

Short communication

River water quality trends and increased dairying in Southland, New Zealand

KEITH D. HAMILL*

Environment Southland
Private Bag 90 116
Invercargill, New Zealand
email: "hamill_k"@wrcplc.co.uk

*Present address: WRC plc, Frankland Road,
Blagrove, Swindon SN5 8YF, United Kingdom.

GRAHAM B. McBRIDE

National Institute of Water and Atmospheric
Research Limited
P.O. Box 11 115
Hamilton, New Zealand
email: g.mcbride@niwa.co.nz

Abstract Trends in six water quality variables measured at 29 Southland, New Zealand, freshwater sites between July 1995 and June 2001 have been assessed. These are compared to changes in livestock numbers to assess whether water quality changes could have been driven by the intensification and expansion of dairy farming and/or by other pressures. Trends were assessed for raw and for flow-adjusted data, the latter being a better indicator of effects of changing external pressures on water quality (an observed water quality trend may be merely the result of changes in river flow). Results indicate that increased dairy farming has been associated with increasing concentrations of dissolved reactive phosphorus. There has been a worsening in other water quality variables (oxidised nitrogen, dissolved oxygen) but these also occurred in non-dairying catchments. An improvement in observed water clarity disappeared after flow adjustment. Continued monthly water quality monitoring is necessary to watch for emerging trends.

Keywords water quality; trends; Southland; dairying; agriculture

INTRODUCTION

Farming practice in Southland, New Zealand, is typified by year-round grazing of pasture, and extensive use of nitrogen (N) and phosphorus (P) fertilisers. There have been considerable changes in the grazing animal populations over the last decade. Between 1990 and 1999 there were increased numbers of dairy cattle (from 38 000 to 233 000), beef cattle (from 187 000 to 230 000), and deer (from 124 000 to 393 000). Also in that decade there were decreased numbers of sheep (8.9 million to 6.7 million), goats (from 73 000 to 7000), and pigs (from 11 000 to 3000) (www.maf.govt.nz/statistics/primaryindustries/regions/graphs/deerpigsgoats/Southland.htm; www.maf.govt.nz/statistics/primaryindustries/regions/tables/southland.htm). At the same time there has been community interest in the effects of these changes on the region's water quality. In particular, in view of the rapid increase in dairying, there has been widespread concern over the effects of dairy farm expansions on Southland's river water quality.

There is growing evidence indicating that agriculture can detrimentally affect water quality (Smith et al. 1993; Harding et al. 1999; Parkyn et al. 2002). In addition, it has been found that the degree of impact can correspond to the density of livestock in the catchment (as reported for an Otago river by Harding et al. 1999). Despite this, few reports in the literature have linked widespread changes in water quality with land use changes in New Zealand. This may be partially because of the difficulty in finding suitable baseline sites when designing a monitoring network—more often than not baseline sites are not "pristine" but are merely less impacted by human activity than are the "impact" sites. It may also reflect the fact that water quality may respond slowly to more rapid changes in land use.

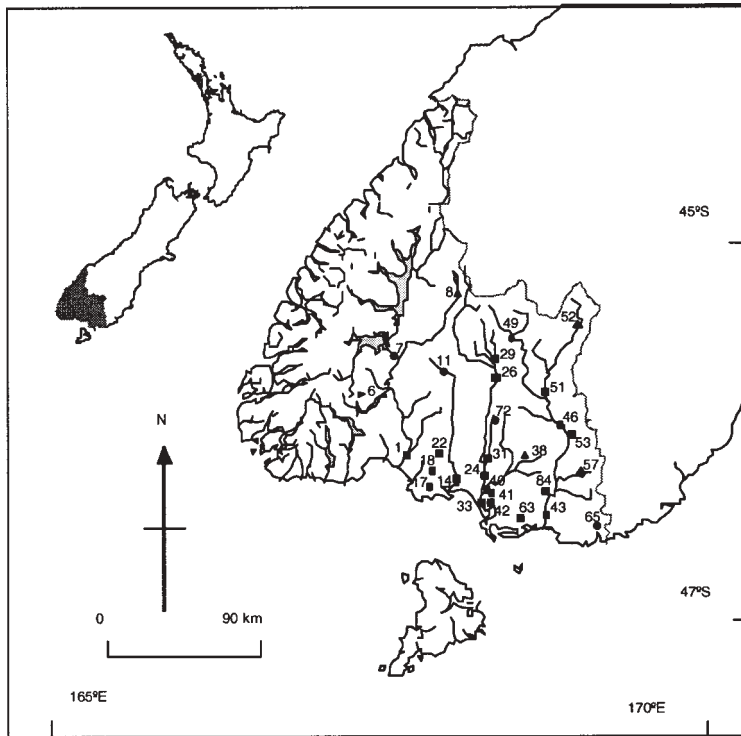


Fig. 1 Site locations. (Names of sites are given on Table 1. Site identifiers: squares, dairying in upstream catchment; circles, pasture but no dairying in upstream catchment; triangles, minimal pasture in upstream catchment.)

An alternative to analysing water quality data within an impact/baseline framework is to directly compare a change in the state of water quality with a change in a pressure on water quality—such as land use. That is the approach adopted here. We use 23 Southland Regional Council and six water quality monitoring sites operated by the National Institute of Water and Atmospheric Research Limited (NIWA) at which nutrients, visual clarity, electrical conductivity, pH, temperature, and dissolved oxygen (DO) have been recorded each month over the period July 1995 – June 2001. Dairy farming was present in the catchments above 19 of these sites, all of which had increases in cow numbers since 1995. Of the other 10 sites, half contain significant pasture development, including areas of deer farming, whereas the rest may be regarded as baseline sites, being either undeveloped catchments or in exotic forest with low animal numbers.

We report on trends in water quality variables at these sites, and relate them to changing land use.

METHODS

Data collection

Data used for trend analysis comes from 23 sites monitored since 1995 in Southland Regional Council's water quality monitoring network (Hamill 2002) and six sites monitored from 1989 in NIWA's National River Water Quality Network (Smith & Maasdam 1994)—see Fig. 1 and Tables 1 and 2. Where rivers had two (Mararoa, Pourakino, Oreti, Waikaia Rivers) or more (Mataura River) sites, we have retained only those sites that are far apart and have differentiated land use. Two sites (numbers 6 and 8) are located below lakes, both being in minimal pasture catchments.

Over the 6-year period from July 1995 to June 2001, monthly data were available from these sites for oxidised N ($\text{NO}_x\text{-N} = \text{NO}_2\text{-N} + \text{NO}_3\text{-N}$), ammoniacal N (i.e., $\text{NH}_4\text{-N} = \text{NH}_4^+\text{-N} + \text{NH}_3\text{-N}$), dissolved reactive phosphorus (DRP), DO (% saturation), visual clarity (measured by horizontal

black disc—Davies-Colley 1988), electrical conductivity (at 25°C), pH, and temperature (some other data, e.g., faecal coliforms and turbidity, were available at these and at other sites, but not at a density suitable for the present analysis). The methods of collecting and analysing water quality samples were constant over this period for all parameters except visual clarity from Southland Regional Council sites. Non-detects were recorded for NO_x (11.3% of all data), DRP (13.8%), and NH₄

(29%). Southland Regional Council improved the quality control of visual clarity measurements in July 1998. Scratches accumulated on the black disk viewers before July 1998 had reduced visual clarity measurements by 8–13% compared with new viewers. Before-analysis clarity data was adjusted for this measurement error by adding 11% to results obtained before July 1998. Data collection and laboratory analysis methods are described in Hamill (2002) and Smith et al. (1996).

Table 1 Percentage pasture, catchment area, and Relative Seasonal Kendall Sen Slope Estimator (RSKSSE) for flow-adjusted data (water temperature results not flow-adjusted). (Temp., water temperature; NH₄-N, ammoniacal N; NO_x-N, oxidised N; DRP, dissolved reactive phosphorus; EC, electrical conductivity at 25°C; % DO, percentage saturation dissolved oxygen; clarity, horizontal black disk visibility.)

Site identifier and name		Pasture (%)	Area (km ²)	RSKSSE (% of median value per annum)							
				Flow	Temp.	NH ₄ -N	NO _x -N	DRP	EC	% DO	Clarity
Non-dairying											
Minimal pasture											
6	Monowai R., d/s gates	0	248	-1.3	0.9	-3.7	-1.9	-7.3	-0.3	0.0	0.9
8	Mararoa R., Mavora Lake	1	342	-5.0	2.0	0.0	-0.3	-0.7	-1.0	-0.2	-1.0
38	Dunsdale Stm, Dunsdale Res.	2	54	-7.7	1.0	-1.2	10.6	0.4	-1.1	-0.4	-3.3
52	Waikaia R., u/s Piano Flat	0	470	3.2	1.2	-0.2	0.1	-3.2	0.0	-0.2	-2.7
57	Mimihau Stm (Venlaw Forest trib.)	1	4	-13.5	0.3	-3.0	6.0	-2.3	1.3	-0.5	-0.6
Non-dairying pasture											
7	Mararoa R., u/s Weir Rd	40	12 162	-4.5	2.0	0.2	5.7	-0.8	-0.2	0.3	-14.9
11	Aparima R., d/s Dunrobin Bridge	23	220	-5.5	2.5	-1.2	2.5	3.1	-0.9	-0.6	-7.7
49	Mataura R., Parawa	33	772	-10.3	1.4	4.0	1.7	2.9	-0.2	-0.3	2.1
65	Waikawa R., Progress Valley	53	179	-8.4	0.1	-4.6	4.2	2.2	-0.3	-1.6	0.5
72	Trenders Ck, Hall Rd	31	32	-7.6	0.1	-2.4	12.3	1.4	-1.2	-0.5	1.9
Median				-6.6	1.1	-1.2	3.4	-0.2	-0.3	-0.4	-0.8
Mean				-6.06	1.15	-1.21	4.09	-0.43	-0.39	-0.4	-2.48
Dairying											
1	Waiau R., Tuatapere	20	7718	-1.9	1.5	-1.9	3.8	-2.7	-0.1	-0.1	6.8
14	Aparima R., Thornbury	62	1254	-7.0	-0.6	-5.1	6.5	0.7	0.0	-1.5	-5.8
17	Pourakino R., Centre Rd	29	188	-11.5	1.0	-1.6	1.3	0.4	-5.2	-0.9	-4.2
18	Pourakino R., Ermedale Rd	8	90	-11.1	1.6	-0.7	4.5	0.3	1.9	-0.7	-3.1
22	Otautau Stm, Tuatapere Rd	79	264	1.0	1.3	-9.1	-2.2	1.6	3.3	-1.5	21.0
24	Oreti R., Wallacetown	57	2141	-9.9	0.1	16.8	5.4	1.9	0.8	0.2	-2.2
26	Oreti R., Lumsden Br.	33	1129	-12.8	2.0	5.5	3.1	6.1	-0.3	-0.2	-0.4
29	Irthing Stm, Ellis Rd	50	456	-9.7	1.4	-6.2	2.3	-2.1	-0.6	-0.4	-2.3
31	Winton Stm, Lochiel	13	141	0.0	1.0	-12.4	-0.1	11.7	0.9	0.3	15.2
33	Makarewa R., Moffett Rd	73	1110	-9.8	1.4	-2.0	3.9	1.1	-2.8	-0.2	1.2
40	Waikiwi Stm, North Rd	99	119	-3.2	-0.5	-3.3	-0.9	-2.9	0.6	0.3	11.5
41	Waihopai Stm, u/s Queens Drive	99	176	-2.6	-0.5	0.7	-1.1	1.3	0.2	0.2	2.2
42	Otepunu Ck, Nith St	94	33	-10.7	-0.9	3.4	-2.7	1.4	1.1	-3.2	1.4
43	Mataura R., Gorge Rd	60	5355	-3.2	0.6	-3.8	4.0	1.2	0.6	-0.8	0.0
46	Mataura R., Otamita Br.	46	3367	-5.6	0.4	-5.4	0.7	3.8	-0.7	-0.5	-6.2
51	Waikaia R., Waipounamu Br.	25	4059	-3.2	1.7	-1.1	1.2	-0.5	-0.6	-0.2	-0.9
53	Waikaka Stm, Gore	95	1281	-8.0	-1.0	0.6	-0.9	5.3	0.6	-0.5	-1.6
63	Waituna Ck, Marshall Rd	87	98	-2.0	-2.7	9.5	1.3	7.6	1.1	-3.7	-3.8
84	Oteramika Stm, d/s Seaward Downs	95	53	-6.7	1.6	-12.5	-0.5	-2.6	0.3	-1.1	9.1
Median				-6.7	1.0	-1.9	1.3	1.2	0.3	-0.5	-0.4
Mean				-6.21	0.49	-1.51	1.56	1.77	0.06	-0.76	1.99

Trend analysis methods

We need a robust measure of trend because water quality data such as we deal with here often come from distinctly non-normal, seasonally varying populations, and contain occasional missing and censored values ("less thans"). Accordingly, we used the seasonal Kendall Sen Slope estimator (Gilbert 1987; Helsel & Hirsch 1992) to assess the magnitude of trends for each variable at each site between July 1995 and June 2001. Monthly sampling over this 6-year period gives adequate trend detection power (for which 5 years of data may suffice, e.g., Smith et al. 1996). This measure is the median of all combinations of within-month slopes. That is, all slopes between each January are computed (for 6 years of data there are 21 such slopes), then for each February, etc. No slopes are computed between-months (e.g., the slope between January 1995 and February 1996 is not computed), and by this means seasonality is accounted for. So for 6 years of monthly data the Sen slope is the median of 252 possible slopes. We calculated it by employing the WQSTAT PLUS software package (IDT 1998), using 12 "seasons" per year. For each water quality variable at each site the trend is reported as a relative slope, being the estimated slope divided by the variable's median (over all 72 data points), denoted as RSKSSE (Relative Seasonal Kendall Sen Slope Estimator).

We have analysed trends in both the raw data and in the flow-adjusted water quality data. Whereas the former reflect the trends actually occurring, the latter are more relevant to the objectives of this paper. This is because trends can be associated with increased pressures, such as agricultural development, but they may also be the result of trends in river flows over

the trend assessment period (water quality variables may change with flow as a result of dilution or wash-off effects, as discussed by Smith et al. 1996). Trend analysis on flow-adjusted data removes a substantial part of the variability associated with flow, and so gives indications of trends being associated with other pressures, particularly changes in agricultural practice or climate. This sequential procedure (first adjust for flow dependency, then test residuals for trend) is permissible unless flows are subject to substantial human manipulation (Helsel & Hirsch 1992, section 12.3). Flow adjustment was carried out for each water quality variable at each site by fitting a smoothing power law relationship between the variable and river flow, and then defining adjusted data as: raw datum – smoothed datum + median of all data for that variable at that site (as provided in the WQSTAT PLUS package). Trends presented for water temperature are not flow-adjusted, as their raw values are important in considering effects of climate changes.

For the purposes of assessing the overall direction of trends in each water quality variable at the regional scale, we tested the null hypothesis that **in fact** there is no regional trend whatsoever. That is, this hypothesis posits that at any one site one is equally likely to obtain an upward or a downward slope for that variable and so the true probability of obtaining an upward or a downward slope is $\frac{1}{2}$. Of course, even if this hypothesis were true (or at least approximately so) statistical sampling error means that one would seldom obtain an equal number of upward and downward slopes for the sites. Recognising this, we have calculated traditional *P* values of the non-parametric (binomial) test to examine the **relative** merits of this hypotheses for

Table 2 Binomial *P* values. Bold values are statistically significant ($P \leq \alpha = 0.05$); trend direction is shown by vertical arrows if $P \leq 0.20$. (Flow-adj., *P* value attained after adjusting for flow dependency; NH₄-N, ammoniacal nitrogen; NO_x-N, oxidised N; DRP, dissolved reactive phosphorus; EC, electrical conductivity at 25°C; % DO, percentage saturation dissolved oxygen; clarity, horizontal black disk visibility.)

Water quality variable	All sites		Non-dairying sites		Dairying sites	
	Raw	Flow-adj.	Raw	Flow-adj.	Raw	Flow-adj.
Flow	<0.0001 ↓	–	0.011 ↓	–	0.0001 ↓	–
Water temp.	0.0012 ↑	–	0.001 ↑	–	0.084↑	–
NH ₄ -N	0.048 ↓	0.018 ↓	0.50	0.090↓	0.059↓	0.084↓
NO _x -N	0.047 ↑	0.031 ↑	0.063↑	0.055↑	0.23	0.18↑
DRP	0.41	0.068↑	0.50	0.62	0.27	0.032 ↑
EC	0.0001 ↑	0.35	0.50	0.020 ↓	<0.0001 ↑	0.24
% DO	0.0002 ↓	0.0005 ↓	0.020 ↓	0.020 ↓	0.0038 ↓	0.0096 ↓
Clarity	0.0002 ↑	0.29	0.090	0.38	0.0007 ↑	0.41

each variable. If one water quality variable has a lower value of P than does a second variable, the evidence for a trend in the first is the stronger. The mechanics of the calculation were as follows. If the obtained proportion of positive RSKSSE values exceeded $\frac{1}{2}$, the P value of this test was calculated (using a binomial calculator) as the probability of getting at least as many positive RSKSSE values as were obtained if the true probability of an upward slope is $\frac{1}{2}$ (and similarly if the obtained proportion of negative slopes exceeded $\frac{1}{2}$). Following standard conventions we have also denoted any analysis with a P value ≤ 0.05 as “statistically significant”. Furthermore, a trend direction has been stated in our aggregated results if $P \leq 0.20$ —a convenient but arbitrary cut-off (i.e., not statistically significant by the usual convention but still considered noteworthy). Finally, if there are an appreciable number of tied values in the data the RSKSSE procedure can produce a slope of exactly zero. In that case the occurrences of zero slope have been excluded from the binomial calculations. Such cases arise when there is a number of non-detects all of which are replaced by a single value, as occurred particularly for raw data for nutrients at un-impacted sites with minimal pasture (as discussed by Stansfield 2001).

Collating dairy cow changes per catchment

The number of dairy cows and total milk production from each water quality catchment was collated for the seasons 1992/93–2000/01, except for the 1996/97 season. These historical data were obtained from Fonterra, Southern Fresh Milk Company and Kiwi, who respectively had 555, 22, and five suppliers in Southland during the 2000/01 season. In general, dairying catchments have experienced a steady annual increase in cow numbers.

For seasons since 1997/98 the estimated number of cows for Fonterra suppliers was based on pre-season estimates of milking cows per farm. These were confirmed using a cow census in November 1999 and November 2000. Cow numbers were not recorded before 1997, so they were estimated by dividing the litres produced by the average litres per cow for each farm. Cow numbers for Southern Fresh Milk suppliers were based on recent Southland Regional Council compliance inspection sheets and estimated for earlier years as above. This may slightly underestimate dairy cow numbers before 1997 since production per cow has improved in the last decade.

GIS (Geographical Information Systems) techniques were used to aggregate data for all farms

within the total catchment upstream of water quality sites. The catchments were delineated from a digital terrain model (DTM) with a 30-m grid size used for the River Environment Classification (Snelder & Biggs 2002). Dairy farms were plotted at the position of the dairy shed using grid references from Southland Regional Council’s consent database. The property boundaries of dairy farms near the edge of each catchment were examined to ensure that farms were assigned appropriately. Farms were included within a catchment if more than half the farm was within the catchment boundary and excluded otherwise, ensuring that farms were not counted twice in adjoining catchments.

RESULTS

Table 1 lists catchment characteristics (upstream per cent pasture and catchment area) for each site. It also lists RSKSSE values for flow, raw water temperature, and for flow-adjusted values of the remaining water quality variables (NO_x , DRP, NH_4 , EC, % DO, visual clarity). These results are separated into dairying and non-dairying groups. Table 2 gives binomial test P values for each of these variables for the raw and flow-adjusted data. These are also presented separately for dairying and non-dairying sites. These P values have been derived from data in Table 1 (and similarly for the raw data, not shown) and convey the basic information about trends in the Southland region. Summary physico-chemical data are presented in Table 3, illustrating the difference between water quality of developed and undeveloped catchments and providing a context to the trend results.

Regional trends

Table 2 (“All sites” columns) indicates strongly that there was a decrease in river flow and an increase in water temperature over the monitoring period. Also indicated are upward trends in raw values of oxidised N, electrical conductivity, and water clarity, with downward trends in ammoniacal N and % DO. No trends in raw DRP data are indicated. However, the table also shows that some of this trend pattern was related to reducing river flows, as the trend in visual clarity virtually disappears in the flow-adjusted results, as it does also for electrical conductivity in the dairying catchments. After flow adjustment, the direction of trends in the remaining three variables (NO_x , NH_4 , and %DO) remained unchanged.

Notably, a low *P* value for DRP appears for the dairying catchments after flow adjustment.

Dairying and non-dairying comparison

In comparing the flow-adjusted values in Table 2 we see that, with one exception, trend directions at both dairying and non-dairying sites are the same as found for the overall regional trend. That exception is electrical conductivity, where there was a consistent downward trend at non-dairying sites. We note that trends in oxidised nitrogen (NO_x) both at dairying and at non-dairying sites were similar, but that there

was a much stronger trend in DRP at dairying sites as compared to the non-dairying sites.

Because the binomial test takes account only of the trend **direction** at each site, it can be instructive to examine the **magnitude** of the trends found. Figure 2 shows all flow-adjusted RSKSSE values for NO_x plotted against increases in cow density from the 1994/95 season to the 2000/01 season. The graph does suggest an increase in oxidised N with increase in cow numbers. However, it also shows an increase in oxidised N in the non-dairying pasture (and in a small catchment site with minimal pasture—

Table 3 Median water quality results for monthly sampling July 1995–June 2001 (temp., water temperature; NH₄-N, ammoniacal N; NO_x-N, oxidised N; DRP, dissolved reactive phosphorus; EC, electrical conductivity at 25°C; % DO, percentage saturation dissolved oxygen; clarity, horizontal black disk visibility.)

Site identifier and name	Number of cows 2000/01	Temp. (°C)	NH ₄ -N (mg m ⁻³)	NO _x -N (mg m ⁻³)	DRP (mg m ⁻³)	EC (mS cm ⁻¹)	% DO	Clarity (m)
Non-dairying								
Minimal pasture								
6 Monowai R., d/s gates	0	11.25	3	6	1	38.4	100.4	6.15
8 Mararoa R., Mavora Lake	0	10.25	10	10	5	30.8	88.5	5.50
38 Dunsdale Stm, Dunsdale Res.	0	7.7	10	180	14	117.2	95.0	1.50
52 Waikaia R., u/s Piano Flat	0	8.25	10	10	8	30.4	92.6	2.96
57 Mimiha Stm (Venlaw Forest trib.)	0	6.7	10	62	16	89.5	92.4	1.69
Non-dairying pasture								
7 Mararoa R., u/s Weir Rd	0	9.8	10	240	6	59.2	92.8	2.03
11 Aparima R., d/s Dunrobin Bridge	0	6.8	10	19	8	54.9	92.6	2.82
49 Mataura R., Parawa	0	9.2	7	240	6	48.6	98.0	1.82
65 Waikawa R., Progress Valley	0	9.3	25	700	16	143.3	91.4	0.62
72 Trenders Ck, Hall Rd	0	9.4	13	335	21	141.5	91.0	0.70
Median		9	10	121	8	57	93	2
Mean		9	11	180	10	75	93	3
Dairying								
1 Waiiau R., Tuatapere	3152	11.8	4	162	2	77.9	102.8	1.51
14 Aparima R., Thornbury	24741	12.4	14	560	10	114.9	91.8	1.36
17 Pourakino R., Centre Rd	2478	9.8	24	280	10	171.4	85.7	0.62
18 Pourakino R., Ermedale Rd	255	8.2	10	89	10	105.9	90.2	1.39
22 Otautau Stm, Tuatapere Rd	10998	9.2	35	840	22	186.4	85.7	0.63
24 Oreti R., Wallacetown	32745	12.35	9	735	6	83.5	101.9	1.14
26 Oreti R., Lumsden Br.	2225	10	5	381	3	51.4	98.2	3.27
29 Irthing Stm, Ellis Rd	2225	9	11	810	8	62.2	86.9	2.20
31 Winton Stm, Lochiel	1888	10.1	48	1400	43	218.0	93.0	0.76
33 Makarewa R., Moffett Rd	21734	10.9	340	1200	51	188.9	85.2	0.43
40 Waikiwi Stm, North Rd	3197	10.2	55	2400	19	201.0	89.9	0.71
41 Waihopai Stm, u/s Queens Drive	10365	10.75	41	1900	16	199.1	95.4	1.08
42 Otepunu Ck, Nith St	826	10.95	63	1350	21	240.5	89.9	0.81
43 Mataura R., Gorge Rd	60637	10.75	34	780	19	91.1	89.8	0.38
46 Mataura R., Otamita Br.	13515	9.7	11	530	12	58.4	89.2	1.12
51 Waikaia R., Waipounamu Br.	2610	9.5	10	420	11	48.3	87.1	1.70
53 Waikaka Stm, Gore	6565	9.9	51	590	26	129.9	88.1	0.76
63 Waituna Ck, Marshall Rd	7397	10	77	1200	28	204.5	82.0	0.50
84 Oteramika Stm, d/s Seaward Downs	5285	9.6	70	1650	40	190.4	87.5	0.56
Median		10	34	780	16	130	90	1
Mean		10	48	909	19	138	91	1

Fig. 2 Relative trend slopes for flow-adjusted oxidised nitrogen in the 29 sites versus cow density (RSKSSE, Relative Seasonal Kendall Sen Slope Estimator).

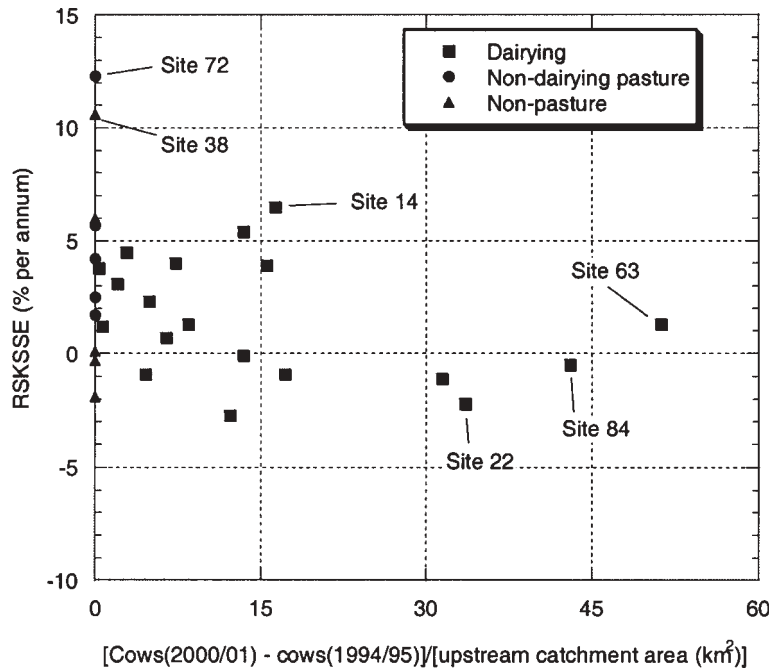
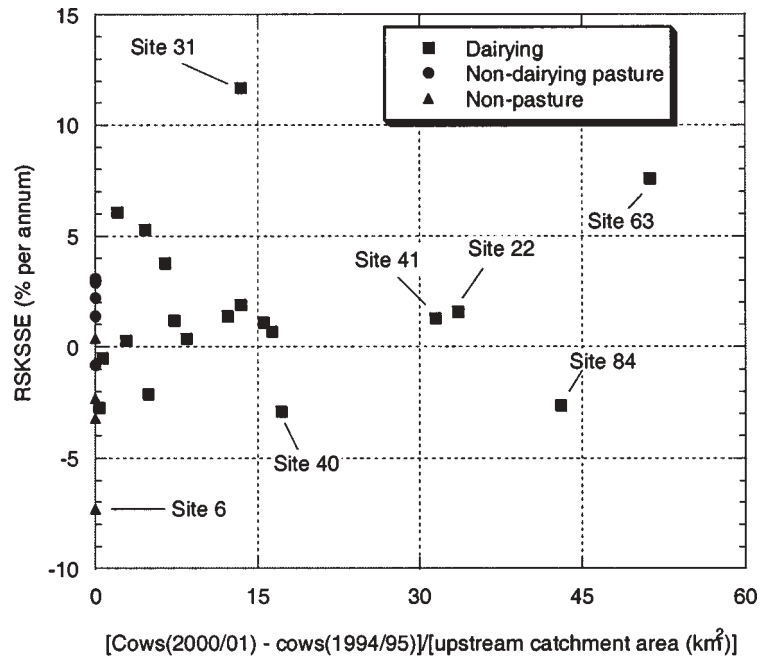


Fig. 3 Relative trend slopes for flow-adjusted dissolved reactive phosphorus in the 29 sites versus cow density (RSKSSE, Relative Seasonal Kendall Sen Slope Estimator).



Dunsdale Stream). A similar pattern emerges if one assumes a lag in the effect of increased dairying, e.g., by using as a metric the increase in cow numbers between the 1992/93 and 1998/1999 seasons. Of

course this does not preclude the possibility that evidence of longer lag periods may eventually emerge, allowing for the travel time of nutrients through soil and groundwater.

Figure 3 plots the DRP in the same way, indicating a distinction between the catchments with increases in cow density and the non-dairying sites. In this case the range of RSKSSE values in the non-dairying sites is smaller than for NO_x , as is reflected in the difference in P values for flow-adjusted DRP between dairying and non-dairying sites. That is, there is evidence for an effect of dairying on DRP.

DISCUSSION

At the regional scale, our results suggest that there was an increase in oxidised N and DRP and also decreases in ammoniacal N and % DO in Southland in the period 1995–2001. This period coincided with large-scale intensification of dairying in the region. However, similar trends were also seen in both dairying and in non-dairying sites for the N species and for dissolved oxygen—but not for phosphorus. The decreasing regional trend in ammoniacal N may be related to the speeding up of nitrification as the region's water temperatures rose in that period—an analysis of the 77-site National Rivers Water Quality network over the period 1989–2001 has demonstrated an association between the El Niño Southern Oscillation Index and water quality variables (Scarsbrook et al. unpubl. data). These findings need to be tempered by noting that our data underestimate the total number of stock units associated with dairy farms by not accounting for drystock and calves. Also, and perhaps more importantly, information was only available on the location of stock during the milking season. It is common for Southland dairy farmers to shift stock to run-off blocks during the winter, so some catchments may have different patterns of cow numbers during winter. Run-off blocks are particularly common in northern Southland, despite there being relatively few dairy farms in this area.

Some statistical issues are now worth stating, as they bear heavily on our interpretation of the results and are occasionally the subject of debate (e.g., Death et al. 2000).

Type I error rates

The analysis presented herein makes use of both trend magnitude estimation and hypothesis testing procedures. Regarding the latter there is at times a debate about the appropriateness of using a multitude of P values, as we have done in Table 2 (e.g., Death

et al. 2000; Scarsbrook et al. 2000). This arises when one is considering a global model and asking the question: is there an effect of increased dairying on Southland's water quality? Some argue that this should be addressed by means of a single model, testing a global trend hypothesis for all the water quality variables. In so doing, the overall Type I error rate (i.e., frequency of rejecting a true hypothesis) is controlled to some *a priori* value (e.g., below 5%, as used for each water quality variable herein). In contrast, by examining many P values arising from individual tests on each variable too many of the results could become "significant" merely by chance if the tested hypothesis (positing no trend whatsoever) is true—i.e., the overall Type I error rate is inflated. However, we are discomfited by the global modelling approach for two reasons. First, we do not believe the tested hypothesis (positing that there is no regional trend whatsoever) to be tenable—to reject it is merely to confirm what is already known. To contemplate it being true is to posit that the substantial increase in dairying and other land uses might have no effect whatsoever, down to the n^{th} decimal place. Because this is untenable, some at least of the "significant" results we see reflect substantial departures from that null hypothesis. Second, it may be claimed (e.g., Death et al. 2000) that one should consider all the other water quality variables measured and adjust significance levels accordingly in analyses for any one of them, so as to control the overall Type I error rate. In so doing there comes a point at which a severe penalty is paid for having gone to the trouble to measure other water quality variables. Accordingly, we have not done this. To do so it would be appropriate to reduce all the water quality variables to a single index and use some form of redundancy analysis (Legendre & Legendre 1998). Instead we make separate probability (P value) statements about each water quality variable.

Nevertheless, we have not used the binomial procedure employed by Smith et al. (1996) in whose study the test was only applied to results where a trend test (the seasonal Kendall test) found the trend to be statistically significant (at the 10% level). In part this was because of the above-noted concern about inflation of Type I error rates. But we were also guided by the desire to obtain good statistical power in the binomial test, by using all the Sen slopes calculated, not just those adjudged to be "significant". In effect we are using those calculated slopes, however large or small, as a random sample of the true regional trend.

Evidential content of P values

If it is accepted that a null hypothesis, positing an exactly zero slope, could ever be true one may then ask what evidence its test result (P) conveys, especially in the light of an increasing literature on the topic showing that such P values do not constitute absolute evidence for or against any meaningful hypothesis (e.g., Royall 1997; Goodman 1999; Johnson 1999; McBride 2002). In particular they have the tendency to become ever smaller (i.e., statistical power increases) as the number of samples increases. This is because the hypothesis tested is actually untenable—there will always be some slope present, although it may be very small (consequently, P values do not behave in this manner for tests of interval hypotheses or of one-sided hypotheses, McBride 1999, because the hypothesis they test may actually be true).

But such P values do serve as a **relative** measure of support for the strength of trends in water quality variables, and this is the manner of our interpretation of them. Note that the statistical power of each test within each of our three site categories (dairying, non-dairying, all sites) is comparable, because our P values are all calculated from the same number of samples (or almost the same, if there is a missing datum). However, a difficulty arises when making comparisons across categories because of the different number of sites available ($n = 10, 19,$ and 29). For example, in the flow-adjusted non-dairying NH_4 nitrogen data we have seven out of nine non-zero slopes in the downward direction, giving $P = 0.090$. Among all the flow-adjusted NH_4 data there are 20 out of 28 non-zero slopes in the downward direction, giving a much lower P value (0.018), yet the proportions of downward slopes in these two groups are almost the same. This demonstrates that similar proportions of upward (or downward) trends in data sets of different sizes will result in divergent P values. Accordingly, if there were in fact a similar pattern of upward and downward trends between the dairying and non-dairying sites we would expect P for the former to be smaller than that for the latter. We never know whether such a pattern would exist of course, but this simple example demonstrates the difficulties in comparing P values with data sets of rather different size. Accordingly, the difference in flow-adjusted P values for DRP between dairying and non-dairying sites (Table 2) is even more dramatic than their face value would imply, whereas the values for NO_x may be quite comparable. Our interpretation has been made in the light of these issues.

CONCLUSION

With the provisos discussed above, and the limited data for non-dairying sites, the analysis presented here suggests that increased dairy farming has been associated with increasing DRP in Southland. There has been a worsening in other water quality variables (oxidised N, DO) but these also occurred in non-dairying catchments. A regional improvement in observed water clarity was related to reductions in river flows—it essentially disappeared after flow adjustment. Continued monthly water quality monitoring is desirable to watch for emerging trends.

ACKNOWLEDGMENTS

Steve Ledington (Southland Regional Council) performed the bulk of the trend calculations. NIWA colleagues Bob Wilcock, Mike Scarsbrook, Rob Davies-Colley, and Bryce Cooper reviewed drafts of this material, as did Dr Jim Barnett (Fonterra Research). Data on cow numbers were supplied by Fonterra, Kiwi and Southern Fresh Milk Company. Mark Weatherhead (NIWA) derived catchment boundaries. This study was funded jointly by Southland Regional Council and the Foundation for Research, Science & Technology (contract C01X0018, "Improved statistical methods for environmental issues").

REFERENCES

- Davies-Colley, R. J. 1988: Measuring water clarity with a black disk. *Limnology and Oceanography* 33: 616–623.
- Death, R. G.; Henderson, I. M.; Joy, M. K. 2000: Letter to the Editor. New Zealand's National River Quality Network Comment: problems with Type I errors in water quality analysis. *New Zealand Journal of Marine and Freshwater Research* 34: 745–746.
- Gilbert, R. O. 1987: Statistical methods for environmental pollution monitoring. New York, Van Nostrand Reinhold. 320 p.
- Goodman, S. N. 1999: Towards evidence-based medical statistics. 1: The p -value fallacy. *Annals of Internal Medicine* 130: 995–1004.
- Hamill, K. D. 2002: Design and procedures of surface water quality monitoring programmes at Environment Southland. *Southland Regional Council Publication No. 2002–5*. 33 p.
- Harding, J. S.; Young, R. G.; Hayes, J. W.; Shearer K. A.; Stark J. D. 1999: Changes in agricultural intensity and river health along a river continuum. *Freshwater Biology* 42: 345–357.

- Helsel, D. R.; Hirsch, R. M. 1992: Statistical methods in water resources. Amsterdam, Elsevier. 522 p.
- IDT1998: WQSTAT Plus Users Guide. Intelligent Decision Technologies Limited. Longmont, Colorado.
- Johnson, D. H. 1999: The insignificance of statistical significance testing. *Journal of Wildlife Management* 63(3): 763–772.
- Legendre, P.; Legendre, L. 1998: Numerical Ecology. 2nd ed. Amsterdam, Elsevier. 853 p.
- McBride, G. B. 1999: Equivalence tests can enhance environmental science and management. *Australian & New Zealand Journal of Statistics* 41(1): 19–29.
- McBride, G. B. 2002: Statistical methods helping and hindering environmental science and management. *Journal of Agricultural, Biological and Environmental Statistics* 7(3): 300–305.
- Parkyn, S.; Matheson, F.; Cooke, J.; Quinn, J. 2002: Review of the environmental effects of agriculture on freshwaters. NIWA Client Report FGC2206, April. 43 p.
- Royall, R. M. 1997: Statistical evidence: a likelihood paradigm. London, Chapman and Hall. 191 p.
- Scarsbrook, M. R.; Quinn, J. M.; Boothroyd, I. K. G. 2000: Letter to the Editor. Reply to letter by Death and others. A balanced approach to interpretation of data. *New Zealand Journal of Marine and Freshwater Research* 34: 745–749.
- Smith, C. M.; Wilcock, R. J.; Vant, W. N.; Smith, D. G.; Cooper, A. B. 1993: Towards sustainable agriculture: freshwater quality in New Zealand and the influence of agriculture. Ministry of Agriculture and Fisheries Technical Policy Paper No. 93/10, Wellington.
- Smith, D. G.; Maasdam, R. 1994: New Zealand's National River Water Quality Monitoring Network 1. Design and physico-chemical characterisation. *New Zealand Journal of Marine and Freshwater Research* 28: 19–35.
- Smith, D. G.; McBride G. B.; Bryers G. G.; Wisse J.; Mink D. F. J. 1996: Trends in New Zealand's national river water quality network. *New Zealand Journal of Marine and Freshwater Research* 30: 485–500.
- Snelder, T.; Biggs, B. 2002: Multi-scale river environment classification for water resource management. *Journal of the American Water Resources Association* 38: 1225–1239.
- Stansfield, B. 2001: Effects of sampling frequency and laboratory detection limits on the determination of time series water quality trends. *New Zealand Journal of Marine and Freshwater Research* 35: 1071–1075.