is difficult to model for all storm events for an annual comparison. However, we have included a sampling strategy that assumes activation of an autosampler to augment the 24/7 solution when flows are >10th percentile ("24/7 + high flow", in Fig. 2). This results in a median over estimation of annual load of 132% for 2006-07 and 121% for 2007-08. This is due to the disproportionate number of high concentration values which, when averaged using the PARCOM method raises the mean flux (eqn. 1). As we cannot properly quantify the inclusion of random extreme flow samples between two 7hr samples over a period of one year, we propose omitting this from the analysis.

"Concluding then that measurement frequency is likely to be higher in smaller catchment. Do you think the same happens in the Monaghan catchment for phosphorous loads estimations? Considering the nonlinearity of discharge- suspended sediment relationships."

"As cited by Kirchner et al (2004), Robson (1993) noted that at Plynlimon, additional pH and conductivity measurements became redundant at about the same sampling frequency as additional discharge measurements did."

- The reviewer queries whether, in the Monaghan catchment, sampling frequencies for TP exhibit similar behaviour. This does not appear to be the case for Monaghan. Discharge does not exhibit the same temporal variability as TP records do; often remaining stable for several hours while the P records rarely stay the same over 2 successive measurements.

"I agree with most of the comments made by C. Duvert & T. Grangeon (LTHE, Grenoble, France). What makes the authors assume that 20-min frequency sampling can be assimilated to continuous records?"

- The reviewer also raises the question of the assimilation of 20 minute frequencies to continuous records. We have addressed this question in our response to Duvert and Grangeon in the preceding section.

Response (bullet points) to the Interactive comment by Prof. Kronvang (Referee)

We thank **Professor Kronvang** for reviewing our manuscript. We respond to his comments and suggestions as follows:

"I believe that the authors have to reflect more clearly and show also in the Title of the manuscript that their proposal of a new monitoring strategy (24/7) is in my opinion only the most cost-effective in catchment up to a certain size (< 25 km^2)."

- We agree that we need to emphasise more strongly that the 24/7 monitoring strategy is most cost effective in small catchments and can address this in the text and in a proposed changed title: "Demonstrating a 24/7 solution for monitoring water quality loads in small catchments". However, we also contend that a single strategy could be used for ease of planning and that the method would introduce less interpolation uncertainty in larger catchments. In Irish catchments, it is important to note that storm and non-storm fluxes are important to P transfer. Decreasing trajectories in both are desirable/necessary under the WFD. The 24/7 approach allows for capturing of diurnal trends that may be abated with point source mitigation and that might not be seen with token baseflow sampling at a much coarser resolution.
- We agree to also include a reference to Kronvang and Bruhn 1996.

"Authors must be more clear on the difference in sampling strategy between streams receiving nutrient inputs from point and diffuse sources and just diffuse sources."

- We acknowledge this and the text can be amended accordingly. However, we show in Cassidy and Jordan (2011, Journal of Hydrology, 405, 182-193) that even with diffuse transfers only, important smaller events can be missed if sampling strategies are biased towards extreme events and can result in significant overestimation of annual fluxes. The time integrated approach demonstrated here, based on practical consideration of sampling equipment and laboratory commitment, is a method that can at least account for these smaller events at scales from 5km^2 upwards.