

## SPATIO-TEMPORAL ASSIMILATION OF MODELLED CATCHMENT LOADS WITH MONITORING DATA IN THE GREAT BARRIER REEF

BY DANIEL W. GLADISH\*, PETRA M. KUHNERT\*,<sup>1</sup>, DANIEL E. PAGENDAM\*,  
CHRISTOPHER K. WIKLE<sup>†</sup>, REBECCA BARTLEY\*, ROSS D. SEARLE\*,  
ROBIN J. ELLIS<sup>‡</sup>, CAMERON DOUGALL<sup>§</sup>, RYAN D. R. TURNER<sup>‡</sup>,  
STEPHEN E. LEWIS<sup>¶</sup>, ZOË T. BAINBRIDGE<sup>¶</sup> AND JON E. BRODIE<sup>¶</sup>

*Commonwealth Scientific and Industrial Research Organisation (CSIRO)\*,  
University of Missouri<sup>†</sup>, Department of Science Information Technology and  
Innovation<sup>‡</sup>, Department of Natural Resources and Mines<sup>§</sup>  
and James Cook University<sup>¶</sup>*

Soil erosion and sediment transport into waterways and the ocean can adversely affect water clarity, leading to the deterioration of marine ecosystems such as the iconic Great Barrier Reef (GBR) in Australia. Quantifying a sediment load and its associated uncertainty is an important task in delineating how changes in management practices can contribute to improvements in water quality, and therefore continued sustainability of the GBR. However, monitoring data are spatially (and often temporally) sparse, making load estimation complicated, particularly when there are lengthy periods between sampling or during peak flow periods of major events when samples cannot be safely taken.

We develop a spatio-temporal statistical model that is mechanistically motivated by a process-based deterministic model called Dynamic SedNet. The model is developed within a Bayesian hierarchical modelling framework that uses dimension reduction to accommodate seasonal and spatial patterns to assimilate monitored sediment concentration and flow data with output from Dynamic SedNet. The approach is applied in the Upper Burdekin catchment in Queensland, Australia, where we obtain daily estimates of sediment concentrations, stream discharge volumes and sediment loads at 411 spatial locations across 20 years. Our approach provides a method for assimilating both monitoring data and modelled output, providing a statistically rigorous method for quantifying uncertainty through space and time that was previously unavailable through process-based models.

**1. Introduction.** The Great Barrier Reef (GBR) off the northeastern coast of Queensland, Australia, is the largest reef system in the world, with an ecosystem spanning an area of 350,000 km<sup>2</sup>, containing 3000 individual reefs having a

---

Received June 2015; revised June 2016.

<sup>1</sup>Supported by the CSIRO through the Julius Award.

*Key words and phrases.* Water quality, Bayesian hierarchical model, SedNet, catchment modelling, spatio-temporal.

combined area of 24,000 km<sup>2</sup>. Listed as a World Heritage Site, the GBR is an Australian icon that adds billions of dollars to the Australian economy through tourism, commercial fishing and agriculture, among other industries [Furnas (2003)]. The GBR catchment area is extensive (~430,000 km<sup>2</sup>) and diverse, covering dry and wet tropical landscapes with a variety of agricultural uses (including forests, grazing, horticulture and sugarcane) [Brodie et al. (2010)]. As noted by De'ath et al. (2012), both global and local disturbances affect the health of the reef. One regional factor is increased suspended sediment in the water column leading to reduced water clarity and a decline in coral biodiversity and seagrass condition. There are several ways to transport suspended sediment to the reef, one of which is from soil erosion induced by overland flow [see Bartley et al. (2014) and the references therein]. Sediment discharge from rivers flowing into the GBR lagoon can have an adverse effect on the marine ecosystem [Brodie et al. (2011), Fabricius et al. (2014)]. There is evidence suggesting that the amount of sediment from overland flow has increased over the last 150 years due to rapid agricultural expansion and clearing [Furnas (2003), McCulloch et al. (2003)]. The Australian and Queensland governments, through the Reef Water Quality Protection Plan (2013), call for a 20% reduction in sediment discharged into the GBR lagoon by 2020. To achieve this, methods for quantifying sediment loads entering into the lagoon with a high level of confidence are required to determine the main sources of erosion and whether changes in land management practices are effective for remediation [Bartley et al. (2010), Brodie et al. (2012), Wilkinson et al. (2013)].

Load quantification methods are well documented and summarised within the literature [Cooper and Watts (2002), Letcher et al. (2002)], ranging from empirically-based approaches to more complex, physically-based models. Popular data-driven methods include ratio estimators and linear interpolation schemes [Littlewood and Marsh (2005)]. Whilst such methods offer a simple solution, they cannot quantify the uncertainties associated with concentration and flow rates, nor make use of the vast knowledge base that exists in the catchment modelling community around the physical processes governing load generation. Rating curve methods provide a mechanism for estimating the concentration of a contaminant (e.g., sediment) given flows, while accommodating uncertainties associated with measurement error in flow and concentration in a regression-based framework [Cohn (1995), Cohn et al. (1992), Walling and Webb (1985)]. Various extensions to these approaches have led to improvements in the estimation of the load and the explicit incorporation of uncertainty through Monte Carlo simulation [Crawford (1991)], bootstrapping [Rustomji and Wilkinson (2008)], semiparametric approaches incorporating uncertainty associated with the hydrological process [Kuhnert et al. (2012), Wang, Kuhnert and Henderson (2011)] and Bayesian methods [Le Coz et al. (2014), Moyeed and Clarke (2005), Reitan and Petersen-Øverleir (2011)]. Yet these approaches are all site based and do not allow estimates to borrow strength across multiple sites in space through time. They also do not incorporate hydrological process uncertainty, with exception to the previously mentioned methods of Kuhnert et al. (2012), Wang, Kuhnert and Henderson (2011).

There exists a number of popular physically based process models used for quantifying loads. In contrast to the statistical approaches outlined above, these models are deterministic in nature and generally require additional statistical tools to quantify errors and uncertainties. For a thorough review of sediment transport models, we refer the reader to [Aksoy and Kavvas \(2005\)](#). Over the last decade in Australia, the catchment modelling community has developed a number of process-based models that can be used for estimating sediment loads specifically suited to the Australian landscape. The two main models employed in this arena are SedNet and Dynamic SedNet (D-SedNet), with the latter model being implemented through the Source software modelling platform [[Armour, Hateley and Pitt \(2009\)](#), [Wilkinson et al. \(2014\)](#)]. The strength of these catchment models lies in providing an understanding of the links between the processes that generate runoff, erosion and the sediment load. However, the output of such models may not always agree with observed flow and sediment concentrations seen in field monitoring. For example, [Wilkinson et al. \(2014\)](#) state that their model may under-predict loads following drought years. Moreover, measures of uncertainty are not readily available within this modelling framework. Two of the more popular approaches for assessing uncertainty in hydrological modelling are the Parameter ESTimation (PEST) tool [e.g., [Doherty and Johnston \(2003\)](#)] and the generalised likelihood uncertainty estimation (GLUE) [e.g., [Beven and Freer \(2001\)](#)]. PEST focuses on estimation through the regularization of the parameter space, seeking to minimise uncertainty in the parameters. However, the approach makes the simplifying assumptions that model outputs are a linear function of model parameters and differentiability with respect to the parameters [[Doherty and Hunt \(2009\)](#)]. Applications of PEST also often rely on calibrating a model to some objective function defined by the user. GLUE focuses on specification of a “less formal likelihood.” However, it has recently been strongly criticised for inconsistent and paradoxical results in models with a complex error structure [[Mantovan and Todini \(2006\)](#)].

A powerful alternative modelling framework for the quantification of uncertainty is the Bayesian paradigm. Unlike frameworks like GLUE, Bayesian statistics has a rigorous probabilistic basis, and has broad appeal in the environmental sciences [[Clark \(2005\)](#)], not just in hydrology. Examples of Bayesian studies in the context of hydrological modelling include pollutant load models in [Liu et al. \(2008\)](#), [Chen et al. \(2012\)](#), and sediment models as described in [Schmelter, Hooten and Stevens \(2011\)](#) and [Pagendam et al. \(2014\)](#). In particular, Bayesian hierarchical modelling (BHM) provides a framework for assimilating process-based model output with monitoring data, whilst acknowledging errors and uncertainties in both information sources [[Berliner \(1996\)](#), [Cressie and Wikle \(2011\)](#), [Wikle and Berliner \(2007\)](#)]. Specific examples of this approach include the following: [Wu, Clark and Vose \(2010\)](#) who blended monitoring data of temperature, precipitation, soil moisture and streamflow with the output of the GR4J model for streamflow [[Perrin, Michel and Andréassian \(2003\)](#)]; and [Pagendam et al. \(2014\)](#) who utilised

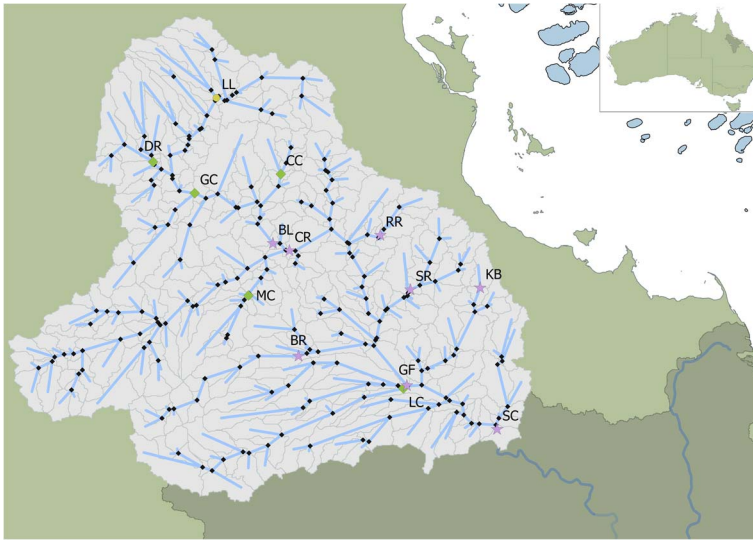


FIG. 1. Map of the Upper Burdekin catchment with 411 spatial sites of interest with *D-SedNet* output corresponding to the 411 drainage regions. The insert map in the upper right shows the entire Burdekin catchment (dark) in relation to the rest of Australia. Pink stars indicate 8 locations with both concentration and flow monitoring data (BL indicates Blue Range, RR indicates Running River, KB indicates Keelbottom River, SR indicates Star River, CR indicates Clarke River, BR indicates Basalt River, GF indicates Gainsford, SC indicates Sellheim). The yellow square shows 1 site with only flow monitoring data (LL indicates Lucy Lake). Green diamonds indicate 5 locations with only concentration monitoring data (LC indicates Lolworth Creek, MC indicates Maryvale Creek, CC indicates Camel Creek, GC indicates Grey Creek, DR indicates Dry River). The black line indicates the rest of the Burdekin River from the end-of-catchment Sellheim site through the Burdekin catchment to the Great Barrier Reef (blue) lagoon.

the BHM framework by blending observations of flow and sediment concentration at a site within the Burdekin catchment on the GBR with the SIMHYD rainfall-runoff model [Chiew, Peel and Western (2002)] and a mechanistically-motivated contaminant generation model [Kuhnert et al. (2012), Wang, Kuhnert and Henderson (2011)].

The goal of this manuscript is to develop a methodology that quantifies sediment loads while incorporating multiple sources of information and accounting for spatial and multiscale temporal dependencies. This paper continues the work of Pagendam et al. (2014) by modelling sediment loads quantification at 411 sites over 20 years in the Upper Burdekin through the BHM framework (see Figure 1). The difficulties overcome by Pagendam et al. (2014), such as missing data, nonstationarity of the process, high dimensionality and extreme events, are compounded by modelling sediment loads through space and time. This becomes especially apparent in the sparse monitoring data which are not missing at random. Process-based model output can help inform the true state of nature at locations and times without monitoring data, but such “data” will also have added uncertainty. Ideally,

multiple process-based model output runs would occur to help reduce uncertainty, but due to various issues (such as financial or computational time), as in our case, only one model run was available. Therefore, to inform the underlying true sediment load, we fuse flow and concentration monitoring data with the respected outputs arising from one run of the process-based model. Due to the variable nature of flow and concentration arising from extreme events, our model also needs to accommodate nonstationarity in space and in time. Furthermore, stream-network systems such as the Upper Burdekin can exhibit multiscale temporal behaviour that extends beyond a simple autoregressive structure. Thus, a novel component in our model is the reduced rank formulation that uses a basis expansion in both space and time. This has the added benefit of alleviating the curse of dimensionality common in such spatio-temporal processes.

The remainder of the paper proceeds as follows. Section 2 describes the computer model output and monitoring data used in estimating sediment loads in the Upper Burdekin catchment. Section 3 introduces the specifics of the model, including transformations of the data, implementation of mechanistic information and adaptations of the Bayesian hierarchical modelling framework. Section 4 applies the methodology developed to the Upper Burdekin catchment. We conclude with a discussion in Section 5 and make recommendations on how this methodology can be used to inform decision-making by managers to assist catchment remediation.

**2. Description of the data.** Cohn (1995) defines a sediment load  $L$  as the mass of sediment transported past a location over a specified time period; that is, for a given time interval  $[t_1, t_2]$ ,  $L = \int_{t_1}^{t_2} c(t)q(t) dt$ , where  $c(t)$  represents sediment concentration (measured as mass per unit volume) and  $q(t)$  represents flow (measured as volume per unit time). As the load  $L$  cannot be measured directly, we estimate it through discrete measurements of the continuous-time processes for concentration  $c(t)$  and stream flow  $q(t)$ . Since the load is computed as an integral and the observations are measured discretely, the computation is approximated by utilising the total daily flow  $v_i$  and the average daily concentration  $\hat{c}_i$ , with  $L \approx \sum_{i=1}^n v_i \hat{c}_i$ . Both the continuous and discrete-time representations of the load cannot be calculated if missing values are present for either the concentration or the flow. Estimating a load is therefore reliant on modelling these quantities where gaps exist in monitoring data, which is why process-based models have emerged as popular tools for modelling sediment loads. Ideally, the estimated load should be based on all available information. BHMs can be considered to achieve this in a statistically rigorous framework.

The Burdekin catchment, located in northeastern Australia, spans an area of approximately 130,000 km<sup>2</sup> and discharges the largest annual sediment load into the GBR lagoon [Bartley et al. (2014)]. The primary land use within the Burdekin catchment is cattle grazing, which accounts for 95% of land use [Furnas (2003)]. We focus on the Upper Burdekin sub-catchment, one of 6 sub-catchments in the

Burdekin, which comprises an area of approximately 36,000 km<sup>2</sup> (see Figure 1). We utilised daily data over a 20 year period from 1 July, 1988, through 30 June, 2008. For our purposes, we ignore leap days, and define a year as consisting of 365 days. Furthermore, our years are based on the Australian financial year (1 July–30 June) that coincides with the reporting of loads [Reef Water Quality Protection Plan, 2013]. This time period corresponds to computer model output for the Burdekin D-SedNet model.

*2.1. The burdekin D-SedNet model.* The Queensland government has invested heavily in the development of D-SedNet within the Source modelling framework for estimating loads from GBR catchments. In the work presented herein, we make use of one such D-SedNet model run for the Upper Burdekin catchment using daily observations. The model makes use of spatial grids of daily rainfall and potential evapotranspiration (interpolated from weather station information) to predict different sources of streamflow (e.g., runoff versus baseflow) using the SIMHYD rainfall runoff model [Wilkinson et al. (2013), Wilkinson et al. (2014)]. Surface runoff is then used in conjunction with constituent generation models to predict sediment loads derived from hillslope, gully and bank erosion. D-SedNet models each of these erosion sources through processes associated with the landscape and specific erosion rate properties. For example, the hillslope erosion generation model uses a spatial Revised Universal Soil Loss Equation (RUSLE) model [Renard et al. (1997)] that accounts for the spatial properties, such as slope length and gradient, erosion control practices, erodability of the soil and vegetation cover. The gully erosion generation model uses similar spatial properties, including the proportion of area and depth of gullies associated with the stream network contributing area, cross-sectional area, gully density, bulk density of the soil, and silt and clay proportions of the gully soil, among other factors. D-SedNet models sediment generated from bank erosion through bank risk factors while accounting for floodplain deposition and channel deposition or remobilisation. Bank risk factors include the presence and effectiveness of riparian vegetation and erodibility of the bank soil. Generated loads are transferred with runoff to the stream network, where loads and flows are accessible, thus allowing calculation of concentrations. For the mathematical details of the D-SedNet model, we refer the reader to Wilkinson et al. (2014).

*2.2. Monitoring in the upper burdekin catchment.* In general, the spatial coverage of monitoring sites across GBR catchments is sparse, with potentially substantial variability in monitoring frequency. Flow data are measured at fixed gauging stations, with measurements typically recorded every 10–15 minutes, which can be used to compute a daily flow volume. Sediment concentration is measured much more infrequently and focuses on high-flow events where the majority of the mass transport takes place. At present, monitoring concentration and flow at regular intervals over a spatially dense network is economically and logistically

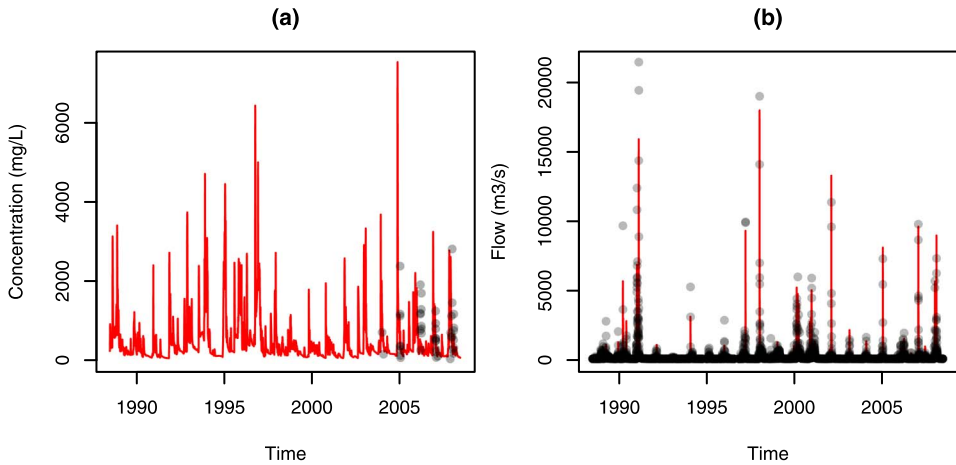


FIG. 2. Time series plots of the D-SedNet model output (red line) and observations (transparent black points) for the Sellheim site for (a) concentration in mg/L, and (b) flow in  $\text{m}^3/\text{s}$ .

infeasible. Monitoring locations are therefore chosen according to ease of access and the cost associated with monitoring either using automatic loggers or manual sampling. Figure 1 shows a river network for the Upper Burdekin sub-catchment consisting of 411 links, connected by nodes where water flows from one stretch of water to another. As highlighted by Figure 1, monitoring was performed at nine links for flow and thirteen links for concentration. Only eight monitoring sites had both flow and concentration [Bainbridge et al. (2014)]. Flow monitoring data were available for the majority of the time period of interest, but concentration monitoring data were available primarily from 2004 through 2008. Figure 2 shows the available daily monitoring data for sediment concentration and flow for the end-of-catchment site, the Sellheim site. Additional figures for sites with monitored flow or concentration are found in the online supplementary material [Gladish et al. (2016)]. Furthermore, the number of available days with monitoring data are found in Table 1. Note that sediment concentration has much fewer total monitored days (176) compared to flow (41,682). Furthermore, 41,551 days were monitored for flow only, 45 for concentration, and 131 for both concentration and flow.

**3. Methods.** In developing our model, we account for spatial variability and temporal variability on both yearly and daily scales. To achieve this, we draw upon the approach developed by Oleson and Wikle (2013) and develop a two-stage dimension reduction approach of both the temporal and spatial domains. Specifically, we assimilate monitoring data with D-SedNet output through two data models, both of which are conditioned on the underlying “true” process. This accommodates the sparse nature of monitoring as highlighted in Section 2 through “in-filling” of missing monitoring observations while still accounting for uncertainty associated at unmonitored sites and days. We then project this process onto a latent

TABLE 1

*Number of days for sites where at least either concentration or flow was monitored, with the total number of monitored days across all sites. Each site has a maximum 7300 possible days of monitoring*

Site	Concentration	Flow
Sellheim	45	7042
Star River	10	779
Keelbottom Creek	36	5125
Running River	7	6478
Blue Range	10	7300
Gainsford	10	1486
Basalt River	7	7059
Clarke River	14	379
Camel Creek	4	0
Dry River	6	0
Grey Creek	9	0
Lolworth Creek	17	0
Maryvale Creek	1	0
Lake Lucy	0	6034
Total	176	41,682

spatial process through a spectral seasonal decomposition with in-season varying error structure. This spatial process is then projected onto another spectral process through a spatial decomposition while influenced by covariates. This second hidden process is then modelled dynamically. The following sections outline the specifics of our approach which are summarised in Table 2.

3.1. *Data model.* We utilise two data sources to inform the underlying latent process: (i) the monitoring data (field observations); and (ii) the D-SedNet model output. Note, as the scale of variability in flow and concentration data increases greatly as one travels downstream due to the accumulation of water and sediment from upstream tributaries, we rescale the monitoring data and D-SedNet model output by the terminal node of the stream network (herein referred to as the Sellheim site) over the entire study period. We further apply a Box–Cox transformation [Box and Cox (1964)] to accommodate positively skewed data resulting from this rescaling. Box–Cox-like transformations have been successfully used to model similar processes [e.g., see Sansó and Guenni (1999); Cressie and Wikle (2011), page 380]. More formally, we define the Box–Cox transformation as

$$z(\mathbf{s}_i; t_{j,k}) = \frac{((z^*(\mathbf{s}_i; t_{j,k}) + \rho_2)^{\rho_1} - 1)}{\rho_1},$$

where  $z^*(\mathbf{s}_i; t_{j,k})$  represents the data (either flow or concentration rescaled with respect to Sellheim) at spatial location ( $i = 1, \dots, n$ ) and day  $t_{j,k} \equiv (k - 1)T + j$



TABLE 2

Specification of the data, process and parameters models of the Bayesian hierarchical model. Note,  $q_\varepsilon(j)$ ,  $r_\varepsilon(j)$ ,  $q_\gamma(l)$ ,  $r_\gamma(l)$ ,  $q_\eta(l)$ ,  $r_\eta(l)$ ,  $\lambda_0$ ,  $\sigma_\lambda^2$ ,  $\mu_0$ ,  $\Sigma_0$ ,  $\mu_M$  and  $\Sigma_M$  are prespecified hyperparameters. Note further that  $\mathbf{H}_{ik}^{\text{obs}}$  is an incidence matrix for the field observations. The variance hyperparameters  $q_\varepsilon(j)$  and  $r_\varepsilon(j)$  for  $j = 1, \dots, T$ ,  $q_\gamma(l)$  and  $r_\gamma(l)$  for  $l = 1, \dots, p$ , and  $q_\eta(l)$  and  $r_\eta(l)$  for  $l = 1, \dots, q$  are all set at 0.1. The mean hyperparameters  $\lambda_0$ ,  $\mu_0$  and  $\mu_M$  are all set at  $\mathbf{0}$ . The hyperparameters  $\sigma_\lambda^2$ ,  $\Sigma_0$  and  $\Sigma_M$  are set at 100,  $100\mathbf{I}$  and  $100\mathbf{I}$ , respectively

Stage	Model
Data model:	$\mathbf{Z}_{ik}^{\text{obs}} = \mathbf{H}_{ik}^{\text{obs}} \mathbf{Y}_{ik} + \mathbf{e}_{ik}^{\text{obs}}, \mathbf{e}_{ik}^{\text{obs}} \sim \text{Gau}(\mathbf{0}, \sigma_{\text{obs}}^2 \mathbf{I}),$ $\mathbf{Z}_{ik}^{\text{src}} = \mathbf{Y}_{ik} + \mathbf{e}_{ik}^{\text{src}}, \mathbf{e}_{ik}^{\text{src}} \sim \text{Gau}(\mathbf{0}, \Sigma_{ik}^{\text{src}}),$
Process model:	$\mathbf{Y}_{ik} = \Psi \alpha_{ik} + \varepsilon_{ik}, \varepsilon_{ik} \sim \text{Gau}(\mathbf{0}, \Sigma_\varepsilon),$ $\tilde{\alpha}_k = \Phi \beta_k + \lambda \otimes \mathbf{X}_k + \gamma_k, \gamma_k \sim \text{Gau}(\mathbf{0}, \Sigma_\gamma),$ $\beta_k = \mathbf{M} \beta_{k-1} + \eta_k, \eta_k \sim \text{Gau}(\mathbf{0}, \Sigma_\eta),$
Parameter model:	$\sigma_\varepsilon^2(j) \sim \text{IG}(q_\varepsilon(j), r_\varepsilon(j)), j = 1, \dots, T,$ $\sigma_\gamma^2(l) \sim \text{IG}(q_\gamma(l), r_\gamma(l)), l = 1, \dots, p,$ $\sigma_\eta^2(l) \sim \text{IG}(q_\eta(l), r_\eta(l)), l = 1, \dots, q,$ $\lambda \sim \text{Gau}(\lambda_0, \sigma_\lambda^2 \mathbf{I}),$ $\beta_0 \sim \text{Gau}(\mu_0, \Sigma_0),$ $\mathbf{m} \equiv \text{diag}(\mathbf{M}) \sim \text{Gau}(\mu_M, \Sigma_M),$ $\sigma_{\text{obs}}^2$ is estimated from the data and fixed for both flow and concentration, $\Sigma_{ik}^{\text{src}}$ is estimated from the data and fixed for both flow and concentration.

( $j = 1, \dots, T$  and  $k = 1, \dots, K$ ), where  $D_s$  is the spatial domain of the stream network,  $i$  is an index over sites,  $j$  is an index over days, and  $k$  is an index over years. For our purposes,  $T = 365$ , as we ignore leap days, and  $K = 20$ . Further,  $\rho_1$  is a power transformation parameter and  $\rho_2$  is a shift parameter set a priori. We estimate  $\rho_1$  empirically for concentration and flow based on the profile likelihood of the D-SedNet output over all sites and times. Additionally,  $\rho_2$  is set at half the minimum nonzero value of  $z^*(\mathbf{s}_i; t_{j,k})$  over all sites and times.

We define  $z_{ijk}^d \equiv z^d(\mathbf{s}_i; t_{j,k})$  as the transformed data (either flow or concentration) for  $d \in \{\text{obs}, \text{src}\}$ , where obs corresponds to the monitoring data and src corresponds to the D-SedNet model output. We then define  $\mathbf{Z}_{ik}^d$  as the vector of length  $m_{ik}^d$  with elements  $z_{ijk}^d$ , where  $m_{ik}^d$  is the number of days with data for location  $i$ , year  $k$ . Note that  $\mathbf{Z}_{ik}^{\text{obs}}$  will be sparse (often with  $m_{ik}^{\text{obs}} \equiv 0$ ), while  $\mathbf{Z}_{ik}^{\text{src}}$  will have no missing values (with  $m_{ik}^{\text{src}} \equiv T$ ). The general data models can be written as

$$(1) \quad \begin{aligned} \mathbf{Z}_{ik}^{\text{obs}} &= \mathbf{H}_{ik}^{\text{obs}} \mathbf{Y}_{ik} + \mathbf{e}_{ik}^{\text{obs}}, & \mathbf{e}_{ik}^{\text{obs}} &\sim \text{Gau}(\mathbf{0}, \sigma_{\text{obs}}^2 \mathbf{I}), \\ \mathbf{Z}_{ik}^{\text{src}} &= \mathbf{Y}_{ik} + \mathbf{e}_{ik}^{\text{src}}, & \mathbf{e}_{ik}^{\text{src}} &\sim \text{Gau}(\mathbf{0}, \Sigma_{ik}^{\text{src}}), \end{aligned}$$

where  $\mathbf{Y}_{ik} \equiv (Y_{i1k}, \dots, Y_{iT k})'$  is the  $T$ -dimensional vector corresponding to the latent process,  $\mathbf{H}_{ik}^{\text{obs}}$  is an incidence (zero/one) matrix accounting for potential miss-

ing data in the monitoring data, and  $\sigma_{\text{obs}}^2 \mathbf{I}$  and  $\Sigma_{ik}^{\text{src}}$  are covariance matrices corresponding to deviations between the true underlying latent process and each of the two types of data. We specify the structure of  $\Sigma_{ik}^{\text{src}}$  in equation (1) differently for the concentration and flow models. For flow,  $\Sigma_{ik}^{\text{src}} \equiv \sigma_{\text{src}}^2 \mathbf{I}$ , and, for concentration,  $\Sigma_{ik}^{\text{src}}$  is a time-varying diagonal matrix with different variances for each element of the vector  $\mathbf{Z}_{ik}^{\text{src}}$  (details of which are given in Section 3.5).

3.2. *Seasonal process model.* We use a reduced rank approach to model the latent process as a function of the spectra of the spatio-temporal dynamics for flow and concentration. Specifically, we model the dynamics of these processes as a projection onto spectral basis functions by means of a truncated Karhunen–Loève (K–L) expansion of the D-SedNet model output, with the dynamics in the latent processes modelled through the random spectral coefficients [see Cressie and Wikle (2011), Section 7.2.6]. Since the hydrological processes under investigation follow a seasonal pattern, we first use a spectral representation of the dynamics within years that is common across all sites. The resulting seasonal spectral expansion coefficients vary across space and can be treated as a latent spatial process, and can be modelled in a reduced rank setting by undertaking a second truncated K–L expansion of the spatial processes. This two-tiered dimension reduction approach is based on the methodology adopted by Oleson and Wikle (2013), and precise details of our implementation are provided in Section 3.8. Under this approach, we model  $\mathbf{Y}_{ik}$  as

$$(2) \quad \mathbf{Y}_{ik} = \Psi \alpha_{ik} + \boldsymbol{\varepsilon}_{ik},$$

where  $\Psi$  is a  $T \times p$  matrix of seasonal (yearly) spectral basis functions,  $\alpha_{ik} \equiv (\alpha_{ik}(1), \dots, \alpha_{ik}(p))'$  is a  $p$ -dimensional vector of expansion coefficients,  $p$  is the number of basis vectors retained when truncating the K–L expansion, and  $\boldsymbol{\varepsilon}_{ik}$  is a zero-mean Gaussian error process independent of  $\alpha_{ik}$ . For our purposes, we let  $\boldsymbol{\varepsilon}_{ik} \sim \text{Gau}(\mathbf{0}, \Sigma_{\varepsilon})$ , where  $\Sigma_{\varepsilon} \equiv \text{diag}(\sigma_{\varepsilon,1}^2, \dots, \sigma_{\varepsilon,T}^2)$ . Critically, this error structure accommodates heteroskedasticity in the error process over the course of a year, allowing for larger errors in the highly dynamic wet season compared to the more stable dry season.

3.3. *Spatial process model.* As pointed out by Oleson and Wikle (2013), the resulting  $\alpha_{ik}$  values are a multivariate spatial field of expansion coefficients over  $K$  years. As such, we reformulate the  $\alpha_{ik}$  process at year  $k$  by constructing a  $p \times n$  matrix  $\mathbf{A}_k$  whose columns are  $\alpha_{ik}$ ; that is,

$$\mathbf{A}_k = (\alpha_{1k}, \dots, \alpha_{nk}).$$

We then define  $\tilde{\boldsymbol{\alpha}}_k$  as the vectorisation of the transpose of  $\mathbf{A}_k$ ,  $\tilde{\boldsymbol{\alpha}}_k \equiv \text{vec}(\mathbf{A}_k')$ . The  $np$ -dimensional vector  $\tilde{\boldsymbol{\alpha}}_k$  is a multivariate spatio-temporal process. Following

Oleson and Wikle (2013), we project this process onto a second rank-reduced basis function expansion (by means of another truncated K–L expansion). We model the yearly dynamics in these spatial fields of spectral coefficients as

$$(3) \quad \tilde{\alpha}_k = \Phi \beta_k + \lambda \otimes \mathbf{X}_k + \boldsymbol{\gamma}_k,$$

where  $\beta_k$  is a  $q$ -dimensional latent random effects process. In this expression,  $q$  is the number of basis functions retained through the truncation of the K–L expansion and  $\Phi$  is a  $np \times q$  matrix of multivariate spatial basis functions. The formulation in equation (3) allows for the inclusion of exogenous explanatory variables that may have some relationship to the latent hydrological process ( $\mathbf{Y}_{ik}$ ). In this expression,  $\mathbf{X}_k$  is an  $n$ -dimensional vector of (potentially spatial) exogenous variables at year  $k$  and  $\lambda$  is a vector of coefficients of length  $p$  whose  $l$ th component  $\lambda_l$  corresponds to the  $l$ th spatial basis function. We further assume the error process  $\boldsymbol{\gamma}_k$  is zero-mean with  $\boldsymbol{\gamma}_k \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\gamma)$ , and independent of  $\beta_k$  and  $\lambda$ . By using spatial basis functions  $\Phi$ , we assume that  $\Phi \beta_k$  captures the spatial variability in  $\tilde{\alpha}_k$  and, furthermore, that  $\Phi \beta_k$  captures the codependency between the components of  $\alpha_{ik}$ . Consequently, we use a diagonal error structure for  $\boldsymbol{\gamma}_k$  with individual errors for each spatial field, whereby  $\boldsymbol{\Sigma}_\gamma = \text{diag}(\sigma_\gamma^2(1), \dots, \sigma_\gamma^2(p)) \otimes \mathbf{I}_n$ . Critically, through using spatial basis functions, we are able to model the spatial dependence of the process in a reduced rank setting, avoiding the computational costs associated with high dimensionality.

3.4. *Process model dynamics.* The underlying dynamics of  $\tilde{\alpha}_k$  are propagated through a first-order Markov process

$$\beta_k = \mathcal{M}(\beta_{k-1}; \boldsymbol{\theta}; \eta_k),$$

which evolves the coefficients  $\beta_k$  according to some function  $\mathcal{M}$ . This expression depends on the coefficients of the previous time  $\beta_{k-1}$ , some underlying error process  $\eta_k$  and parameters  $\boldsymbol{\theta}$ . For our purposes, we follow the formulation of Oleson and Wikle (2013) and utilise a first-order linear autoregressive structure, but we note that more complex formulations [such as the generalised quadratic nonlinear framework outlined by Wikle and Hooten (2010)] are also possible candidate models. Specifically, we follow the formulation, for  $k = 1, \dots, K$ ,

$$\beta_k = \mathbf{M}\beta_{k-1} + \eta_k,$$

where  $\eta_k \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ ,  $\mathbf{M}$  is a  $q \times q$  propagator matrix, and  $\boldsymbol{\Sigma}_\eta$  is a covariance matrix associated with the autoregressive model. We make the further simplifying assumptions that  $\mathbf{M}$  and  $\boldsymbol{\Sigma}_\eta$  are both diagonal matrices, though a general vector autoregressive structure could be used as well. We specify  $\mathbf{m} \equiv \text{diag}(\mathbf{M})$  and  $\sigma_\eta^2(l)$  as element  $(l, l)$  of the matrix  $\boldsymbol{\Sigma}_\eta$ . The initial dynamics are specified by  $\beta_0 \sim \text{Gau}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ , where  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\Sigma}_0$  are prespecified hyperparameters.

3.5. *Parameter model.* The parameter model provides an avenue for including prior information (from previous studies or expert opinion) about parameters into the current study through prior distributions. In our analyses, we use conjugate prior distributions in order to use Gibbs sampling from the full conditional distributions [Geman and Geman (1984)]. The conjugate priors used in this study were as follows:

$$\begin{aligned}\sigma_\varepsilon^2(j) &\sim \text{IG}(q_\varepsilon(j), r_\varepsilon(j)), & j = 1, \dots, T, \\ \sigma_\gamma^2(l) &\sim \text{IG}(q_\gamma(l), r_\gamma(l)), & l = 1, \dots, p, \\ \boldsymbol{\lambda} &\sim \text{Gau}(\boldsymbol{\lambda}_0, \sigma_\lambda^2 \mathbf{I}), \\ \mathbf{m} &\sim \text{Gau}(\boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M), \\ \sigma_\eta^2(l) &\sim \text{IG}(q_\eta(l), r_\eta(l)), & l = 1, \dots, q,\end{aligned}$$

where  $q_\varepsilon(j), r_\varepsilon(j), q_\gamma(l), r_\gamma(l), \boldsymbol{\lambda}_0, \sigma_\lambda^2, \boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M, q_\eta(l)$  and  $r_\eta(l)$  are all prespecified hyperparameters.

3.6. *Data model variance specification.* Due to potential nonidentifiability, some parameters need to be fixed. With sparse monitoring data, it is unlikely that the model will inform  $\sigma_{\text{obs}}^2$ . However, we can elicit this value using previous work. Further, we also assume the data model variance from D-SedNet,  $\boldsymbol{\Sigma}_{ik}^{\text{src}}$ , is set a priori. We compute the measurement error variance for flow observations,  $\sigma_{\text{obs},Q}^2$ , by first obtaining deviations between gaugings and flow rating curves [similar in approach to Tomkins (2014)] on the transformed scale. Note that the rescaling and Box–Cox transformations that were applied to the data (outlined in Section 3.1) resulted in relatively homoscedastic deviations on the transformed scale, which made our use of a single variance parameter across sites possible. For the error variance for D-SedNet modelled flow,  $\boldsymbol{\Sigma}_{ik}^{\text{src},Q} \equiv \sigma_{\text{src},Q}^2 \mathbf{I}$ , we use a similar approach in estimating  $\sigma_{\text{src},Q}^2$ , but using the deviations between gaugings and the corresponding modelled flow at the location and time that the gauging was recorded to estimate this parameter.

The measurement error variance in sediment concentration,  $\sigma_{\text{obs},C}^2$ , is estimated using a set of sediment concentration monitoring data that is supplementary to the concentration data used in the rest of the analysis. These data were collected in adjacent sub-catchments on the Belyando, Burdekin, Cape and Suttor Rivers [see Bainbridge et al. (2014)]. At each of these sites three samples were collected at each of three different locations: the right bank, the left bank and the middle of the stream. A random effects linear model of the form

$$z_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$$

was then fit, where  $z_{ijk}$  represented the transformed observed concentration at river  $i$ , location  $j$  (i.e., right, left and middle), and sample  $k$ ,  $\mu$  is the grand mean,  $\alpha_i$

is a random effect for river  $i$ ,  $\beta_{ij}$  is a random effect for location  $j$  within river  $i$ , and  $\varepsilon_{ijk} \sim \text{Gau}(0, \sigma_{\text{obs},C}^2)$  is the residual. The estimate of the residual error variance from the analysis was then used for  $\sigma_{\text{obs},C}^2$  and quantifies the unexplained variability that is exhibited in sediment concentration samples in the field.

Last, we estimated the error variance for sediment concentration  $\Sigma_{ik}^{\text{src},C}$  using our estimate of the error variance of D-SedNet flow output ( $\sigma_{\text{src},Q}^2$ ) and the deterministic relationship  $C_{\text{src}}^{\text{nat}} = L_{\text{src}}^{\text{nat}}/Q_{\text{src}}^{\text{nat}}$ , where  $C_{\text{src}}^{\text{nat}}$  is the daily concentration,  $L_{\text{src}}^{\text{nat}}$  is the daily load, and  $Q_{\text{src}}^{\text{nat}}$  is the daily flow volume (all from D-SedNet and all on the natural scale). We then employ the delta method to approximate the elements of  $\Sigma_{ik}^{\text{src},C}$  using the transformed daily outputs from D-SedNet:

$$\text{Var}(C_{\text{src}}) \approx \left( \frac{\partial C_{\text{src}}}{\partial Q_{\text{src}}} \right)^2 \text{Var}(Q_{\text{src}}),$$

where  $C_{\text{src}}$ ,  $L_{\text{src}}$  and  $Q_{\text{src}}$  are the daily concentration, load and flow volume from D-SedNet on the transformed scale. As discussed above, the variance in D-SedNet flow has been quantified on a transformed scale, following a rescaling of the data and a subsequent Box–Cox transformation. It is easy to verify that

$$\begin{aligned} \frac{\partial C_{\text{src}}}{\partial Q_{\text{src}}} &= -(L_{\text{src}} \rho_3^Q \rho_3^C ((\rho_1^Q Q_{\text{src}} + 1)^{1/\rho_1^Q} - \rho_2^Q)^{-1} + \rho_2^C) (\rho_1^C - 1) \\ &\quad \times -L_{\text{src}} \rho_3^Q \rho_3^C ((\rho_1^Q Q_{\text{src}} + 1)^{1/\rho_1^Q} - \rho_2^Q)^{-2} \\ &\quad \times (\rho_1^Q Q_{\text{src}} + 1)^{(1/\rho_1^Q - 1)}, \end{aligned}$$

where  $\rho_1^Q$ ,  $\rho_2^Q$ ,  $\rho_1^C$  and  $\rho_2^C$  represent the Box–Cox parameters used for flow ( $Q$ ) and concentration ( $C$ ), respectively, and  $\rho_3^Q$  and  $\rho_3^C$  are parameters used in the end-of-catchment rescaling discussed in Section 3.1 (i.e.,  $Q_{\text{src}} = \rho_3^Q Q_{\text{src}}^{\text{nat}}$  and  $C_{\text{src}} = \rho_3^C C_{\text{src}}^{\text{nat}}$ ). In practice, the delta method potentially results in some elements of  $\Sigma_{ik}^{\text{src},C}$  lower than  $\sigma_{\text{obs},C}^2$ , indicating that there was less uncertainty in D-SedNet output than in observations. We note that 99% of the time this occurs, D-SedNet estimates flow at less than  $1 \text{ m}^3/\text{s}$ , indicating zero to extremely low flow. Because of this and based on expert opinion, we assume that the variance of outputs from the D-SedNet model would never be lower than the variance of observations since measurements are considered more reliable than model output. Therefore, we set  $\Sigma_{ik}^{\text{src},C}$  to  $\sigma_{\text{obs},C}^2$  on days at sites where this occurred.

**3.7. Covariate specification.** We include one covariate for both the concentration and flow models. For concentration, we use C-Factor, a value that is a function of vegetative ground cover. C-factor is an input variable typically used in USLE models, such as the hillslope erosion model described in Section 2.1 [Renard et al. (1997)], and has an inverse relationship with cover. Furthermore, C-factor is a spatial covariate that varies by financial year. In the current application, C-factor was

used as a predictor of hillslope erosion in open grazing land to estimate sediment generation. Specifically, C-factor is defined as in [Rosewell \(1993\)](#) as

$$\text{C-factor} = \exp(-0.799 - 0.0474 \text{ cover} + 0.000449 \text{ cover}^2 - 0.000052 \text{ cover}^3).$$

We include a measure related to vegetative cover, due to the fact that studies show it to be a key driving factor in sediment transportation for the concentration model [e.g., [Bartley et al. \(2010\)](#)], and the annual mean Pacific Decadal Oscillation (PDO), a climate variable that potentially impacts rainfall runoff, and hence flow [[Lough, Lewis and Cantin \(2015\)](#)]. Details of these covariates are found in the supplementary material [[Gladish et al. \(2016\)](#)]. We note that, although we utilise cover and PDO, alternative exogenous covariates could be used as well as inclusion of multiple covariates for either flow or concentration.

3.8. *Basis function specification.* As outlined in the process model, our approach relies upon dimension reduction of a high-dimensional spatio-temporal modelling problem by projecting the latent processes onto a set of basis functions. Whilst a variety of such basis functions could be employed, we utilise Empirical Orthogonal Functions (EOFs) because of: (i) their successful application in a variety of spatio-temporal modelling problems [recent examples include [Brynjarsdóttir and Berliner \(2014\)](#), [Gladish and Wikle \(2014\)](#), [Graf et al. \(2014\)](#), [Wu, Holan and Wikle \(2013\)](#)]; and (ii) the truncated K–L expansion results in an approximation of the original process with the lowest possible mean square error [[Cressie and Wikle \(2011\)](#), Section 5.3]. The Karhunen–Loève expansion implemented here is for stochastic processes over a discrete spatial or temporal domain so that obtaining the basis functions is as simple as performing an eigenvalue decomposition on the empirical covariance matrix. For the first K–L expansion performed in the temporal domain, the EOFs for flow and concentration are based on output obtained from the D-SedNet model. This allows the underlying latent process to adopt the spectral properties of the D-SedNet model. We follow the same procedure as [Oleson and Wikle \(2013\)](#), and outline the details here.

To obtain  $\Psi$  for either concentration or flow, we calculate the seasonal empirical covariance matrix  $\hat{\Sigma}$ , where

$$\hat{\Sigma} = \frac{1}{nK - 1} \sum_{k=1}^K \sum_{i=1}^n (\mathbf{Z}_{ik}^{\text{src}} - \bar{\mathbf{Z}})(\mathbf{Z}_{ik}^{\text{src}} - \bar{\mathbf{Z}})',$$

where  $n$  is the number of spatial locations,  $K$  is the number of years,  $\mathbf{Z}_{ik}^{\text{src}}$  is defined as in Section 3.1, and  $\bar{\mathbf{Z}} \equiv \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n \mathbf{Z}_{ik}^{\text{src}}$ . Based on this expression,  $\hat{\Sigma}$  represents the empirical covariance matrix averaged over all 411 links in the stream network and over the  $K$  years the model was run. The first  $p$  eigenvectors of  $\hat{\Sigma}$  then form the columns of the  $T \times p$  matrix  $\Psi$ , where  $p \in \{1, \dots, T\}$  is chosen as

the smallest integer for which  $\text{argmin}_p \sum_{i=1}^p \lambda_i^\Psi / \sum_{i=1}^T \lambda_i^\Psi \geq \kappa$ , where  $\lambda_1^\Psi, \dots, \lambda_T^\Psi$  are the eigenvalues of  $\widehat{\Sigma}$  in descending order and  $\kappa \in [0, 1]$  is user specified to account for a set percentage of the variability in the output when truncating the K–L expansion.

We estimate  $\Phi$  by first estimating  $\widehat{\alpha}_{ik} \equiv \Psi' \mathbf{Z}_{ik}^{\text{src}}$ , restructuring as in equation (3) to obtain  $\widehat{\alpha}_k$ . We then determine  $\widehat{\Sigma}_{\widehat{\alpha}}$  by

$$\widehat{\Sigma}_{\widehat{\alpha}} = \frac{1}{K-1} \sum_{k=1}^K (\widehat{\alpha}_k - \widetilde{\alpha})(\widehat{\alpha}_k - \widetilde{\alpha})',$$

where  $\widetilde{\alpha} \equiv \frac{1}{K} \sum_{k=1}^K \widehat{\alpha}_k$ . The first  $q$  eigenvectors of  $\widehat{\Sigma}_{\widehat{\alpha}}$  then form the column vectors of the  $np \times q$  matrix  $\Phi$ . The truncation point  $q$  is then chosen similarly as for  $p$ .

**4. Application to the Upper Burdekin.** We apply the methodology developed in Section 3 to the 20 financial years (FY) of data in the Upper Burdekin catchment (see Section 2). Specifically, using daily D-SedNet output and monitoring data from 1 July, 1988, to 30 June, 2008, the approach is applied on 7300 days over 411 sites (Figure 1). This resulted in 3,000,300 parameters of the latent process of interest,  $Y_{ijk}$  for both flow and concentration. These can then be used to obtain estimates of sediment loads with a measure of uncertainty. Due to the number of parameters of interest, we predominantly focus on a select set of years and Sellheim, the end-of-catchment site.

*4.1. Upper burdekin implementation.* D-SedNet was initially run over the stream node-network shown in Figure 1 and described in Section 2.1, generating one D-SedNet output set for concentration and flow. Using the resulting daily sediment concentration and flow, we utilise the Box–Cox transformation described in Section 3.1. Since sediment concentration and flow from the D-SedNet output both contain zero values, we set the shift parameter  $\rho_2$  at half the minimum nonzero value, being  $\rho_2^C = 5.59 \times 10^{-17}$  for concentration and  $\rho_2^Q = 1.51 \times 10^{-12}$  for flow. The value of  $\rho_1$  is chosen from the region that maximised the profile log-likelihood from the D-SedNet output for flow and concentration. We found  $\rho_1 = 0.1$  suitable for both flow and concentration.

We first construct EOFs described in Section 3.8 for both concentration and flow to account for 80% of the variability in the D-SedNet output. EOFs for the seasonal  $\Psi$  and spatial  $\Phi$  representations of concentration are based on  $p = 13$  and  $q = 11$  components, respectively. EOFs for flow considered  $p = 8$  and  $q = 6$  components. Remaining EOFs accounted for less than 1% of the total variability in either the seasonal or spatial decompositions for both concentration and flow, and their exclusion was considered to have a negligible effect on the results.

We next set the data model variances using the methodology outlined in Section 3.5. For the concentration formulation, we set the monitoring data model variance  $\sigma_{\text{obs},C}^2 = 0.0443$ , and the D-SedNet variance  $\Sigma_{ik}^{\text{src},C}$  has values that range

from 0.0443 to 1249.879, with 95% of the values lying between 0.427 and 162.842. These values reflect variability in our confidence in D-SedNet model output and monitoring data at different stages of stream flow. For the flow model formulation, we set the monitoring data variance  $\sigma_{\text{obs},C}^2 = 0.0443$ , and the D-SedNet variance  $\sigma_{\text{src},Q}^2 = 13.414$ .

In order to sample from the posterior distribution, we implemented a Markov Chain Monte Carlo (MCMC) algorithm using Gibbs sampling of the full conditional distributions described in the Appendix. We ran three separate MCMC chains with different initial values for the parameters for 10,000 iterations, discarding the first 1000 as burn-in. All three chains converged to the same distribution quickly. Convergence was assessed by visual inspection with no evidence of lack of convergence from all 3 chains. We further thinned our chains by storing every 10th iteration to assist with the digital storage of the results. Storing every 10th iteration of all parameters sampled in the MCMC resulted in 163.82 gigabytes. The algorithm was implemented in R [R Core Team (2015)] and took approximately 72 hours to run on a single Dual Xeon 8-core CPU with 16 GB of RAM. We make a special note that sampling  $\mathbf{Y}_{ik}$  and  $\boldsymbol{\alpha}_{ik}$  was particularly slow due to the number of parameters for  $\mathbf{Y}_{ik}$  and the  $np \times np$  covariance matrix for  $\tilde{\boldsymbol{\alpha}}_k$ . We note that considerable speedup could be achieved using an alternative programming language.

*4.2. Posterior distribution results.* The dynamics of  $\mathbf{Y}_{ik}$  are modelled in the latent spectral space through the EOF decomposition. We refer the reader to the online supplementary material for the posterior distribution results of the parameters associated with the decomposition [Gladish et al. (2016)].

*4.2.1. Covariate results.* The posterior values for the covariate components are presented in the online supplementary material [Gladish et al. (2016)]. For C-factor, all of the components are significantly different from zero, while, for PDO, all but one are significantly different. This indicates C-factor and PDO significantly influence concentration and flow, respectively. However, since we model the covariates in the spectral space, care must be taken in the interpretation. As  $\boldsymbol{\lambda}$  influences the latent process  $\mathbf{Y}_{ik}$  through the seasonal EOFs  $\boldsymbol{\Psi}$ , the covariate will affect each day of the financial year differently. Therefore, we illustrate the influence of the covariate by averaging over the financial year. Note that  $X_{ik} \sum_{l=1}^P \psi_j(l)\lambda_l$  is the contribution of  $\boldsymbol{\lambda}$  to  $Y_{ijk}$ , where  $X_{ik}$  is the  $i$ th component of  $\mathbf{X}_k$ ,  $\psi_j(l)$  is the  $j, l$ th element of the matrix  $\boldsymbol{\Psi}$ ,  $\lambda_l$  is the  $l$ th component of  $\boldsymbol{\lambda}$ , and  $Y_{ijk}$  is the  $j$ th day of the latent process  $\mathbf{Y}_{ik}$ . Define  $\tilde{Y}_{ik} \equiv (1/T) \sum_j \sum_{l=1}^P \psi_j(l)\lambda_l$ . Using samples of  $\boldsymbol{\lambda}$  from the posterior distribution,  $\tilde{Y}_{ik}$  has a mean value 12.529 with a 95% credible interval (11.279, 13.692) for concentration. For flow,  $\tilde{Y}_{ik}$  has a mean value of 0.547 with a 95% credible interval (0.448, 0.644); that is, our approach shows evidence that increasing C-factor (and hence decreasing cover) and increasing PDO increase concentration and flow, respectively. It is important to note that this does not give an indication on the magnitude of  $\mathbf{Y}_{ik}$ .



4.2.2. *Latent process results.* Of primary interest to this analysis is the posterior distribution of  $\mathbf{Y}_{ik}$ , which assimilates monitoring data and D-SedNet output together with mechanistic information from the spectral parametrization. We show the results for the 1990/1991 financial year, during which there was no monitoring for concentration at Sellheim and a major rain depression associated with a cyclone crossed the Upper Burdekin, and the 2006/2007 financial year, in which monitoring data for concentration was more consistent at Sellheim and followed years of drought (Figure 3). At days where there are observations, the 95% credible interval is relatively tight compared to days without observations. Moreover, while the posterior mean for both flow and concentration tend to weight observations higher than D-SedNet output at locations with observations, we do see a blending of the two data sources together. Furthermore, we notice that there is more uncertainty in these predictions from December through to February, which coincides with the Australian wet season. Another interesting feature of Figures 3(c) and (d) is observed in December 2006, where D-SedNet estimates a large event with higher concentration, but the posterior estimate does not. Note that the D-SedNet output lies within the 95% credible interval during this time. Another interesting feature of the model is seen in Figures 3(e) and (f), where the posterior mean favours the monitored data more than the D-SedNet model output. This is reasonable, considering there is more certainty associated with  $\sigma_{\text{obs},Q}^2$  than with  $\sigma_{\text{src},Q}^2$ . This is particularly interesting given that  $\mathbf{Y}_{ik}$  is also informed by the process-based model through the seasonal basis decomposition  $\Phi$ .

4.2.3. *Sediment load results.* Using estimates of  $\mathbf{Y}_{ik}$  from the posterior distributions for both flow and concentration, we obtain an estimate of sediment load as described in Section 2. Figure 4 shows sediment load by day for the financial years 1990/1991 (top left) and 2006/2007 (top right) at the Sellheim site, where the posterior estimates of sediment load are in blue with 95% credible intervals in grey, and the estimated sediment load from D-SedNet output are in red for reference. The credible interval for the 2006/2007 financial year is much narrower when compared to the 1990/1991 financial year, and can be attributed to the lack of observations collected for concentration at the Sellheim site. We note that FY 1990/1991 was abnormal in the sense that multiple major rain events occurred, including Tropical Cyclone Joy.

For reporting purposes, catchment managers are often interested in estimates of sediment loads with a measure of uncertainty on a yearly basis. We aggregate the posterior estimates from our methodology by financial year. The bottom plot in Figure 4 shows the annual load (black closed circle) and 95% (black segments) that were constructed for each aggregated to financial year. The D-SedNet model estimate is overlaid with a red cross. The load estimated for the FY 1990/1991 shows a wide credible interval surrounding the estimate of 11.58 megatonnes (Mt), which appears to be accompanied by large annual flow creating this large event. The estimated sediment load from D-SedNet, although much lower at 6.37

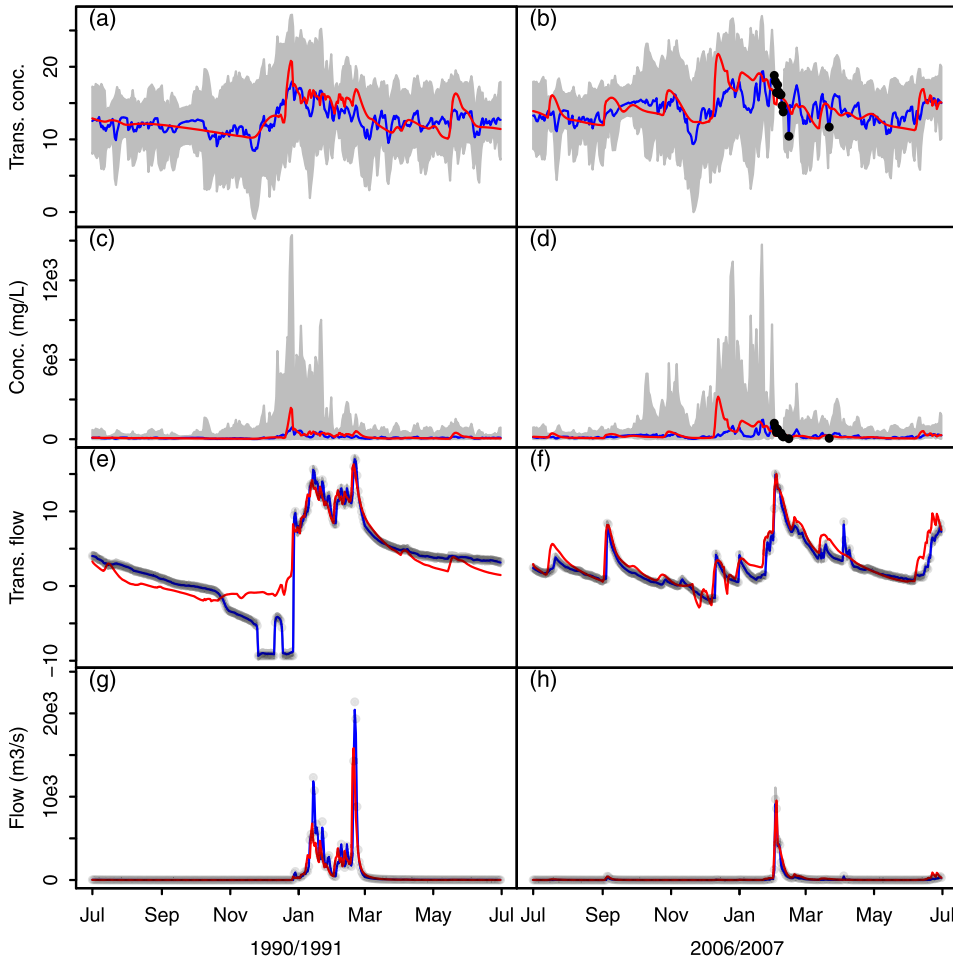


FIG. 3. Posterior distribution results for the Sellheim site for FYs 1990/1991 and 2006/2007. The posterior mean is in blue, 95% credible interval in grey, D-SedNet output in red and observations as black points. Plots (a) and (b) show the Box–Cox transformed concentration. Plots (c) and (d) show the natural scale concentration (in mg/L). Plots (e) and (f) show the Box–Cox transformed flow. Plots (g) and (h) show the natural scale flow (in m<sup>3</sup>/s). Plots (a), (c), (e) and (g) show the posterior results for 1990/1991. Plots (b), (d), (f) and (h) show the posterior results for 2006/2007.

Mt, resides within the 95% credible interval. For the majority of the years, the estimated load from D-SedNet is within the estimated posterior distribution. However, a few years exhibit loads estimates from D-SedNet output that are noticeably outside the interval. Furthermore, the estimated load using our methodology tends to be higher than that of the D-SedNet output, which agrees with Wilkinson et al. (2014).

In addition to the temporal patterns, estimates (with uncertainties) can be obtained across all 411 sites as well as seen in Figure 5. The left column of Figure 5

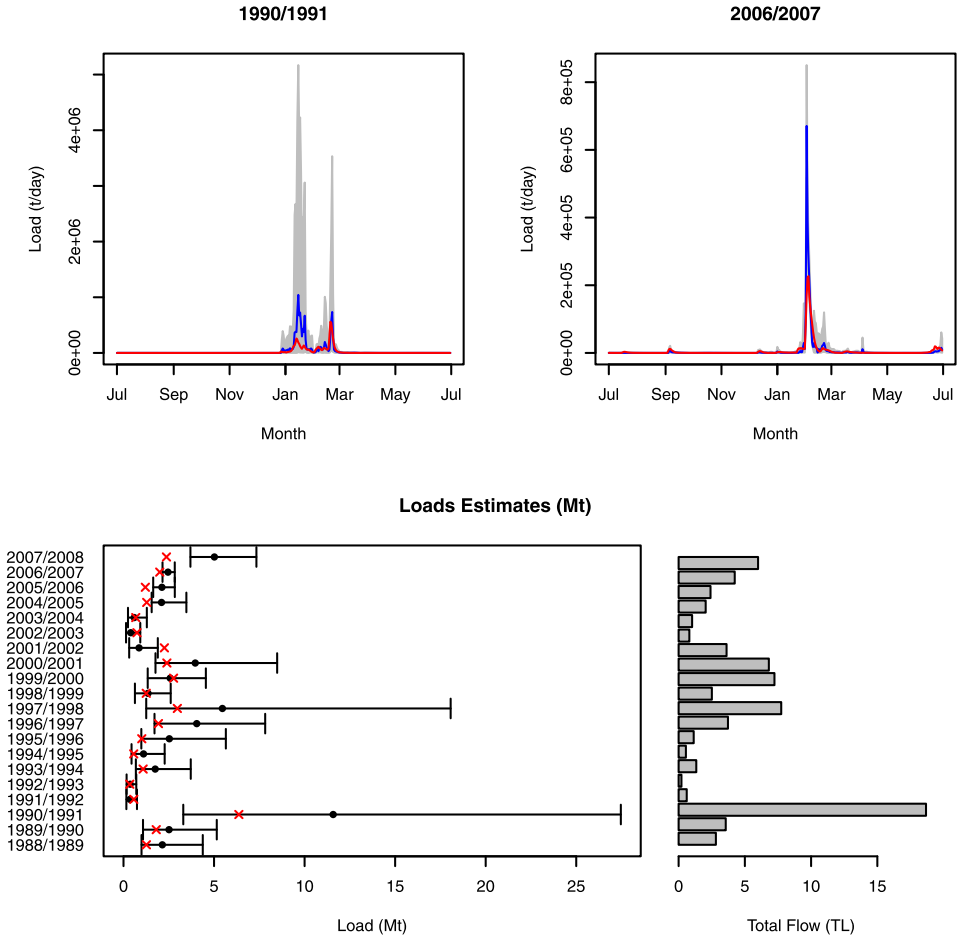


FIG. 4. Estimates of sediment load per day (tonnes/day) for FY 1990/1991 (top left) and FY 2006/2007 (top right) for the Sellheim site. The posterior results for load are in blue with 95% credible intervals in grey. The estimate of load from the D-SedNet output is in red. The bottom plot shows annual sediment load (bottom left) estimates from the posterior distribution per financial year with a comparison of estimated total flow (bottom right). The black point indicates posterior mean with lines indicating 95% credible interval from the posterior estimate. The red X are the sediment load estimates from the D-SedNet model.

shows the estimated loads on 26 December, 1990 (the day Tropical Cyclone Joy made landfall near Townsville, Queensland), for the entire Upper Burdekin with 95% credible intervals. As expected, we see higher estimates of load forming along the Burdekin River, particularly as water accumulates towards the Sellheim site. The right-hand column of Figure 5 shows the aggregated sediment load to the 1990/1991 financial year over the entire Upper Burdekin. While the results for the 1990/1991 financial year look similar to that of the estimated loads for 26 Decem-

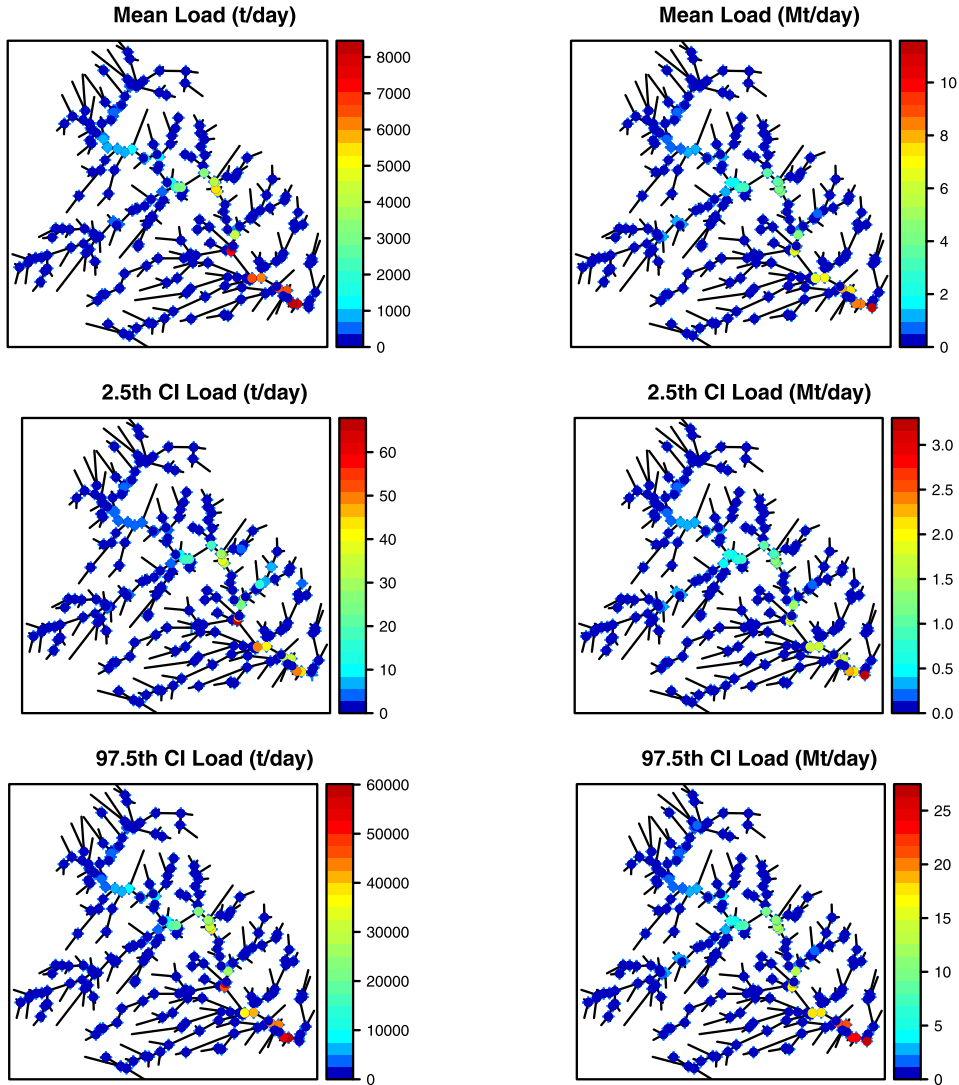


FIG. 5. Posterior estimates of sediment load (tonnes/day) for 26 December, 1990, per site in the left column, and total sediment load for the 1990/1991 financial year in the right column (mega-tonnes/year). The top row shows the posterior mean. The second row shows the lower bound, and the bottom row the upper bound of the 95% credible interval.

ber, 1990, an interesting feature of the yearly aggregate shows some estimates of higher loads in the western portion of the Upper Burdekin.

Due to the sparse nature of the monitoring data, output generated from the D-SedNet model is critical to interpolate sediment concentration and flow, which provide an estimate of sediment loads. However, at sites and days where no mon-

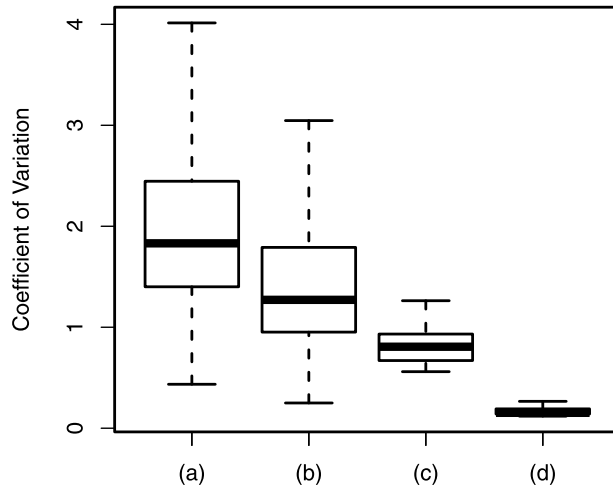


FIG. 6. The coefficient of variation of the posterior estimates of sediment loads (tonnes/day) for sites and days with (a) unmonitored concentration and flow, (b) monitored flow only, (c) monitored concentration only, and (d) monitored flow and concentration.

itoring data exists, the posterior distribution shows greater uncertainty for either flow or concentration and the estimated sediment loads. Land managers may be interested in determining how uncertainty in sediment loads can be reduced through additional monitoring data. One simple measure for comparing monitored sites against unmonitored sites is the coefficient of variation, which is a ratio of the standard deviation and the mean. We present the coefficient of variation resulting from the posterior standard deviation and mean of the daily estimated sediment loads. Figure 6 shows 4 boxplots of the coefficient of variation for days and sites of (a) unmonitored concentration and flow, (b) monitored flow with unmonitored concentration, (c) monitored concentration with unmonitored flow, and (d) monitored flow and concentration. While we cannot say the four are significantly different from each other, we can see a clear decreasing trend in the coefficient of variation when monitoring occurs versus when monitoring does not occur. Of particular note is the considerable decrease in the coefficient of variation when concentration is monitored, and when both concentration and flow are monitored.

4.3. *Model verification.* To determine the goodness of fit of the model, we evaluate the model when leaving out all monitored observations and D-SedNet output at Sellheim. We then determine the posterior distribution at Sellheim given data used for the model; that is, we find  $[\mathbf{Y}_{\text{sel},k} | \mathbf{Z}_{ik}^{\text{obs}}, \mathbf{Z}_{ik}^{\text{src}}]$ , for  $k = 1, \dots, K$  and  $i$  indexing all spatial locations included in the model with sel representing the index at Sellheim. We use the same model setup as previously defined, including the same seasonal and spatial EOFs  $\Psi$  and  $\Phi$  for concentration and flow, respectively. Figure 7 shows the posterior mean and 95% credible intervals for

TABLE 3

*The mean width of the 95% credible intervals for the posterior distribution when the model is run with and without monitoring observations and D-SedNet output at Sellheim during the 1990/1991 and 2006/2007 Financial Years*

	1990/1991		2006/2007	
	With data	Without data	With data	Without data
Concentration	10.637	14.281	10.069	16.468
Flow	0.982	7.504	0.981	7.228

the posterior distribution of Sellheim for FYs 1990/1991 and 2006/2007, as well as overlaying the available monitoring data and D-SedNet output not used. The plots of Figure 7 show wider credible intervals for concentration and flow when compared to Figure 3. This is expected, particularly given that the variance associated with the monitoring data is considerably smaller than the variance associated with D-SedNet. To illustrate these differences more clearly, we have computed the mean width of the 95% credible intervals for concentration and flow during FYs 1990/1991 and 2006/2007 which are found in Table 3. Overall, 93.66% of the D-SedNet concentration output lies within the 95% credible interval, 93.01% of D-SedNet flow, 86.67% of monitored concentration and 83.72% of monitored flow. While not capturing 95% of the process model output or monitoring, we note that we do not consider the additional variation of the data model.

From these results, we can conclude that our model does fit reasonably well. The D-SedNet output predominantly lies within the 95% credible interval in both FYs, while flow monitoring data does not appear successful for the 1990/1991 financial year and very successful for the 2006/2007 financial year. We do note that 1990/1991 was an abnormal year with regards to flow. Further, we note that the model does reasonably well in an average year, but becomes complicated under abnormal conditions. This does seem reasonable with regard to extreme events, but does highlight the importance of monitoring, especially in the presence of abnormal events. We make special note that the posterior results of  $\mathbf{Y}_{ik}$  are sensitive to choices of the empirically specified variances, which highlights the need estimate these variances from any available sources of data.

**5. Discussion.** We presented a spatio-temporal model for quantifying sediment loads in the Upper Burdekin catchment. Our approach allowed for modelling of sediment concentration and daily discharge volume that can be used to quantify sediment loads with a measure of uncertainty while accounting for spatio-temporal dependencies. The approach we developed is critical to achieving a more accurate estimate of loads that can assist in the prioritisation of areas within the catchment that (i) exhibit high sediment loads and could be targeted for improved management, (ii) can be used to help detect and evaluate improvements to water quality

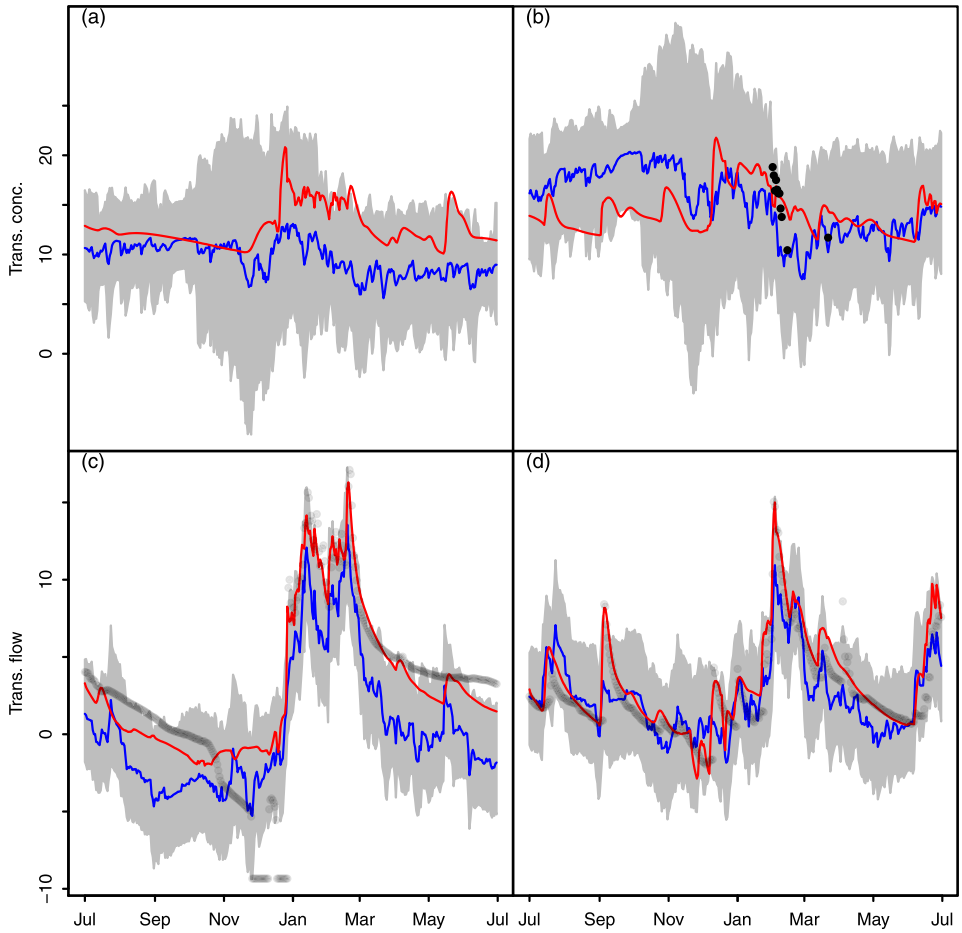


FIG. 7. Posterior distribution results for the Sellheim site for FYs 1990/1991 and 2006/2007 without use of monitoring data or D-SedNet output from Sellheim for all 20 years in the model. The posterior mean is in blue and the 95% credible interval in grey. For reference, the D-SedNet output is overlaid in red and monitoring data as black points. Plots (a) and (b) show the Box–Cox transformed concentration for FYs 1990/1991 and 2006/2007, respectively. Plots (c) and (d) show the Box–Cox transformed flow for FYs 1990/1991 and 2006/2007, respectively.

following improved land management, and (iii) identify where monitoring data are lacking and investment in gauges or water quality samplers could be beneficial. Critically, our methodology allows for blending of process model output with monitored data, allowing for estimates in the presence of sparse monitoring.

Due to the “big data” nature of the problem with high dimensionality in both space and time, it is difficult to capture the underlying dynamical structure. Modelling in a reduced rank setting is critical to capturing the spatial and temporal dependencies. The methodology utilised a two-tiered reduced rank approach by

expanding and modifying the methods developed by Oleson and Wikle (2013). We modelled the underlying dynamics of the spatio-temporal process through a basis function expansion for both seasonal and spatial structure. The expansion coefficients were then modelled through an autoregressive structure, propagating one time-step ahead on a yearly scale. In doing so, we improved computational efficiency by avoiding modelling the dynamics of the process in the large full-rank dimensional space. Additionally, this method of dimension reduction allows for multiscale temporal load quantification, which is of interest to land managers.

The use of multiple data sources was a critical strength of the methodology. Though we included only two sources (monitoring data and D-SedNet model output), we could easily have allowed additional data models. For example, remotely sensed spectral properties of the water in these rivers could also be included as a separate data model conditioned on the underlying latent process  $\mathbf{Y}_{ik}$ . Through the two data models employed in this work, observations and computer model output are both used to make inferences about the underlying true sediment concentration and flow whilst accounting for their associated uncertainties. At a given location and time, this also allowed for one data source to “in-fill” where the other was missing and inform us about the state of the underlying latent process.

Estimates produced for the Upper Burdekin catchment agreed on many occasions with the estimated annual sediment load from the D-SedNet model; that is, D-SedNet modelled loads resided in the credible interval of the estimated annual load from our model, as demonstrated by the Sellheim site. Disagreement between outputs from the D-SedNet model and the posterior estimate of annual sediment load is noted for later years in the period modelled after 2004. This agreed with Wilkinson et al. (2014), where D-SedNet modelled output underestimated a load after a drought year. Notably, our methodology showed less uncertainty during the period after 2004. This coincided with the period during which monitoring occurred more frequently. We make special note that the goal of our model was not to correct D-SedNet, but rather use D-SedNet as a way of infilling unmonitored days and sites. If the goal was to correct the process-based D-SedNet model, then one could include a bias term in the D-SedNet Data Model similar to Salazar et al. (2011). However, this is challenging in our application due to sparsity of the data and complexity of the latent process.

Our methodology also showed the importance of monitoring flow and concentration. The amount of uncertainty tended to be lower at sites where monitoring occurred for both concentration and flow, as noted by the lower coefficient of variation at these sites during days that had recorded flow and concentration. While it is not economically feasible to monitor the entire Upper Burdekin, our results suggested that more monitoring would be beneficial at reducing uncertainty, particularly with monitoring sediment concentration or a suitably correlated surrogate. One possible surrogate for concentration could be turbidity, which may be more



cost effective to measure, and that was considered in the single site analysis of Pagendam et al. (2014). Alternative sampling methods that are economically feasible could also be considered. Additionally, our methodology could guide future research in determining how frequently monitoring should occur to achieve an acceptable reduction of uncertainty at various temporal aggregation levels including monthly, seasonally and yearly.

While our model does a good job of estimating uncertainty of sediment loads while accounting for spatial and temporal dependencies, we do see scope for improvement in future work. For example, we have not presented forecasting results, as we have found the seasonal variation in flows to be too variable in the Upper Burdekin. In the current modelling framework our approach had a tendency to oversmooth forecasted estimates of the flow, concentration and the subsequent load.

As noted in our results, our methodology is able to quantify sediment load with uncertainty at different time scales and locations. The result provides valuable information, previously unavailable to catchment managers for assessing sediment load generation within the catchment and making decisions at the desired temporal scale (e.g., daily, monthly, yearly). Furthermore, the approach presented here provides a measure of uncertainty at various time scales. As such, our methodology provides a flexible and powerful framework for estimating sediment loads with a measure of uncertainty in large catchment models.

**6. Acknowledgements.** We would like to thank the Queensland government for providing the monitoring data and D-SedNet model run that made this work possible. We also would like to thank the Reef Rescue and CSIRO Land and Water and Digital Productivity Flagships for their support and research funding which make development of this framework possible. We also thank two anonymous referees and the Associated Editor, whose comments have greatly improved this manuscript. Finally, we thank Dr. Geoff Hosack for his helpful and insightful comments to this manuscript.

#### SUPPLEMENTARY MATERIAL

**Supplement to “Spatio-temporal assimilation of modelled catchment loads with monitoring data in the Great Barrier Reef”** (DOI: [10.1214/16-AOAS950SUPP](https://doi.org/10.1214/16-AOAS950SUPP); .pdf). The supplementary material contains additional information related to the development and results of our model. We present figures for the available monitoring data described in Section 2. Additional posterior results not discussed in Section 4 are shown, including results for the dynamic parameters and covariate coefficients. Lastly, the details of the Markov Chain Monte Carlo algorithm are included.

#### REFERENCES

- AKSOY, H. and KAVVAS, M. L. (2005). A review of hillslope and watershed scale erosion and sediment transport models. *Catena* **64** 247–271.

- ARMOUR, J. D., HATELEY, L. R. and PITT, G. L. (2009). Catchment modelling of sediment, nitrogen and phosphorus nutrient loads with SedNet/ANNEX in the Tully–Murray basin. *Marine and Freshwater Research* **60** 1091–1096.
- BAINBRIDGE, Z. T., LEWIS, S. E., SMITHERS, S. G., KUHNERT, P. M., HENDERSON, B. L. and BRODIE, J. E. (2014). Fine-suspended sediment and water budgets for a large, seasonally dry tropical catchment: Burdekin river catchment, Queensland, Australia. *Water Resources Research* **50** 9067–9087.
- BARTLEY, R., WILKINSON, S. N., HAWDON, A., ABBOTT, B. N. and POST, D. A. (2010). Impacts of improved grazing land management on sediment yields. Part 2: Catchment response. *Journal of Hydrology* **389** 249–259.
- BARTLEY, R., BAINBRIDGE, Z. T., LEWIS, S. E., KROON, F. J., WILKINSON, S. N., BRODIE, J. E. and SILBURN, D. M. (2014). Relating sediment impacts on coral reefs to watershed sources, processes and management: A review. *Sci. Total Environ.* **468–469** 1138–1153.
- BERLINER, L. M. (1996). Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods (Santa Fe, NM, 1995)* (K. M. Hanson and R. N. Silver, eds.). *Fund. Theories Phys.* **79** 15–22. Kluwer Academic, Dordrecht. [MR1446713](#)
- BEVEN, K. and FREER, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249** 11–29.
- BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. *J. R. Stat. Soc., Ser. B Stat. Methodol.* **26** 211–252.
- BRODIE, J., SCHROEDER, T., ROHDE, K., FAITHFUL, J., MASTERS, B., DEKKER, A., BRANDO, V. and MAUGHAN, M. (2010). Dispersal of suspended sediments and nutrients in the Great Barrier Reef lagoon during river-discharge events: Conclusions from satellite remote sensing and concurrent flood-plume sampling. *Marine and Freshwater Research* **61** 651–664.
- BRODIE, J. E., DEVLIN, M., HAYNES, D. and WATERHOUSE, J. (2011). Assessment of the eutrophication status of the Great Barrier Reef lagoon (Australia). *Biogeochemistry* **106** 281–302.
- BRODIE, J. E., KROON, F. J., SCHAFFELKE, B., WOLANSKI, E. C., LEWIS, S. E., DEVLIN, M. J., BOHNET, I. C., BAINBRIDGE, Z. T., WATERHOUSE, J. and DAVIS, A. M. (2012). Terrestrial pollutant runoff to the Great Barrier Reef: An update of issues, priorities and management responses. *Mar. Pollut. Bull.* **65** 81–100.
- BRYNJARSÓTTIR, J. and BERLINER, L. M. (2014). Dimension-reduced modeling of spatio-temporal processes. *J. Amer. Statist. Assoc.* **109** 1647–1659. [MR3293617](#)
- CHEN, D., DAHLGREN, R. A., SHEN, Y. and LU, J. (2012). A Bayesian approach for calculating variable total maximum daily loads and uncertainty assessment. *Science of the Total Environment* **430** 59–67.
- CHIEW, F. H. S., PEEL, M. C. and WESTERN, A. W. (2002). Application and testing of the simple rainfall-runoff model SIMHYD. In *Mathematical Models of Small Watershed Hydrology and Applications* (V. P. Singh and D. K. Frevert, eds.) 335–367. Water Resources Publications, Littleton, CO.
- CLARK, J. S. (2005). Why environmental scientists are becoming bayesians. *Ecology Letters* **8** 2–14.
- COHN, T. A. (1995). Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers. *Reviews of Geophysics* **33** 1117–1123.
- COHN, T. A., CAULDER, D. L., GILROY, E. J., ZYNYUK, L. D. and SUMMERS, R. M. (1992). The validity of a simple statistical model for estimating fluvial constituent loads: An empirical study involving nutrient loads entering Chesapeake Bay. *Water Resources Research* **28** 2353–2363.
- COOPER, D. M. and WATTS, C. D. (2002). A comparison of river load estimation techniques: Application to dissolved organic carbon. *Environmetrics* **13** 733–750.
- CRAWFORD, C. G. (1991). Estimation of suspended-sediment rating curves and mean suspended-sediment loads. *Journal of Hydrology* **129** 331–348.

- CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ. MR2848400
- DE'ATH, G., FABRICIUS, K. E., SWEATMAN, H. and PUOTINEN, M. (2012). The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc. Natl. Acad. Sci. USA* **109** 17995–17999.
- DOHERTY, J. and HUNT, R. J. (2009). Two statistics for evaluating parameter identifiability and error reduction. *Journal of Hydrology* **366** 119–127.
- DOHERTY, J. and JOHNSTON, J. M. (2003). Methodologies for calibration and predictive analysis of a watershed model. *Journal of the American Water Resources Association* **39** 251–265.
- FABRICIUS, K. E., LOGAN, M., WEEKS, S. and BRODIE, J. (2014). The effects of river run-off on water clarity across the central Great Barrier Reef. *Mar. Pollut. Bull.* **84** 191–200.
- FURNAS, M. J. (2003). Catchments and corals: Terrestrial runoff to the Great Barrier Reef, report, Australian Institute of Marine Science and CRC Reef, Townsville, Australia.
- GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6** 721–741.
- GLADISH, D. W. and WIKLE, C. K. (2014). Physically motivated scale interaction parameterization in reduced rank quadratic nonlinear dynamic spatio-temporal models. *Environmetrics* **25** 230–244. MR3256458
- GLADISH, D. W., KUHNERT, P. M., PAGENDAM, D. E., WIKLE, C. K., BARTLEY, R. B., SEARLE, R. D., ELLIS, R. J., DOUGALL, C., TURNER, R. D. R., LEWIS, S. E., BAINBRIDGE, Z. T. and BRODIE, J. E. (2016). Supplement to “Spatio-temporal assimilation of modelled catchment loads with monitoring data in the Great Barrier Reef.” DOI:10.1214/16-AOAS950SUPP.
- GRAF, A., BOGENA, H. R., DRÜE, C., HARDELAUF, H., PÜTZ, T., HEINEMANN, G. and VERECKEN, H. (2014). Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment. *Water Resources Research* **50** 4837–4857.
- KUHNERT, P. M., HENDERSON, B. L., LEWIS, S. E., BAINBRIDGE, Z. T., WILKINSON, S. N. and BRODIE, J. E. (2012). Quantifying total suspended sediment export from the Burdekin River catchment using the loads regression estimator tool. *Water Resources Research* **48** W04533.
- LETCHER, R. A., JAKEMAN, A. J., CALFAS, M., LINFORTH, S., BAGINSKA, B. and LAWRENCE, I. (2002). A comparison of catchment water quality models and direct estimation techniques. *Environmental Modelling and Software* **17** 77–85.
- LE COZ, J., RENARD, B., BONNIFAIT, L., BRANGER, F. and LE BOURSICAUD, R. (2014). Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach. *Journal of Hydrology* **509** 573–587.
- LITTLEWOOD, I. G. and MARSH, T. J. (2005). Annual freshwater river mass loads from Great Britain, 1975–1994: Estimation algorithm, database and monitoring network issues. *Journal of Hydrology* **304** 221–237.
- LIU, Y., YANG, P., HU, C. and GUO, H. (2008). Water quality modeling for load reduction under uncertainty: A Bayesian approach. *Water Research* **42** 3305–3314.
- LOUGH, J. M., LEWIS, S. E. and CANTIN, N. E. (2015). Freshwater impacts in the central Great Barrier Reef: 1648–2011. *Coral Reefs* **34** 1–13.
- MANTOVAN, P. and TODINI, E. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology* **330** 368–381.
- MCCULLOCH, M., FALLON, S., WYNDHAM, T., HENDY, E., LOUGH, J. and BARNES, D. (2003). Coral record of increased sediment flux to the inner Great Barrier Reef since European settlement. *Nature* **421** 727–730.
- MOYED, R. A. and CLARKE, R. T. (2005). The use of Bayesian methods for fitting rating curves, with case studies. *Advances in Water Resources* **28** 807–818.

- OLESON, J. J. and WIKLE, C. K. (2013). Predicting infectious disease outbreak risk via migratory waterfowl vectors. *J. Appl. Stat.* **40** 656–673. [MR3047308](#)
- PAGENDAM, D. E., KUHNERT, P. M., LEEDS, W. B., WIKLE, C. K., BARTLEY, R. and PETERSON, E. E. (2014). Assimilating catchment processes with monitoring data to estimate sediment loads to the Great Barrier Reef. *Environmetrics* **25** 214–229. [MR3256457](#)
- PERRIN, C., MICHEL, C. and ANDRÉASSIAN, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* **279** 275–289.
- R CORE TEAM (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- REEF WATER QUALITY PROTECTION PLAN SECRETARIAT (2013). Reef Water Quality Protection Plan. Available at <http://www.reefplan.qld.gov.au/resources/assets/reef-plan-2013.pdf>.
- REITAN, T. and PETERSEN-ØVERLEIR, A. (2011). Dynamic rating curve assessment in unstable rivers using Ornstein–Uhlenbeck processes. *Water Resources Research* **47** W02524.
- RENARD, K. G., FOSTER, G. R., WEESIES, G. A., MCCOOL, D. K. and YODER, D. C. (1997). Predicting soil erosion by water: A guide to conservation planning with the revised universal soil loss equation (RUSLE). U.S. Department of Agriculture, Agriculture Handbook No. 703, 404 pp.
- ROSEWELL, C. J. (1993). SOILLOSS—A program to assist in the selection of management practices to reduce erosion. Technical Report No. Technical Handbook No.11, 2nd edition, Soil Conservation Service, NSW.
- RUSTOMJI, P. and WILKINSON, S. N. (2008). Applying bootstrap resampling to quantify uncertainty in fluvial suspended sediment loads estimated using rating curves. *Water Resources Research* **44** W09435.
- SALAZAR, E., SANSÓ, B., FINLEY, A. O., HAMMERLING, D., STEINSLAND, I., WANG, X. and DELAMATER, P. (2011). Comparing and blending regional climate model predictions for the American Southwest. *J. Agric. Biol. Environ. Stat.* **16** 586–605. [MR2862300](#)
- SANSÓ, B. and GUENNI, L. (1999). Venezuelan rainfall data analysed by using a Bayesian space–time model. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **48** 345–362.
- SCHMELTER, M. L., HOOTEN, M. B. and STEVENS, D. K. (2011). Bayesian sediment transport model for unisize bed load. *Water Resources Research* **47** W11514.
- TOMKINS, K. M. (2014). Uncertainty in streamflow rating curves: Methods, controls and consequences. *Hydrological Processes* **28** 464–481.
- WALLING, D. E. and WEBB, B. W. (1985). Estimating the discharge of contaminants to coastal waters by rivers: Some cautionary comments. *Mar. Pollut. Bull.* **16** 488–492.
- WANG, Y. G., KUHNERT, P. and HENDERSON, B. (2011). Load estimation with uncertainties from opportunistic sampling data—A semiparametric approach. *Journal of Hydrology* **396** 148–157.
- WIKLE, C. K. and BERLINER, L. M. (2007). A Bayesian tutorial for data assimilation. *Phys. D* **230** 1–16. [MR2345198](#)
- WIKLE, C. K. and HOOTEN, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST* **19** 417–451. [MR2745992](#)
- WILKINSON, S. N., HANCOCK, G. J., BARTLEY, R., HAWDON, A. A. and KEEN, R. J. (2013). Using sediment tracing to assess processes and spatial patterns of erosion in grazed rangelands, burdekin river basin, Australia. *Agriculture, Ecosystems and Environment* **180** 90–102.
- WILKINSON, S. N., DOUGALL, C., KINSEY-HENDERSON, A. E., SEARLE, R. D., ELLIS, R. J. and BARTLEY, R. (2014). Development of a time-stepping sediment budget model for assessing land use impacts in large river basins. *Sci. Total Environ.* **468–469** 1210–1224.
- WU, W., CLARK, J. S. and VOSE, J. M. (2010). Assimilating multi-source uncertainties of a parsimonious conceptual hydrological model using hierarchical Bayesian modeling. *Journal of Hydrology* **394** 436–446.

WU, G., HOLAN, S. H. and WIKLE, C. K. (2013). Hierarchical Bayesian spatio-temporal Conway-Maxwell Poisson models with dynamic dispersion. *J. Agric. Biol. Environ. Stat.* **18** 335–356. MR3110897

D. W. GLADISH  
D. E. PAGENDAM  
R. BARTLEY  
R. D. SEARLE  
ECOSCIENCES PRECINCT  
41 BOGGO ROAD  
DUTTON PARK, QLD 4102  
AUSTRALIA  
E-MAIL: [dan.gladish@csiro.au](mailto:dan.gladish@csiro.au)  
[dan.pagendam@csiro.au](mailto:dan.pagendam@csiro.au)  
[rebecca.bartley@csiro.au](mailto:rebecca.bartley@csiro.au)  
[ross.searle@csiro.au](mailto:ross.searle@csiro.au)  
URL: <http://www.csiro.au>

C. K. WIKLE  
DEPARTMENT OF STATISTICS  
UNIVERSITY OF MISSOURI  
146 MIDDLEBUSH HALL  
COLUMBIA, MISSOURI 65211  
USA  
E-MAIL: [wiklec@missouri.edu](mailto:wiklec@missouri.edu)  
URL: <http://www.stat.missouri.edu/~wikle/>

R. D. R. TURNER  
ECOSCIENCES PRECINCT  
41 BOGGO ROAD  
DUTTON PARK, QLD 4102  
AUSTRALIA  
E-MAIL: [ryan.turner@dsiti.qld.gov.au](mailto:ryan.turner@dsiti.qld.gov.au)

P. M. KUHNERT  
GPO BOX 664  
CANBERRA, ACT 2601  
AUSTRALIA  
E-MAIL: [petra.kuhnert@csiro.au](mailto:petra.kuhnert@csiro.au)  
URL: <http://www.csiro.au>

R. J. ELLIS  
C. DOUGALL  
61 MARY STREET  
BRISBANE, QLD 4000  
AUSTRALIA  
E-MAIL: [robin.ellis@dsiti.qld.gov.au](mailto:robin.ellis@dsiti.qld.gov.au)  
[cameron.dougall@dnrm.qld.gov.au](mailto:cameron.dougall@dnrm.qld.gov.au)

S. E. LEWIS  
Z. T. BAINBRIDGE  
J. E. BRODIE  
CENTRE FOR TROPICAL WATER  
AND AQUATIC ECOSYSTEM RESEARCH  
ATSIP BUILDING 145  
JAMES COOK UNIVERSITY  
TOWNSVILLE, QLD 4811  
AUSTRALIA  
E-MAIL: [stephen.lewis@jcu.edu.au](mailto:stephen.lewis@jcu.edu.au)  
[zoe.bainbridge@jcu.edu.au](mailto:zoe.bainbridge@jcu.edu.au)  
[jon.brodie@jcu.edu.au](mailto:jon.brodie@jcu.edu.au)