Accepted Manuscript

Population density controls on microbial pollution across the Ganga catchment

D.G. Milledge, S.K. Gurjar, J.T. Bunce, V. Tare, R. Sinha, P.E. Carbonneau

PII:	S0043-1354(17)30870-9
DOI:	10.1016/j.watres.2017.10.033
Reference:	WR 13290
To appear in:	*Water Research*
Received Date:	17 May 2017
Revised Date:	12 October 2017
Accepted Date:	17 October 2017

Please cite this article as: D.G. Milledge, S.K. Gurjar, J.T. Bunce, V. Tare, R. Sinha, P.E. Carbonneau, Population density controls on microbial pollution across the Ganga catchment, *Water Research* (2017), doi: 10.1016/j.watres.2017.10.033

Population density controls on microbial pollution across the Ganga catchment

D.G. Milledge[1*], S.K. Gurjar[2], J.T. Bunce[3], V. Tare[2], R. Sinha[2,4], P.E. Carbonneau[1]

1. Institute of Hazard, Risk, and Resilience and Department of Geography, Durham University, Durham, UK.

2. Department of Civil Engineering, Indian Institute of Technology Kanpur, India

3. Department of Civil Engineering and Geoscience, Newcastle University, Newcastle, UK.

4. Department of Earth Sciences, Indian Institute of Technology Kanpur, India

* Corresponding author: David Milledge, Department of Geography, Durham University, Science Laboratories, Durham, DH1 3LE, UK. d.g.milledge@durham.ac.uk

Abstract

For millions of people worldwide, sewage-polluted surface waters threaten water security, food security and human health. Yet the extent of the problem and its causes are poorly understood. Given rapid widespread global urbanisation, the impact of urban versus rural populations is particularly important but unknown. Exploiting previously unpublished archival data for the Ganga (Ganges) catchment, we find a strong non-linear relationship between upstream population density and microbial pollution, and predict that these river systems would fail faecal coliform standards for irrigation waters available to 79% of the catchment's 500 million inhabitants. Overall, this work shows that microbial pollution is conditioned by the continental-scale network structure of rivers, compounded by the location of cities whose growing populations contribute c. 100 times more microbial pollutants per capita than their rural counterparts.

Highlights

Faecal coliform concentration is strongly related to upstream population density [80]

Local rivers predicted to fail WHO irrigation standards for 79% of the population [83]

Rivers receive c. 100 times more sewage per capita from urban than rural populations [84]

Microbial pollution is conditioned by river network structure and settlement pattern [84]

Himalayan headwaters continue to dilute microbial pollution far downstream [74]

Keywords: faecal coliform; river network; population density; catchment-scale;

Background

Rising demands on water resources raise concerns about the sustainable provision of clean water worldwide. Unclean water poses significant risks of diarrhoea, opportunistic infections, and consequent malnutrition accounting for ~1.7 million deaths annually; of which >90% are

in developing countries and almost half are children [Prüss-Ustün et al., 2014]. These deaths are primarily due to ingestion of faecal pathogens from humans or animals [Ashbolt et al., 2004; Kotloff et al., 2013; Prüss-Ustün et al., 2014].

India's growing population and economy are driving rapid urbanisation (30% of the population now live in urban areas [Census of India, 2011a]) and exerting increased pressure on surface and groundwater availability. In rural areas ~67% of the population defecate in the open [Census of India, 2011b], a practice that poses severe risk to health and safety [Clasen et al., 2010; Mara et al., 2010; Ziegelbauer et al., 2012; Kotloff et al., 2013]. While in urban areas ~80% of the population have access to a toilet [Census of India, 2011b], but only ~30% are connected to a sewage pipeline and few pipelines are connected to a treatment plant [Narain, 2012]. The impact of these sanitation problems on surface water quality has been documented for many years at individual sample locations or river reaches across India [Bhargava, 1983; Mukherjee et al., 1993; Baghel et al., 2005; Mishra et al., 2009; Central Pollution Control Board, 2010]. However, there has been no catchment-wide quantification of the problem and limited indication of what is driving it. The former is essential to fully understand the scale of intervention required, while the latter might inform decision-making on 'what to do where'. Urban areas often dominate the microbial pollution signal in rivers [Tchobanoglous et al., 1991; Kay et al., 2008; McGrane et al., 2014] but there is little consensus on the extent to which this reflects an increased impact per capita or simply a larger population and thus source. This difference is important since a higher per capita impact indicates reduced attenuation, perhaps due to more efficient delivery to the river system or less efficient treatment. If the difference can be attributed to per capita contribution this will define the extent to which urban or rural focused interventions will improve surface water quality.

We address this question using archival water quality data from across the Ganga (Ganges) catchment and show the pattern of microbial pollution in its major rivers. We compare instream concentrations of a pollution proxy with upstream densities of the two major sources of faecal pathogens (humans and livestock) at 100 sites spanning an approximate surface area of 10^6 km^2.

Faecal pathogens are difficult to measure, however thermo-tolerant coliforms, which originate in faeces (i.e. faecal coliforms, FC), are easily detectable and routinely monitored as indicator organisms [Ashbolt et al., 2001]. FCs are not a perfect predictor of human pathogen presence, rather they establish connectivity between defecation and some receiving environment which could be contributed to by a pathogen carrier. New host-specific tracing techniques allow more precise tracking of microbial pollution sources that can help to better assess risks to human health [Harwood et al., 2014, Field and Samadpour, 2007]. However, such techniques are not used within routine monitoring in India and thus do not have the spatial coverage required for our analysis. Furthermore, the use of FCs for monitoring pollution is still regarded as a viable measure of drinking and irrigation water quality [WHO, 2017].

Two key issues that must be addressed are: 1) the extent to which the FC signal that we observe reflects human sources; and 2) the potential impact of FC die-off in our pollution tracer. Upstream livestock and human population densities are strongly correlated at the catchment scale limiting our capacity to identify the source of the pollution signal. To address this, we seek to de-correlate the predictor variables by using a mixing model to estimate contributions from each non-overlapping segment of the catchment (our sub-catchments). To address the impact of die-off in our pollution tracer we adjust the population and livestock densities using a distance decay function then seek decay parameters that will maximise performance of our statistical model.

In the sections that follow we first introduce our null hypothesis that pollution should be linearly related to source density (both with and without accounting for die-off). We then detail our data sources and methods for their analysis; and introduce the mixing model that we use to calculate effective FC concentrations and source densities for each sub-catchment (the non-overlapping segments of the catchment).

Theory: Expected relationship between FC concentration and upstream source density with and without die-off

The FC concentration (C_{FC}) at a given location is defined by the ratio of the FC flux (Q_{FC}) to the water flux (Q_w):

$$C_{FC} = \frac{Q_{FC}}{Q_w} \quad (1)$$

Under the assumption that there is no die-off in FCs over time, the FC flux is calculated from:

$$Q_{FC} = (P_h N_h + P_a N_a) = (P_h \rho_h + P_a \rho_a) A \quad (2)$$

where: P_h is the production rate of FCs per human head [MPN $\#^{-1}$ T^{-1}]; P_a is the production rate per head of livestock [MPN $\#^{-1}$ T^{-1}]; N_h and N_a are the total upstream population of humans and livestock respectively [#]; ρ_h and ρ_a are the upstream population densities of humans and livestock respectively [# L^{-2}]; and A is the catchment area [L^2]. Under the assumption of spatially uniform and time invariant runoff R_w [$L\ T^{-1}$] the water flux Q_w [$L^3\ T^{-1}$] is calculated from:

$$Q_w = R_w A \quad (3)$$

Substituting equations 2 and 3 into equation 1 gives the following equation for FC concentration at each measurement point as a function of upstream population density.

$$C_{FC} = \frac{(P_h \rho_h + P_a \rho_a)}{R_w} = k_h \rho_h + k_a \rho_a \quad (4)$$

where: $k_h=P_h/R_w$ and $k_a=P_a/R_w$. It is clear from this relationship that under these assumptions C_{FC} should be a linear function of upstream population and livestock density with the gradients defined by the ratio of production rate, P, to runoff, R_w.

The assumption of no FC die-off is unlikely to be true but controls on die-off remain poorly understood. Given the uncertainties, die-off is most often represented using an exponential decay based on first order kinetics [Crane and Moore, 1986; Sadeghi and Arnold, 2002; Cho et al., 2012]:

$$Q_{FC} = Q_0 e^{-k_1 t} \quad (5)$$

where: Q_0 is the FC flux at time t_0 (the time of exit from the gut) [MPN T^{-1}], t is time since exit [T], k_1 is a decay coefficient [T^{-1}]. Assuming uniform time invariant FC velocity from source to measurement point the FC flux Q_{FC} can be expressed as a function of distance:

$$Q_{FC} = Q_0 e^{-k_1\left(\frac{x}{v}\right)} \quad (6)$$

where: x is the travel distance from source to measurement point [L] and v is the characteristic velocity [L T^{-1}]. Changing population (of people or livestock) with distance x upstream of the sampling point can be calculated as the derivative of N(x):

$$n(x) = -\frac{dN}{dx} = -\rho(x)\frac{dA}{dx} - A(x)\frac{d\rho}{dx} \quad (7)$$

Assuming that FC production rates are time invariant and incorporating characteristic velocity into the decay coefficient to express decay in terms of distance, the FC flux can be calculated by combining equations 2, 6 and 7 and integrating over the range of travel distances from the measurement point to the furthest point upstream:

$$Q_{FC} = \int_0^{x_{max}} \left(\left(P_h\, n_h(x) + P_a\, n_a(x)\right) e^{-k\,x} \right) dx \quad (8)$$

where change in population (for both humans and livestock) and area are a function of travel distance; and $k=k_1/v$ the distance decay coefficient [L^{-1}]. Substituting equations 3 and 8 into equation 1 gives the following equation for FC concentration:

$$C_{FC} = \int_0^{x_{max}} \left(\frac{(P_h\, n_h(x) + P_a\, n_a(x))\, e^{-k\,x}}{R_w\, A} \right) dx \qquad (9)$$

This can be implemented in discrete form by summing over the ncells upslope of the measurement point where for each cell the flow path lengths and routes are derived from digital elevation data, and human and livestock population data from the sources described below.

$$C_{FC} = \sum_{i=1}^{ncells} \left(\frac{(P_h\, \rho_{hi} + P_a\, \rho_{ai}) A_i\, e^{-k\,x_i}}{R_w\, A_i} \right) \qquad (10)$$

where: ρ_{hi} and ρ_{ai} are the density of human and animal populations respectively in cell i; A_i is the area of cell i; and x_i is the average flowpath length from cell i to the measurement point. Rearranging and simplifying equation 10 gives:

$$C_{FC} = k_h \sum_{i=1}^{ncells} \left(\rho_{hi}\, e^{-k\,x_i} \right) + k_a \sum_{i=1}^{ncells} \left(\rho_{ai}\, e^{-k\,x_i} \right) \qquad (11)$$

where: $k_h=P_h/R_w$ and $k_a=P_a/R_w$. Re-arranged in this form, equation 11 shows that accounting for FC die-off, C_{FC} remains a linear function of population and livestock density transformed to account for flowpath length. As in the no die-off case (equation 4), the linear coefficients reflect the ratio of (human or livestock) production rate to runoff.

Methods

We used water quality samples from 100 locations across the Ganga catchment (Figure 1), collected and analysed by six agencies following a uniform protocol. Total and faecal coliform concentrations were estimated using the standardised 9221B and 9221E multiple tube fermentation techniques [APHA, 1995] to establish the most probable number (MPN) of

faecal coliforms per 100ml. At each site, we collated 10 years of data (2002-2012). The frequency with which this data was sampled varies between sites, from three samples per year at the two most remote Himalayan sites, to quarterly for 24 more Himalayan sites and one or two samples per month at the remaining sites. At ~30 sites, samples were collected at two locations across the river in some years in order to improve representation. This data was quality checked for potential data entry or measurement errors. We removed a total of 63 observations where FC concentrations exceeded Total Coliform (TC) concentrations (since FC is a subset of TC). We also removed two observations at a single site on the same date where FC concentration exceeded 10^{10} MPN / 100 ml. We consider this to be suspicious given that the concentration is ~100 times the upper end of the range of observed concentrations for sewage influent [Tchobanoglous et al., 1991]. Removing suspicious observations results in a loss of <0.5% of the full dataset and <3% at any individual site. The error-checked FC data at each site were poorly approximated by a normal distribution but were generally well approximated by a log-normal distribution, thus we used geometric means to summarise FC concentration for each site throughout our analysis.

To estimate upstream population density we used the GPWv3 gridded synthesis of census data from 2000 [Balk and Yetman, 2004; Balk et al., 2010]. To estimate livestock density we used the FAO global gridded livestock density data [Wint and Robinson, 2007; Robinson et al., 2014], weighted by estimates of FC production rates for each livestock type (cow and buffalo: 10^{11} MPN/# day; goats and sheep: 1.2 x 10^{10} MPN/# day; pigs: 1.1 x 10^{10} MPN/# day; poultry: 1.4 x 10^{8} MPN/# day) [ASAE Standards, 1998]. Upstream area, upstream population density (UPD) and upstream livestock density (ULD) for each sample point were calculated using a D8 flow routing algorithm [O'Callaghan and Mark, 1984; Schwanghart and Scherler, 2014] and the hydrologically corrected 90 m SRTM DEM [Farr et al., 2007]. To examine the influence of coliform die-off in transit and thus relax the assumption that

coliforms behave as conservative tracers we introduced an exponential decay in coliform concentration with distance from the source. We sampled the shape parameter that defines the rate of distance decay at 500 logarithmic intervals from 10^{-8} to 10^{-1} km^{-1} testing model performance in each case using ordinary least squares regression.

Mixing model

The observation locations form a nested set of catchments where 82% of observation sites have at least one observation site upstream. We deal with this nested sampling in two ways. First, by assessing the results for only non-nested (independent) catchments, however this considerably reduces the number of available observations. Second, by performing the analysis using sub-catchments, where these are defined as the part of the catchment that drains to the current sample site without first draining through any upstream sample site. The result of this definition is segmentation of the entire Ganga catchment into 100 non-overlapping sub-catchments.

Effective source density and FC concentration are then calculated for each sub-catchment using an approach similar to that of Granger et al. [1996] and Portenga et al. [2015] for effective erosion rates in nested catchments. To do this we assume that catchment area can be used as a proxy for discharge (equation 3) and use a mixing model to calculate the concentration of the FC input for the sub-catchment (C_{FCr}) given the catchment area and FC concentration at the upstream and downstream boundaries:

$$C_{FCr} = \frac{Q_{FCr}}{A_r} = \frac{C_{FCd}\,A_d - \Sigma_{i=1}^{n}(C_{FCui}\,A_{ui})}{A_d - \Sigma_{i=1}^{n}(A_{ui})} \quad (12)$$

where: C_{FCui} is the FC concentration at upstream boundary i; and C_{FCd} is the FC concentration at the downstream boundary of the sub-catchment; A_{ui} is the catchment area for upstream boundary i; A_d is the catchment area for the downstream boundary of the sub-catchment; and

n is the number of upstream boundaries. We repeat the same process to calculate the human and animal populations densities (ρ_r) within the catchment area that drains into this sub-catchment:

$$\rho_r = \frac{N_r}{A_r} = \frac{\rho_d A_d - \Sigma_{i=1}^{n}(\rho_{ui} A_{ui})}{A_d - \Sigma_{i=1}^{n}(A_{ui})} \quad (13)$$

where: ρ_{ui} is the upstream population density of upstream boundary i; and ρ_d is the upstream population density of the downstream boundary of the sub-catchment.

Results

Observed pattern of FC concentrations

Our results suggest that high FC concentrations previously reported at the reach and sub-catchment scale [Mukherjee et al., 1993; Baghel et al., 2005; Mishra et al., 2009; Central Pollution Control Board, 2010] do not reflect isolated pockets of poor water quality but extensive pollution across the catchment. Decadal mean FC concentrations at sites across the Ganga catchment range from $3x10^0$ to $2.5x10^6$ MPN/100ml. 70% of sites fail Indian Government desirable bathing limits [Central Pollution Control Board, 2008] with those that pass located almost exclusively in the sparsely populated catchment headwaters. On the more populous plains, 70 of the 80 sites fail the desirable limits and 63 of the 80 sites fail the maximum permissible 2500 MPN/100ml limit [Central Pollution Control Board, 2008]. Locally high FC concentrations are generally associated with large population centres (Figure 1), most markedly for rivers with smaller catchment areas (e.g. the Varuna at Varanasi). FC concentrations are moderately reduced downstream of the Yamuna-Ganga confluence as tributaries with lower FC concentrations dilute the main stem. Further downstream, even large cities (e.g. Patna) have limited influence and many samples on the main stem have very similar FC concentration, reflecting the central tendency of water quality with increasing catchment area for nested catchments.

239

240 **Catchment scale relationships between FC concentration and upstream source density**

241 Since people and livestock are the primary sources of FCs, we expect FC concentration to
242 increase with the upstream density of these sources. Figure 2 suggests that the data fit this
243 expectation. If FC production per capita is spatially uniform, delivery to the river is
244 independent of population density, and if water flux is a linear function of catchment area,
245 then we expect FC concentration to be a linear function of upstream source density with the
246 form $y=\beta x$ (see equations 1-4 for a full derivation). Variability in delivery to the river
247 network or transit time through the river network that is uncorrelated with population density
248 will introduce scatter to the relationship but should not alter its functional form. However,
249 comparing the data with linear contours in Figure 2 shows that the data are not a good fit to a
250 linear function ($r^2 < 0.1$). Power functions are a better fit ($r^2=0.69$ for UPD and 0.62 for ULD)
251 but over-predict high and low FC values and under-predict central values with both UPD and
252 ULD. Quadratic relationships offer a further improvement ($r^2=0.71$ for UPD and 0.68 for
253 ULD) suggesting positive curvature in log-log space but have a physically unreasonable
254 negative slope at low population densities. Residuals from the quadratic function, fitted by
255 ordinary least squares regression, for both population and livestock show some
256 heteroscedasticity, though White [1980] and BPK [Breusch and Pagan, 1979; Koenker, 1980]
257 tests return p-values that are always below 0.1. Given this moderate heteroscedasticity and
258 the insensitivity of ordinary least squares coefficients to heteroscedasticity we do not pursue
259 more complex variance weighted analyses. UPD alone explains slightly more of the variance
260 in FC concentration than ULD, but there is little difference between the explanatory power of
261 these predictors, and their combination in a multiple quadratic regression offers little
262 improvement ($R^2=0.71$). This is consistent with the strong correlation between upstream
263 population and livestock densities (Figure 2c). A cubic function constrained to monotonic

increase over the range of the data gives a similar performance to the quadratic (r^2=0.71 for UPD and 0.68 for ULD). A linear spline (in log-log space) with a single interior knot (i.e. piecewise power function) is the best-fit for both individual predictors (r^2=0.73 for UPD and 0.71 for ULD), suggesting a threshold rather than continuous change in power relationship between UPD and FC concentration. Finally, we test one further null hypothesis that there are two ranges of source density (population or livestock) with FC concentrations represented by their average value over each range. This model is important to exclude given the appearance of clustered points within Figure 2 but has difficult physical implications. It implies a step change in contribution at some source density and a constant contribution independent of source density change (i.e. a declining per-head contribution) within each range. The 'step model' (r^2=0.69) does not outperform any of the curved functions (quadratic, cubic or linear spline) for UPD, though it is a slight improvement on quadratic and cubic spline functions for ULD (r^2=0.70). These results demonstrate that there is positive curvature to the FC-UPD relationship independent of the particular functional form (quadratic, cubic or linear spline) under consideration; and that the FC-ULD relationship also contains positive curvature but can be almost as well described as two FC distributions at high and low population density.

Sub-catchment relationships between FC concentration and upstream source density

As for catchment analysis, the sub-catchment analysis suggests that people and livestock are the primary sources of FCs with FC concentration increasing with the upstream density of these sources (Figure 3). The relationship between source density and FC concentration is not linear for sub-catchment based analysis or catchment based analysis. Figure 3 shows that as in the catchment analysis the data are not a good fit to any linear model (r^2 <0.2). Power functions are a better fit (r^2=0.54 for UPD and 0.17 for ULD) but over-predict high and low FC values and under-predict central values for UPD. For ULD the fit is very poor, suggesting

that in the sub-catchment based analysis livestock density is only a weak control on FC concentration. Quadratic relationships (in loglog space) offer further improvement for UPD (r^2=0.63) but not for ULD (r^2=0.16). UPD alone explains considerably more of the variance in FC concentration than ULD. Their combination in a multiple quadratic regression offers some improvement (R^2=0.72). This reflects the reduced correlation between UPD and ULD for sub-catchment (r^2=0.66) rather than catchment (r^2=0.95) analysis (compare Figure 2c with 3c). The linear spline with a single knot (i.e. piecewise power function) or cubic function (in loglog space) constrained to monotonic increase result in similar fits relative to a quadratic for both UPD (r^2=0.63 in both cases) and ULD (r^2=0.16 and 0.15 respectively). As in the catchment analysis this suggests that there is not clear evidence for a threshold rather than continuous change in power relationship between UPD and FC concentration when examined at the sub-catchment scale. The results from these three (quadratic, cubic and linear spline) approaches demonstrate that there is positive curvature to the FC-UPD relationship independent of the particular functional form under consideration. They also demonstrate that UPD is a far better predictor than ULD for sub-catchment scale analysis and that there is some merit in considering the two in combination. This suggests that most instream FCs are human derived.

Per capita impact on instream FC concentration

Positive curvature to the FC-UPD and FC-ULD relationships indicate that FC concentration increases with upstream source density at an increasing rate per unit increase in upstream source density. This can be interpreted as the change in FC per capita with increasing upstream source density. The gradient of the line in logarithmic space reflects its exponent in linear space thus: values >1 indicate positive curvature and increasing per capita impact, those <1 indicate negative curvature and decreasing per capita impact with increased source

density. At low upstream source densities (<10 people or 6 livestock per km^2), FC concentrations are low and the gradient of all three best-fit curves is slightly less than one indicating a slight decline in per capita impact with increasing upstream source density. At source densities from 10-60 people or 6-30 livestock per km^2 the gradient of all three best-fit curves reaches then exceeds unity, indicating that per capita impact reaches a minimum and begins increasing with increasing upstream source density.

For population density, quadratic, cubic and linear spline fits all predict a very similar relationship between UPD and FC concentration for 10^2<UPD<10^3 #/km^2 (Figure 2a). Over this range the predicted FC concentration increases by three orders of magnitude (from 10^2 to 10^5 MPN/100ml), indicating a 100-fold increase in per capita impact. Over the same range in population density (10^2<UPD<10^3 #/km^2) there is considerable variability in the per capita contribution from no change at the lower limit to a 10,000-fold increase at the upper limit.

A similar comparison can be made for individual sites, with the linear trend lines in Figure 2a acting as contours for per capita impact. For example, moving downstream from the catchment with lowest population density, UPD increases 10-fold from Badrinath to Srinagar (7-77 #/km^2) but FC concentration increases only three-fold (3-10 MPN/100ml), thus per capita impact declines by a factor of 3. Continuing downstream from Srinagar to Kanpur UPD increases by a factor of 6 (77 to 450 #/km^2) while the FC concentration increases by a factor of 1600 (10 to $1.6x10^4$ MPN/100ml), thus impact per capita increases by a factor of 300. Per capita impact increases by a factor of 60,000 from its minimum for the rural Pindar catchment (B) to its maximum for the densely populated Yamuna at Delhi (A). These results indicate that urban populations contribute more sewage to the river per capita than rural populations and that this increase: 1) depends on the difference in population densities, rather than changing sharply at a particular density; 2) is large on average (a factor of 100); and 3) is highly, and asymmetrically, variable (ranging from a factor of 1 to 10,000).

Discussion

The relative importance of human or livestock FC sources

Both UPD and ULD are good predictors of FC concentration based on catchment scale analysis. This may reflect the importance of both sources, but is also very likely due to the strong positive correlation between UPD and ULD in the catchment based analysis (Figure 2c), which makes it difficult to distinguish between the sources based on these data alone. When calculated over large areas population and livestock density are highly correlated. However, at small scales population and livestock density can become de-correlated (e.g. in cities, where population density is high but livestock density low). Our sub-catchment based analysis breaks the catchment into smaller non-nested segments, disrupting the correlation between UPD and ULD (Figure 3c). This analysis shows a small reduction in the percentage of variance in FC concentration explained by UPD and a large reduction in that explained by ULD. In the sub-catchment based analysis UPD is a much better predictor of FC concentration than ULD.

This is consistent with simple accounting estimates of export coefficients calculated using population and livestock densities with estimated FC production rates for the loading terms and observed FC concentration as the output. Assuming a human production rate of 2×10^9 MPN/# day [Tchobanoglous, 1991] and livestock production rates detailed in the methods section, livestock-derived FC loads produced on any given day range from 2×10^{10} MPN/km^2 day (for ULD = 3 #/km^2) to 1.5×10^{13} MPN/km^2 day (for ULD = 200 #/km^2) while population derived FC loads range from 1.4×10^{10} MPN/ km^2 day (for UPD = 7 #/km^2) to 2×10^{12} MPN/ km^2 day (for UPD = 1000 #/km^2). Yet over this range of source densities FC concentrations increase from 2×10^0 to 1×10^5 MPN/100 ml on average. This results in export coefficients >100 times larger at high livestock and population densities than at low densities. It is

364 difficult to conceive of a mechanism for such an increase in export coefficient for livestock-
365 derived FCs as a function of source density.
366
367 **The relative importance of local or non-local FC sources**
368 UPD is a good predictor of instream FC concentrations across the Ganga catchment,
369 explaining 73% of the observed variance in decadal mean FC concentrations from a
370 catchment scale analysis and 63% from a sub-catchment scale analysis (Figure 2a and 3a).
371 This is consistent with findings from catchments across the world [Tchobanoglous et al.,
372 1991; Kay et al., 2008; McGrane et al., 2014], and with previous reach-scale findings in the
373 Ganga Catchment [Mukherjee et al., 1993; Baghel et al., 2005; Mishra et al., 2009; Central
374 Pollution Control Board, 2010]. However, there remains considerable variance in FC
375 concentration unexplained by either UPD or ULD, particularly at high population densities,
376 >100 people/km^2 (Figures 2 and 3). Previous reach-scale studies did not account for the
377 upstream boundary condition either in terms of FC flux or upstream population [Mukherjee et
378 al., 1993; Baghel et al., 2005; Mishra et al., 2009]. These studies implicitly assumed that
379 point sources proximal to sample sites dominated the FC signal (perhaps due to coliform die-
380 off in transit). However, while many of our sites near larger settlements have high coliform
381 concentrations, these concentrations are better explained by upstream population density (r^2
382 >0.7) than population of the nearest settlement (r^2=0.25). Examining paired samples above
383 and below settlements suggests that in some cases, positive residuals (where FC
384 concentration is greater than predicted) may reflect sites immediately downstream of
385 population centres. However, including a distance-decay function in our analysis did not
386 improve our ability to predict FC concentrations. Figure 4 shows that model performance is
387 initially stable as the rate at which FCs decay with distance increases, but that the
388 performance is never better than that without distance decay, and that performance declines

markedly for decay rates greater than 0.01 %/km. This reduction in performance relates to a reduction in decay-adjusted population density primarily at sites with intermediate or dense populations (Figure 5). These results suggest that, UPD is an important but not singular factor in defining the connectivity between sources and receiving waters that defines the timescales and thus efficiency of delivery. Our approach neglects many processes that should be important in the transport of coliforms from source to the point of measurement (e.g. weather dependent die-off rates, hydrological connectivity, hydraulics at the cross section and reach scale). However, it is encouraging that even our simple empirical model explains a large fraction of the variance in microbial pollution concentrations.

Implications of the FC-UPD relationship

The increase in per capita impact as UPD increases likely reflects an increase in the efficiency of delivery rather than FC production, perhaps due to changes in individual or corporate waste management decisions as population density increases. At low population densities, much of the population defecate in the open or in pit latrines [Census of India, 2011b] where faeces are less likely to be washed into the river and FCs are more likely to die *in situ*. As population density increases and towns and cities grow, the distance to open fields increases and there is a need for an alternative strategy to manage faeces. This problem has historically confronted communities across the world, leading to degradation of sanitary conditions and construction of sewers [Gandy, 2004; Allen, 2008; Benzerzour et al., 2011]. Sewage systems vary in sophistication but generally involve the movement of excreta by water out of the population centre; often made possible by piped domestic water. The faeces have a much shorter residence time in the environment and FCs will be removed primarily by sewage treatment rather than die-off in the environment. In many Indian cities, the flux of sewage that is, and must be, removed from the population centre through a growing network

414 of sewers and storm water drains is many times higher than the capacity of the sewage
415 treatment facilities [Ansari et al., 2000]. In this case the predominant impact of the sewage
416 network is to remove the sewage from the population centre and rapidly deliver it to the river
417 untreated. Sewage removal is essential for the public health of the city, but without effective
418 treatment it comes at the cost of accentuated river pollution with associated public health
419 implications for the population downstream. Here we demonstrate as others have [Central
420 Pollution Control Board, 2010] the severe river pollution that results. The extent to which this
421 can be addressed by following the same trajectory towards centralised 'end-of-pipe' sewage
422 treatment has been called into question for practical and economic reasons [Jha, 2003;
423 Bracken et al., 2007; Katukiza, 2012]. However, there is a growing range of innovative,
424 water and energy efficient, on-site alternatives [Jha, 2003; Bracken et al., 2007; Gates
425 Foundation, 2016] as well as a growing recognition that this is a social as well as physical or
426 technical issue [Burra, 2003; Sharma and Bhide, 2005; McFarlane, 2008].
427 It is important to emphasise that our results do not imply that open defecation is a safe
428 approach to sewage management. Water is not the only vector for faecal-oral disease;
429 transmission can also occur through food, insects, and direct contact [Wagner and Lanoix,
430 1958]. Thus safely disposing of faeces involves more than simply ensuring that they do not
431 enter the watercourse. There is good evidence to suggest that open defecation is extremely
432 problematic for public health and safety [Clasen et al., 2010; Mara et al., 2010; Ziegelbauer et
433 al., 2012].
434
435 **Network structure controls the spatial pattern of microbial pollution**
436 The relationship between upstream population density and FC concentration enables a simple
437 predictive relationship, albeit with considerable scatter. This model predicts that 33–48% of
438 rivers in the Ganga catchment fail the Indian Government's safe bathing standards,

439 depending on the choice of standard (Figure 6). However, many of those rivers that pass are
440 in sparsely populated headwaters. For 70-85% of the catchment's population, their nearest
441 river fails safe bathing standards [Central Pollution Control Board, 2008]; for 79% it should
442 not be used for flood irrigation, irrigation of crops eaten raw or where children are involved
443 in farming [WHO, 1989; Blumenthal et al., 2000]; and for 51% it should not be used for
444 irrigation with sprinklers [Blumenthal et al., 2000].
445 The pattern of predicted FC concentration from this empirical model is strongly influenced
446 by the catchment's network structure (Figure 6). Sparsely populated Himalayan headwaters
447 produce high discharges of clean water suppressing FC concentrations far downstream;
448 without this discharge, plains-fed rivers (e.g. Kali) have high FC concentrations throughout.
449 The most polluted reach of the Ganga is predicted to be between Kanpur and Allahabad.
450 Upstream of Kanpur the diluting effect of the headwaters persists while downstream of
451 Allahabad the Ganga is diluted first by the less polluted Yamuna (strongly influenced by the
452 Chambal) and then by the large left bank tributaries with their headwaters in the Himalaya.
453 This may be the result of not only the topology but also the geometry of the network, since
454 the Ganga at Allahabad is at its furthest point from the mountain front meaning cleaner
455 Himalayan water must travel over a larger expanse of populated plain to reach that point.
456 Interventions high up the river network have the highest potential for impacting FC
457 concentration for a given FC flux reduction because: 1) lower discharge on these rivers
458 means that the same FC flux reduction will lead to a larger concentration reduction; and 2)
459 rivers are directed networks (i.e. they accumulate) thus a reduction in FC flux at a given
460 location will impact only reaches downstream of it. Decisions of what to do where are
461 difficult and necessarily political, with many drivers [Bulkeley and Mol, 2003], but the
462 findings of this study can help guide strategic investment in pollution reduction.
463

Conclusions

The rivers of the Ganga catchment are subject to widespread and, in places, severe microbial pollution. 52-67% of measured sites fall below the Indian Government's upper and desirable limits for safe bathing; and for 61-70% of the population, model results suggest that their nearest river falls below these same bathing standards. The network structure of the Ganga catchment pre-conditions certain rivers to be highly polluted, and others (with large Himalayan headwaters) to be more robust against pollution, despite their location on the densely populated plains. The entire population upstream (not only those nearby) contribute to microbial river pollution but urban populations contribute more pollution per capita than rural populations. How much more depends on their respective population densities. A person living in an area with 1000 #/km^2 contributes on average 100 times more pollution to the river than they would in an area with 100 #/km^2. While this is an average in the presence of considerable (asymmetric) variability, the denser population in this case contribute at least as much pollution per capita at the lower limit and up to 10,000 times more at the upper limit. Densely populated areas dominate surface water pollution in the Ganga catchment not only because of they contain many people but because their faeces are more efficiently delivered to the river network. We suggest that this increasing efficiency reflects: the transmission speed of urban sewerage systems, delivering the coliforms to the river more quickly with less die-off; and the limited capacity for sewage treatment within these systems. Addressing this problem requires investment in both sewage removal and treatment whether by increasing existing sewerage capacity or implementing decentralised treatment solutions.

Acknowledgements

Data for the project was drawn from the Ganga River Basin Management Plan (GRBMP) GeoSpatial Database; the authors thank the members of the GRBMP consortium for their

time and effort in sample collection and analysis. The authors would also like to thank Mariano Moreno-de las Heras, Mary Power and Nick Cox for informative conversations on this project and five anonymous reviewers for their constructive comments. The project was supported by: NERC Fellowship NE/H015949/1 and a grant from UKIERI.

References

Allen, M.E. *Cleansing the city: sanitary geographies in Victorian London*. Ohio University Press (2008).

American Public Health Association. Standard methods for the examination of water and wastewater. *American Public Health Association, Washington, DC* (1995).

American Society of Agricultural Engineers. ASAE Standards, 45th Edition. Standards, Engineering Practices, Data (1998).

Ansari, A. et al. Role of monsoon rain on concentrations and dispersion patterns of metal pollutants in sediments and soils of the Ganga Plain, India. *Environmental Geology*, **39**, 221-237 (2000).

Ashbolt, N.J. et al. Indicators of microbial water quality. *IWA Publishing*, 289-316 (2001).

Ashbolt, N.J. Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology*, **198**, 229-238 (2004).

Ashbolt, N.J., Dorsch, M.R., Cox, P.T. and Banens, B. Blooming E. coli, what do they mean? In Coliforms and E. coli, Problem or Solution? (eds D. Kay and C. Fricker), pp. 78–85, The Royal Society of Chemistry, Cambridge (1997).

Baghel, V.S. et al. Bacterial indicators of faecal contamination of the Gangetic river system right at its source. *Ecological Indicators*, ***5***, 49-56 (2005).

Balk, D. & Yetman, G. The global distribution of population: evaluating the gains in resolution refinement. *New York: Center for International Earth Science Information Network (CIESIN), Columbia University* (2004).

Balk, D. et al. Construction of gridded population and poverty data sets from different data sources. In *Proceedings of European Forum for Geostatistics Conference* (2010).

Beaureau of Indian Standards. Drinking water specification, IS1500:1991 eighth reprint, UDC 628.1.033.Manak Bhavan, New Delhi.
https://law.resource.org/pub/in/bis/S06/is.10500.1991.pdf (2010)

Benzerzour, M. et al. Simulation of the urban climate variations in connection with the transformations of the city of Nantes since the 17th century. *Building and Environment*, ***46***, 1545-1557 (2011).

Bhargava, D.S. Use of water quality index for river classification and zoning of Ganga River. *Environmental Pollution Series B, Chemical and Physical*,6, 51-67 (1983).

Blumenthal, U.J. et al. Guidelines for the microbiological quality of treated wastewater used in agriculture: recommendations for revising WHO guidelines. *Bulletin of the World Health Organization*, ***78***, 1104-1116 (2000).

Bracken, P. et al. The road not taken: how traditional excreta and greywater management may point the way to a sustainable future. *Water Science and Technology: Water Supply*, **7**, 219-227 (2007).

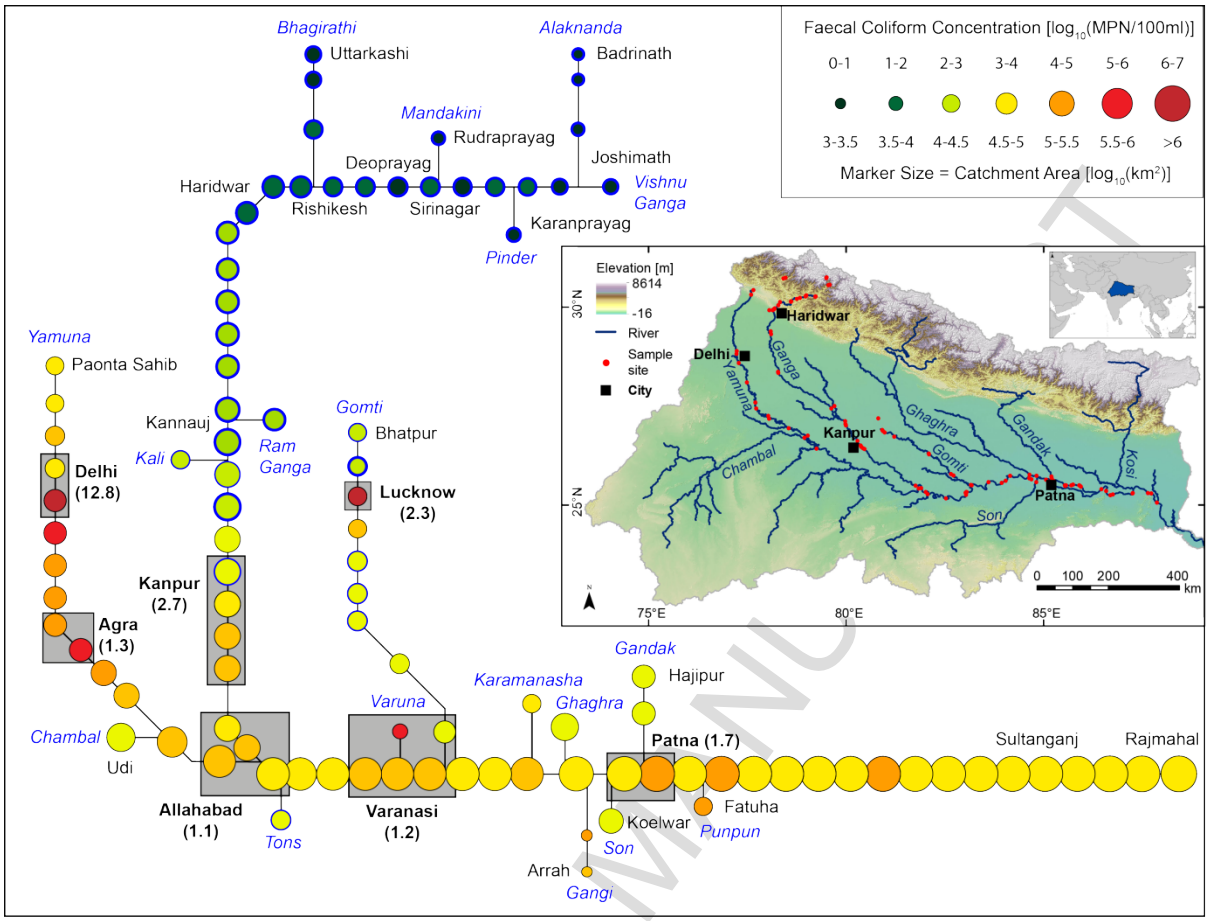
Breusch, T.S. and Pagan, A.R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pp.1287-1294 (1979).

Bulkeley, H., & Mol, A. P. Participation and environmental governance: consensus, ambivalence and debate. *Environmental Values*, ***12***, 143-154 (2003).
Burra, S., Patel, S., and Kerr, T. Community-designed, built and managed toilet blocks in Indian cities. *Environment and Urbanization*, **15**(2), 11-32 (2003).
Census of India: http://censusindia.gov.in/2011census/hlo/Data_sheet/India/Latrine.pdf (2011).
Census of India: http://censusindia.gov.in/2011-prov-results/paper2/data_files/India2/1.%20Data%20Highlight.pdf (2011).
Central Pollution Control Board. Ganga water quality trend. Monitoring of Indian Aquatic Resources Series MINARS/31/2009-10. http://cpcb.nic.in/upload/NewItems/NewItem_168_CPCB-Ganga_Trend%20Report-Final.pdf2 (2010).
Central Pollution Control Board. Performance evaluation of sewage treatment plants under NRCD (2013).
Central Pollution Control Board. Performance of sewage treatment plants – coliform reduction. http://www.cpcb.nic.in/divisionsofheadoffice/pams/pstp-colliform.pdf (2008).
Central Pollution Control Board. *Status of Water Supply, Wastewater Generation and Treatment in Class-I cities and Class-II towns of India* (2009).
Cho, K.H., Pachepsky, Y.A., Kim, J.H., Kim, J.W. and Park, M.H., 2012. The modified SWAT model for predicting fecal coliforms in the Wachusett Reservoir Watershed, USA. *water research*, *46*(15), pp.4750-4760.
Clasen T. F. et al. Interventions to improve disposal of human excreta for preventing diarrhoea. *Cochrane Database Syst Rev* **6**, doi: 10.1002/14651858.cd007180.pub2 (2010).
Crane, B.S. and Moore, J.A., 1986. Modeling enteric bacterial die-off: a review. *Water, air, and soil Pollution*, *27*(3-4), pp.411-439.
Farr, T. G., et al. The Shuttle Radar Topography Mission, Rev. Geophys., 45, RG2004, doi:10.1029/2005RG000183 (2007).
Field, K.G. and Samadpour, M. Fecal source tracking, the indicator paradigm, and managing water quality. Water research, **41**(16), pp.3517-3538 (2007).
Gandy, M. Water, modernity and emancipatory urbanism. *The emancipatory city. London: Sage*, 178-191 (2004).
Gates Foundation: https://docs.gatesfoundation.org/documents/Reinvent%20the%20Toilet%20Fair%20India%202014%20Technical%20Guide.pdf
Granger, D.E., Kirchner, J.W. and Finkel, R. Spatially averaged long-term erosion rates measured from in situ-produced cosmogenic nuclides in alluvial sediment. *The Journal of Geology*, *104*(3), pp.249-257 (1996).
Harwood, V.J., Staley, C., Badgley, B.D., Borges, K. and Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS microbiology reviews*, *38*(1), pp.1-40 (2014).
Jha, P. K. Health and social benefits from improving community hygiene and sanitation: an Indian experience. *International journal of environmental health research*, **13**, S133-S140 (2003).
Katukiza, A.Y. Sustainable sanitation technology options for urban slums. *Biotechnology advances*, ***30***, 964-978 (2012).
Kay, D. et al. Faecal indicator organism concentrations and catchment export coefficients in the UK. *Water Research*, ***42***, 2649-2661 (2008).
Koenker, R. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, *17*(1), pp.107-112 (1981).

583 Kotloff, K.L. et al. Burden and aetiology of diarrhoeal disease in infants and young children
584 in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-
585 control study. *The Lancet*, **382**, 209-222 (2013).
586 Kramer, M.H., Herwaldt, B.L., Craun, G.F., Calderon, R.L. and Juranek, D.D. Surveillance
587 for waterborne-disease outbreaks: United States, 1993–4. *Morbid. Mortal Weekly Rep*. **45**, 1–
588 33 (1996).
589 Levy, D.A., Bens, M.S., Craun, G.F., Calderon, R.L. and Herwaldt, B.L. Surveillance for
590 waterborne-disease outbreaks: United States, 1995–6. *Morbid. Mortal. Weekly Rep.* **47**, 1–34
591 (1998).
592 Mara, D. et al. Sanitation and health. *PLoS Med*, *7*, p.e1000363 (2010).
593 Mara, D., *Domestic wastewater treatment in developing countries*. Routledge (2013).
594 McFarlane, C. Governing the contaminated city: infrastructure and sanitation in colonial and
595 postcolonial Bombay. *International journal of urban and regional research*, **32**(2), 415-435
596 (2008).
597 McGrane, S.J. et al. Application of a linear regression model to assess the influence of
598 urbanised areas and grazing pastures on the microbiological quality of rural
599 streams. *Environmental monitoring and assessment*, **186**, 7141-7155 (2014).
600 Mishra, A. et al. Seasonal and temporal variations in physico-chemical and bacteriological
601 characteristics of River Ganga in Varanasi. Int. J. Environ. Res., **3**, 395-402 (2009)
602 Mukherjee, D. et al. Water quality of the River Ganga (The Ganges) and some of its physico-
603 chemical properties. *Environmentalist*, *13*, 199-210 (1993).
604 Narain, S. Sanitation for all. *Nature*,***486***, 185-185 (2012)
605 Niemi, R.M., Niemelä, S.I., Lahti, K. and Niemi, J.S. Coliforms and E. coli in Finnish surface
606 waters. In Coliforms and E. coli. Problems or Solution? (eds D. Kay and C. Fricker), pp.
607 112–119, The Royal Society of Chemistry, Cambridge (1997).
608 O'Callaghan, J.F. & Mark, D.M. The extraction of drainage networks from digital elevation
609 data. *Computer vision, graphics, and image processing*, ***28***, 323-344 (1984).
610 Portenga, E.W., Bierman, P.R., Duncan, C., Corbett, L.B., Kehrwald, N.M. and Rood, D.H.
611 Erosion rates of the Bhutanese Himalaya determined using in situ-produced 10
612 Be. *Geomorphology*, *233*, pp.112-126 (2015).
613 Prüss-Ustün, A. et al. Burden of disease from inadequate water, sanitation and hygiene in
614 low-and middle-income settings: a retrospective analysis of data from 145 countries. *Tropical*
615 *Medicine and International Health*, **19**, 894-905 (2014).
616 Robinson, T.P. et al. Mapping the global distribution of livestock. *PloS one*, ***9***, p.e96084
617 (2014).
618 Sadeghi, A.M. and Arnold, J.G., 2002. A SWAT/microbial sub-model for predicting
619 pathogen loadings in surface and groundwater at watershed and basin scales. In *Total*
620 *Maximum Daily Load (TMDL): Environmental Regulations, Proceedings of 2002*
621 *Conference* (p. 56). American Society of Agricultural and Biological Engineers.
622 Schwanghart, W. & Scherler, D. Short Communication: TopoToolbox 2-MATLAB-based
623 software for topographic analysis and modeling in Earth surface sciences. *Earth Surface*
624 *Dynamics*, ***2***, 1 (2014).
625 Sharma, R.N., and Bhide, A. World Bank Funded Slum Sanitation Programme in Mumbai:
626 Participatory Approach and Lessons Learnt. *Economic and Political Weekly*, 1784-1789
627 (2005).
628 Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R. and Palmer, C.J. Sources of
629 Escherichia coli in a coastal subtropical environment. *Appl. Environ. Microbiol.* **66**, 230–237
630 (2000).

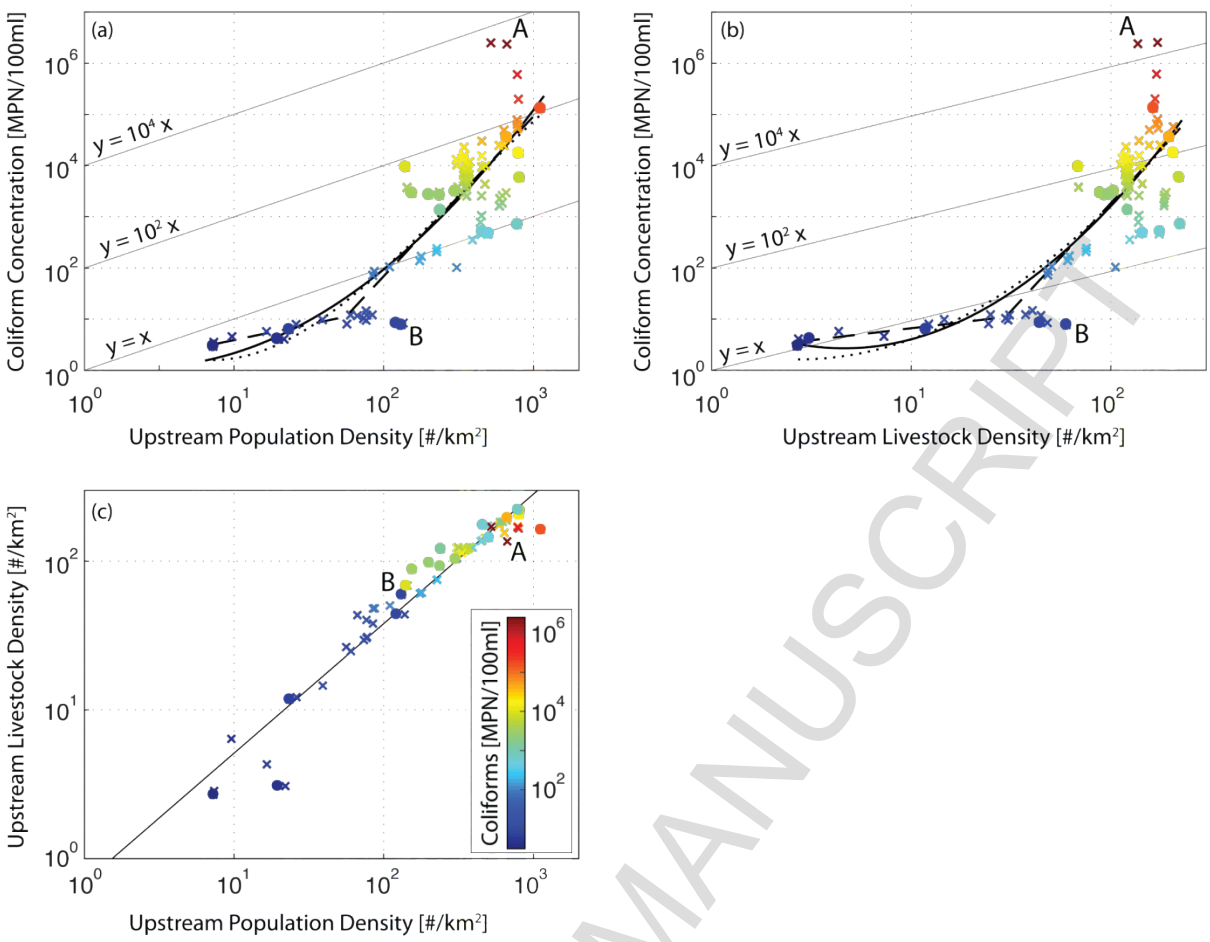
631 Sowah, R. et al. Evaluating the influence of septic systems and watershed characteristics on
632 stream faecal pollution in suburban watersheds in Georgia, USA. *Journal of applied*
633 *microbiology*, ***117***, 1500-1512 (2014).
634 Tchobanoglous et al. Wastewater Engineering: Treatment, Disposal, Reuse. 4th ed. McGraw-
635 Hill, Inc., New York (1991).
636 U.S. EPA. Quality Criteria for Water. U.S. Environmental Protection Agency: Washington
637 DC (1976).
638 Wagner, E. G. & Lanoix, J. N. Excreta disposal in rural areas and small communities.
639 Geneva: World Health Organization. 327 (1958).
640 White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for
641 heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pp.817-838 (1980).
642 Winslow, C.E.A. and Walker, L.T. *Science* **26**, 797 (1907).
643 Wint G.R.W. & Robinson T.P. Gridded Livestock of the World, Food and Agriculture
644 Organization of the United Nations, Rome. 131 (2007).
645 World Health Organisation. Guidelines for drinking-water quality, fourth edition
646 incorporating the first addendum. *World Health Organization.* 380 (2017).
647 World Health Organization. Health guidelines for the use of wastewater in agriculture and
648 aquaculture: report of a WHO scientific group (1989).
649 World Health Organisation and United Nation Children's Fund. Progress on Sanitation and
650 Drinking Water – 2015 update and MDG assessment. New York: WHO/UNICEF Joint
651 Monitoring Programme for Water Supply and Sanitation.
652 http://apps.who.int/iris/bitstream/10665/177752/1/9789241509145_eng.pdf?ua=1 (2015).
653 Young, K.D. & Thackston, E.L. Housing density and bacterial loading in urban
654 streams. *Journal of Environmental Engineering*, ***125***, 1177-1180 (1999).
655 Ziegelbauer K. et al. Effect of sanitation on soil-transmitted helminth infection: systematic
656 review and meta-analysis. *PLoS Med* **9,** e1001162. doi: 10.1371/journal.pmed.1001162
657 (2012).
658

659 **Figures**



660

661 Figure 1. Network graph of decadal mean FC concentrations (circle colour) and catchment area (circle size).
662 Large red circles indicate high FC concentration and water discharge (thus high FC flux); smaller green circles
663 indicate lower concentration and discharge (thus low FC flux). Sites with thick blue outlines pass Indian
664 Government desirable standards of <500 MPN / 100 ml; those with thin blue outlines pass the upper limit of
665 <2500 MPN / 100 ml [Central Pollution Control Board, 2008]. Rivers are labelled in blue; cities are labelled in
666 black, with approximate populations, in millions, in brackets and grey boxes to show approximate extent. Inset
667 shows a location map of the Ganga catchment.



668

669 Figure 2: catchment scale analysis of faecal coliform concentration against: a) upstream population density; b)
670 upstream livestock density adjusted for variable coliform production rates; c) co-variation between upstream
671 population and livestock density. Trend lines show quadratic (solid), cubic (dotted) and linear spline (dashed)
672 regressions for a and b, and linear regression for c. Solid circles show non-nested (i.e. independent)
673 observations, n=18; crosses show the full dataset, n=100. Labelled points are: A) Yamuna catchment at Delhi;
674 and B) Pinder catchment at Karanprayag.

675

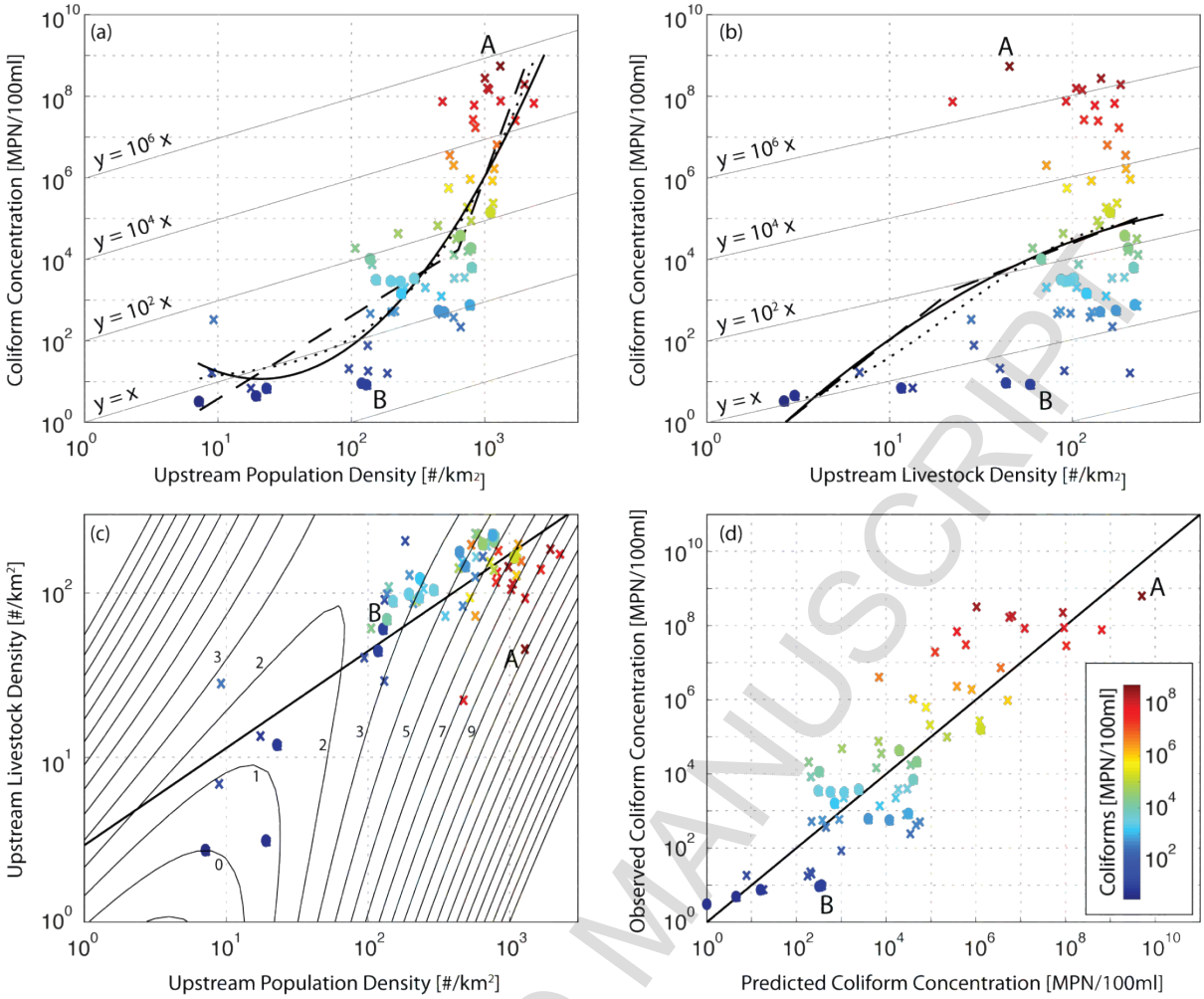
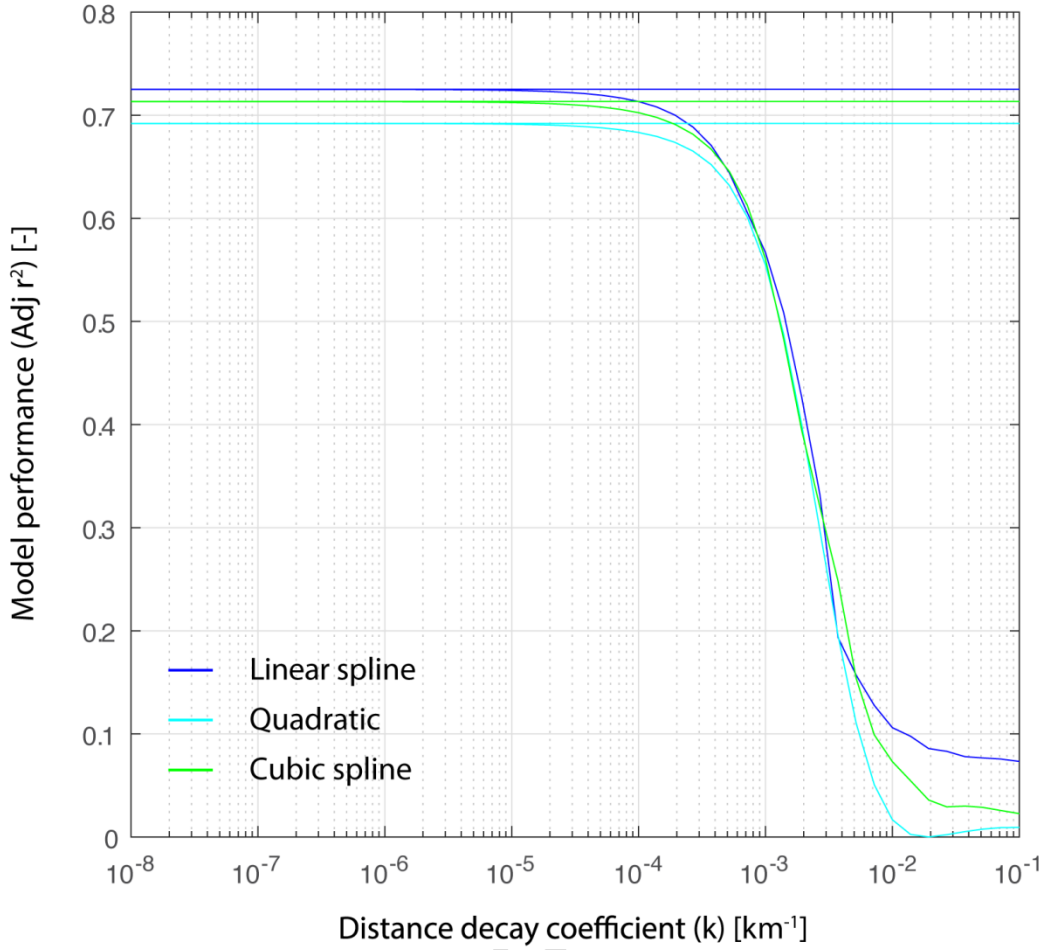
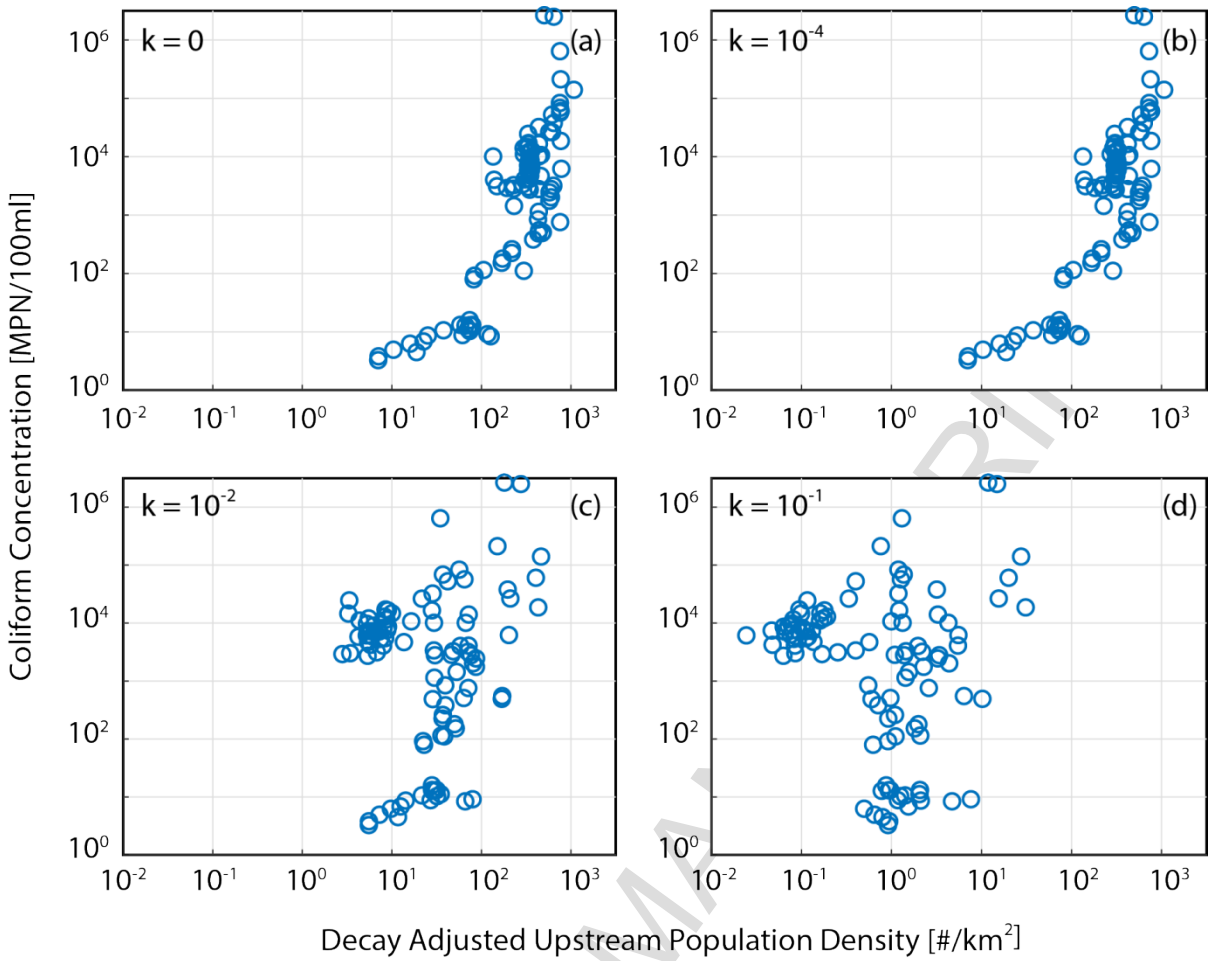


Figure 3: sub-catchment based Faecal Coliform concentration against: a) upstream population density and b) upstream livestock density adjusted for variable coliform production rates; c) co-variation between upstream population and livestock density; d) predicted v observed coliform concentrations from multiple cubic regression with upstream population and livestock density. Trend lines show quadratic (solid), cubic (dotted) and linear spline (dashed) regressions for a and b, and linear regression for c. Solid circles show non-nested (i.e. independent) observations, n=18; crosses show the full dataset, n=100. Labelled points are: A) Yamuna catchment at Delhi; and B) Pinder catchment at Karanprayag. Contours in c show prediction surface from multiple regression.



686

687 Figure 4. Model performance (Adjusted r^2 for FC concentration v decay-adjusted UPD) with varying distance
688 decay coefficient (k) for the three empirical functions fitted in Figure 2. Best performance is always for no
689 decay (k=0); small coefficients ($k < 10^{-4}$) have little effect; larger coefficients result in a breakdown in model
690 performance.



691

692 Figure 5. Scatter plots of faecal coliform (FC) concentration against upstream population density (UPD)
693 adjusted with an exponential distance decay using a range of decay coefficients (k). Panels reflect decay rates
694 of: a) 0 %/km, b) 0.01 %/km, c) 1 %/km and d) 10 %/km. Best model performance is for no decay (k=0); small
695 coefficients ($k < 10^{-4}$) have little effect; larger coefficients result in a breakdown in the relationship between
696 UPD and FC concentration.

697

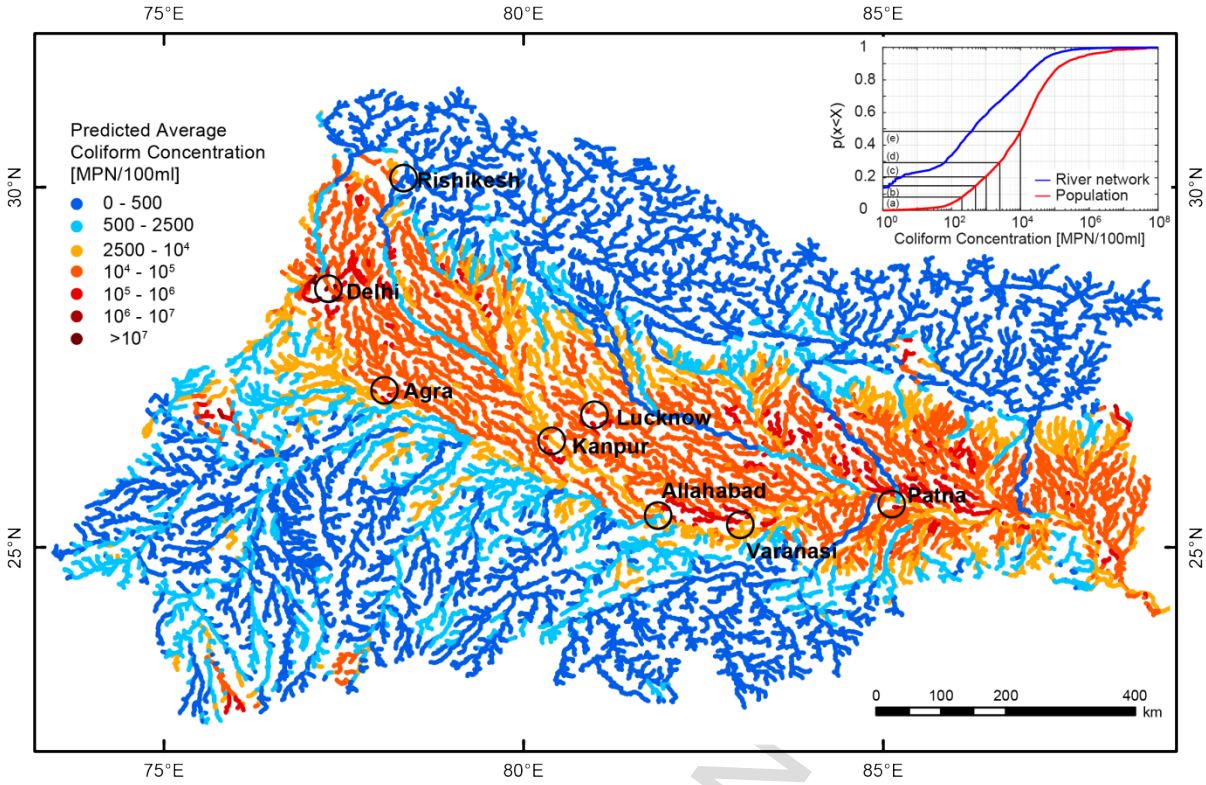


Figure 6. Spatial pattern of predicted coliform concentration. Dark blue areas have concentrations below 500 MPN/100ml, the Indian Government's desirable limit for safe bathing [Central Pollution Control Board, 2008]; light blue areas have concentrations below 2500 MPN/100ml, the upper limit for safe bathing [Central Pollution Control Board, 2008]. Inset shows the fraction of the river network (blue) and population (red) for which the nearest river has an FC concentration less than the x-axis value. Letters signify: (a) USA limit for safe bathing [U.S. EPA, 1976]; (b) Indian government desirable limit for safe bathing [Central Pollution Control Board, 2008]; (c) WHO recommended limit for flood irrigation, or for crops eaten raw, or where children are involved in farming [WHO, 1989; Blumenthal et al., 2000]; (d) Indian government upper limit for safe bathing [Central Pollution Control Board, 2008]; (e) WHO limit for sprinkler irrigation [Blumenthal et al., 2000].