AGRICULTURE RESEARCH GROUP ON SUSTAINABILITY

Research Report: Number 13/02

ISSN 2324-5751 (Print)
ISSN 2324-5700 (Online)

The New Zealand Sustainability Dashboard

Evaluating the ARGOS soil monitoring scheme on kiwifruit orchards

Adrian Monks and Catriona J. MacLeod

Landcare Research

March 2013

Acknowledgements

This work was funded by the Ministry of Business, Innovation and Employment (Contract Number AGRB1201).

Author Affiliations:

Landcare Research, Private Bag 1930, Dunedin 9054

Suggested citation for this report: Monks A, MacLeod CJ 2013. Evaluating the ARGOS soil monitoring scheme on kiwifruit orchards. ARGOS, The Sustainable Dashboard Contract Report 13/02. 48 p. www.argos.org.nz

Reviewed by:	*Approved for release by:*
Deb Wilson Scientist Landcare Research	Chris Phillips Portfolio Leader Realising Land's Potential
Landcare Research Contract Report:	LC 1238

Disclaimer

This report has been prepared by Landcare Research for Zespri International Ltd as part of the ARGOS project. If used by other parties, no warranty or representation is given as to its accuracy and no liability is accepted for loss or damage arising directly or indirectly from reliance on the information in it.

Executive Summary

Value of environmental monitoring: Environmental monitoring is valuable for documenting responses to land use change, engaging public awareness in environmental issues, and providing the necessary evidential basis for market access. In New Zealand, there is very little information available on the long-term trends for environmental indicators associated with farmland at either regional or national scales. This makes it difficult to identify which land practices are sustainable and which species or environmental measures could be used as sustainability indicators.

Soil monitoring in kiwifruit: Soil monitoring was initiated by the Agricultural Research Group on Sustainability (ARGOS) in 2004. Soil quality was selected for monitoring because it is fundamental for sustaining production and livelihoods as well as maintaining diverse and abundant ecological communities on orchards. The long-term goal of our study is to identify the optimal survey design for monitoring long-term trends in soil quality and health in the kiwifruit sector.

Research objective: In this report, we evaluate whether the existing ARGOS soil survey design will be able to detect detrimental changes in soil quality indicators at the industry level for New Zealand's kiwifruit orchards in the future, or if alternative designs perform better. Meaningful threshold trend values and timelines for raising a 'red-alert' alarm if soil quality was declining at the industry level were determined from the literature. Our analysis focused on six soil quality indicators (available mineralisable nitrogen, bulk density, carbon, nitrogen, Olsen phosphorus, and pH). ARGOS data were used to simulate red-alert trends in soil quality in the kiwifruit sector over a 25-year period.

Evaluating the power of the ARGOS design to detect a red-alert trend: A power analysis was used to test the likelihood of detecting a 'red-alert' trend at the industry level, given that the change occurred. We assumed field surveys were carried out following the existing ARGOS survey design, but explored the effect of varying the number and interval between sampling events (2, 5 and 10 years). We set the extreme bounds on the power of the design using different variance scenarios estimated from the data. We assumed (1) the variation around the red-alert trend (i.e. change in the trend through time) was either small or large and (2) the trend was either consistent or variable among orchards.

Power of the ARGOS survey design depends on the red-alert trend characteristics: For all the target indicators, the likelihood (or power) of detecting the simulated red-alert trend was:

- High (95%) when variation around the red-alert trend was assumed to be small irrespective of whether the trend was consistent or variable among orchards. Assuming that sampling occurred every 2 years under these scenarios, there was a 95% chance that the red-alert would be detected within 4 years, or two sample events, for all variables except Olsen P, where three or five sample events were required for a consistent or variable trend among orchards respectively.

- Considerably reduced for most soil indicators when the trend was allowed to vary among the orchards and the variation around the simulated trend was large; if sampling occurred every 2 years under this scenario, the current ARGOS survey design would be able to accurately detect a red-alert trend within 10–25 years for all variables, except available mineralisable nitrogen and pH, where it should be feasible within a 4-year and 10-year period respectively.

- Lowest (<70%) for all indicators when the trend was consistent among orchards but the variance around it was large, with the ARGOS survey design unable to accurately detect a red-alert trend within a 25-year period, even if sampling occurred every 2 years.

With more data, our analysis could be refined to determine if the observed trend for a soil indicator is indeed real, rather than an artefact of the sampling times in relation to some unknown cycle in the nutrients. This would allow us to determine which of the four variance scenarios considered is the most realistic.

Evaluating alternative monitoring designs: The effect of varying the total sampling effort employed per sampling event was also examined for two monitoring design scenarios where: (1) the ARGOS design was maintained but the total sampling effort employed per sampling event was either equivalent to, two-thirds, or one-half of the current ARGOS design; and (2) the ARGOS design was implemented but a new (unique) set of randomly-selected orchards were sampled at each sampling event; the total sampling effort employed per sampling event was either equivalent to, two-thirds, or one-half of the current ARGOS design. For both design scenarios, variance around the trend was considered small but trends varied among orchards, with sampling occurring at 2-yearly intervals.

Trade-offs in survey design evaluation: Assuming the ARGOS design was implemented but the sampling effort was halved, the power to detect red-alert trends was high after 4 years for all indicators, except Olsen P (which had low power even after 25 years of monitoring). When surveying a new set of orchards at each sampling event, the power to detect red-alert trends was initially reduced for all six soil indicators, but then increased (with the time interval varying among the indicators). Overall, these results highlight the value of using repeated measures from the same orchards over time to increase the power of a given monitoring design to detect changes in trends of soil indicators, relative to a design that measures soil quality at a new set of orchards at each sampling event.

Contents

Acknowledgements

The Agricultural Research Group on Sustainability (ARGOS) is primarily funded by the Ministry of Business, Innovation and Employment (Contract No. AGRB0301). ARGOS also acknowledges financial assistance from Industry stakeholders including ZESPRI, Fonterra, Merino New Zealand Inc., COKA (Certified Organic Kiwifruit Growers Association) and in-kind support from Te Rūnanga O Ngāi Tahu.

The ARGOS programme has been designed and implemented with the intention of providing quality information to both farmers and orchardists and their associated industries to ensure that they are broadly sustainable, internationally competitive and profitable. To facilitate this we greatly value the inputs provided by all the participants and industry partners.

We also thank Henrik Moller, Jayson Benge, Peter Carey and Dan Tompkins for helpful discussion; Heather North and Christine Bezar for proof reading the report; and Andrew Gormley, Al Glen, Bruce Warburton, Murray Efford, Deb Wilson and Chris Phillips for their constructive reviews.

ENVIRONMENTAL MONITORING

Value of monitoring

Successful environmental outcomes on production lands require a combination of environmental monitoring, diagnostic research, testing of management solutions, and incorporation of those solutions into sector-wide management schemes[1,2]. In recent years, the value of monitoring environmental responses to land use change[3,4,5,6], engaging public awareness in environmental issues[7] and providing the necessary evidential basis for market access[8] has clearly been demonstrated. For example, on the basis of population monitoring demonstrating farmland bird decline in the United Kingdom[9,10,11], media and public interest was successfully engaged[5], targeted research to understand the mechanisms of decline (and thus identify approaches to reverse decline) was motivated[12,13], and initiatives and incentives put in place to foster the uptake of management solutions by farmers[2].

In New Zealand, where production lands account for 58% of the total area, recent studies have identified an accelerating trend for agricultural intensification[14,15]. However, despite various calls to develop a monitoring scheme that provides reliable biodiversity and environmental indicators of the impact of land use changes on biodiversity or ecosystem services[14,16,17,18,19,20,21], neither the nature of this threat nor the extent of its impact on biodiversity and ecosystem services is known[21,22]. This is because little or no information is available on the long-term trends in environmental indicators associated with farmland habitats at regional or national scales for New Zealand agricultural sectors[22,23,24]. Knowledge of biodiversity and ecosystem services status in farmland areas, or the factors impacting them, is also lacking. Hence, it is currently difficult to identify which land practices are sustainable and which species or environmental measures could be used as sustainability indicators.

New Zealand's kiwifruit sector is the nation's largest horticultural export industry, and a major contributor to the global market[25]. Since 1997, the total export crop has been produced using an integrated pest management system[26,27,28] (KiwiGreen). This system was introduced in the early 1990s to address the international market's concerns about spray residues on fruit. However, because no environmental monitoring occurred on the orchards or the surrounding landscape in parallel with the significant reduction in agrochemical use, the environmental impact of this land use change is unknown. Monitoring is typically expensive and time-consuming. A key challenge, therefore, is to develop robust sampling designs that address specific research or management objectives in a cost-effective manner[1].

Recently, the Agricultural Research Group on Sustainability (ARGOS) demonstrated that orchards managed under an organic system support enhanced biodiversity[29,30,31] and soil quality[32] relative to those managed using integrated systems. There is growing evidence internationally to show that organic systems on farms can enhance biodiversity[33,34,35,36] and have less negative environmental impacts[37,38], and it appears likely that integrated management systems can also contribute to biodiversity conservation. One ARGOS study demonstrated that reducing the frequency and toxicity of pesticide applications within kiwifruit orchards not only addressed consumers' concerns about adverse health impacts of spray residue on fruit in the international market[27], but probably also alleviated adverse impacts on biodiversity[31]. In general the regulations that organic and integrated farms need to meet to secure market access do not necessarily have a strong ecological basis, with biodiversity benefits often being assumed rather than demonstrated[27,34]. The lack of any long-term environmental monitoring in New Zealand's kiwifruit orchards means the industry has had only limited information available for demonstrating to their international market the environmental benefits of reducing use of toxic pesticides.

Designing and evaluating monitoring schemes

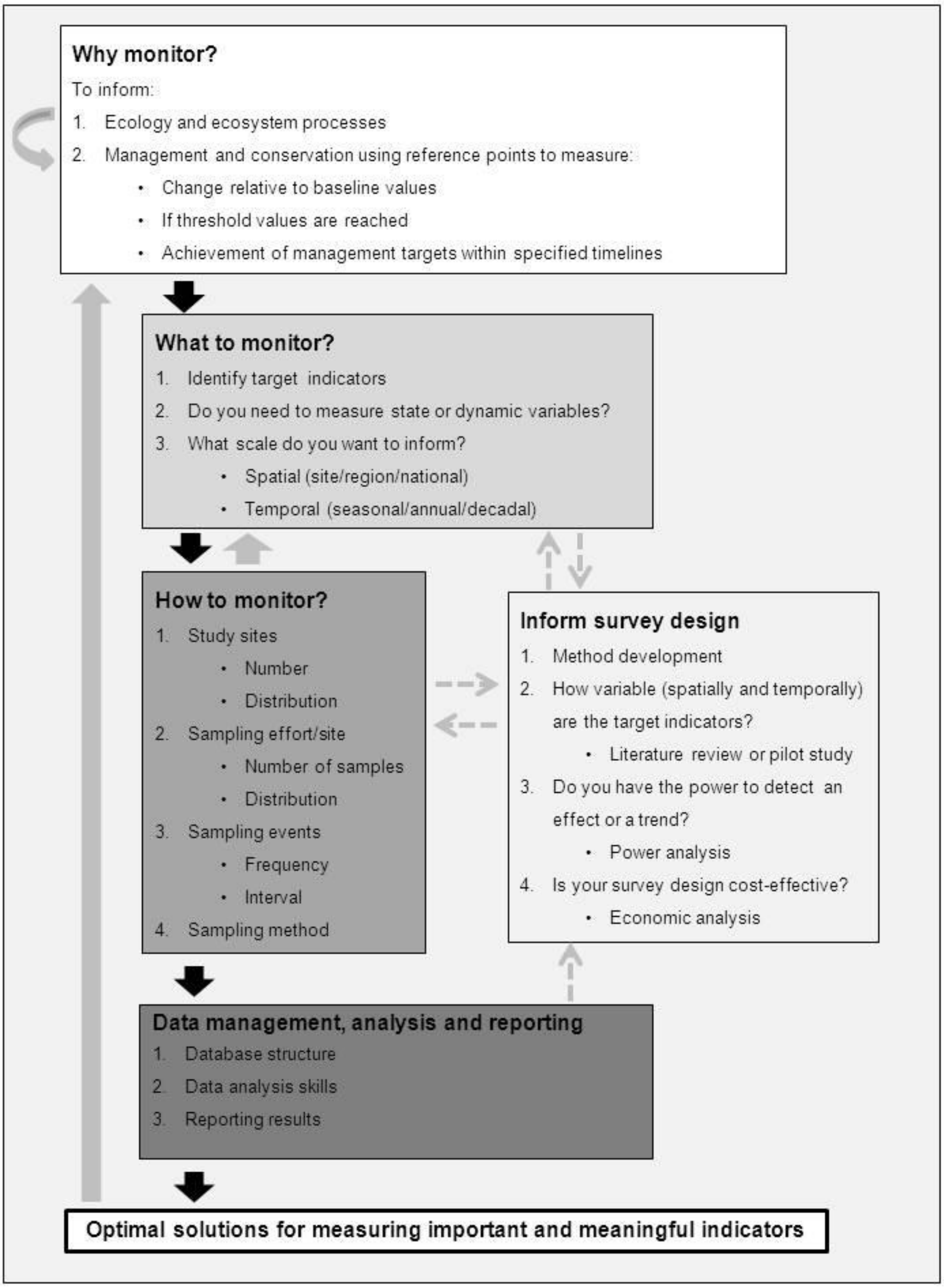
Many existing environmental monitoring schemes suffer from design deficiencies because three basic questions (Box 1) have not been clearly addressed[1,39,40,41]:

1. For what specific purpose are we monitoring?
2. What should we be monitoring to best achieve this purpose?
3. How should we be monitoring to best collect the required data?

Addressing these questions (Why? What? How?) at the design stage prior to the start of monitoring is essential in order to identify the appropriate scale, design and intensity of the monitoring scheme. The extent and strength of inferences drawn from a monitoring scheme depend on the design used[1]. Having selected a particular sampling design, it is important to determine whether that design will have the ability (or power) to detect a specified level of change in the environmental indicator of interest (Boxes 1 and 2). The power of a sampling design depends on:

1. How variable the indicator is over space and time.
2. The sampling effort implemented in the field:
    a. The number and distribution of study sites
    b. The number of sampling locations within sites
    c. The frequency of and interval between sampling events.
3. The magnitude of change that the monitoring scheme aims to measure (given that the change occurred).

Box 1: Key design steps for developing an environmental monitoring scheme

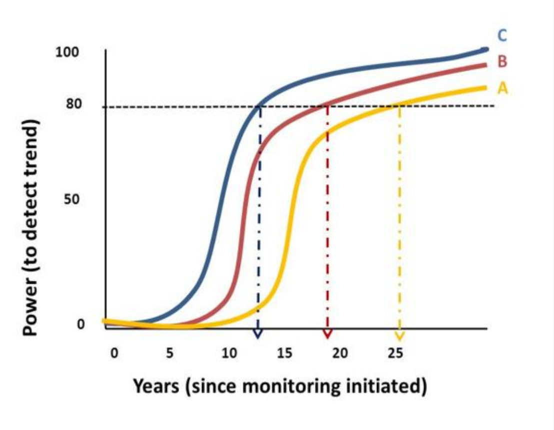


Box 2: What is a power analysis?

A power analysis is a valuable method for determining whether a particular sampling design is likely to detect a specified level of change in an environmental indicator (given that the change occurred). Ideally, this method should be used at the outset of a monitoring scheme; it is not recommended for evaluating why a pre-existing study design failed to detect a change (or effect).

The power to accurately detect a specified level of change can be enhanced (from A to C) by increasing sampling effort through one or more of the following methods:

- Increasing the frequency of sampling events
- Increasing the number of study sites sampled
- Increasing the number of samples collected from each study site



Overall, power to detect a change increases with sampling effort, decreases with increasing variability in the indicator, and increases with the magnitude of change you are aiming to measure.

The strongest inferences are typically made when the measured variables have low bias (little systematic over- or under-estimation) and high precision (a low level of uncertainty)[42,43]. Sources of bias arise when there is non-random selection of sampling units or field data are recorded inaccurately. If precision is low, the scale or intensity of sampling need to be increased. Whether this is most effectively achieved by having more sampling sites or more intensive sampling at existing sites depends on how variable environmental indicators are over space and time. Overall, the power to detect change increases with sampling effort (e.g. Box 3), decreases with increasing variability in the indicator, and increases with the magnitude of change you are aiming to measure. The optimal sampling design will balance the ability of the scheme to detect the specified level of change with the level of investment (financial and time) available for monitoring.

Pilot trials are often essential to identify and confirm the best approach to address the scheme’s objective prior to commencing full-scale monitoring. Important points to consider include:

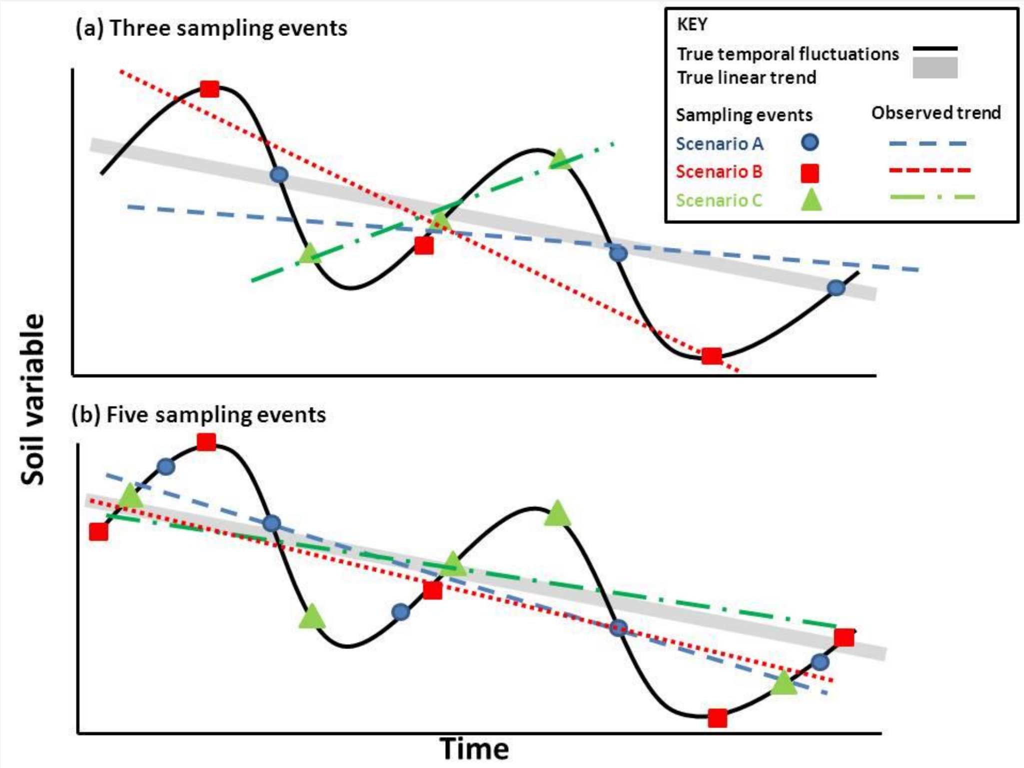
a) whether the survey aims to maximise the quality of the variable estimated or minimise the effort employed[44];
b) what level of difference or change is biologically significant versus statistically significant[45,46]; and
c) what is the target level of significance and power[47].

A sampling design is typically considered robust when it has high power (i.e. ≥80% chance or probability) to accurately detect a specified magnitude of change, but weak when it has low power (i.e. the likelihood of detecting change would be <80%). However, the power thresholds used to determine whether a sampling design is robust or not may be modified depending on the seriousness of erroneously detecting an effect that is not present (false positive or Type I error; e.g. Box 3) versus not detecting an existing effect of given size (false negative or Type II error) and their consequent costs (biological, social or economic). Following the precautionary principle, for example, minimising the risk of missing a species decline will often be desirable for a threatened species (as extinction is irreversible), but when distributing limited conservation resources among species, minimising the risk of accepting effects that are not real can be equally important. Alternatively, it may be desirable to minimise the total cost of the two kinds of error combined[47].

With these points in mind, statistical power analyses can be essential tools for comparing the capacity of different sampling strategies to achieve the purpose for which they are to be applied[48,49,50]. Power analyses can also help identify the optimal balance between resources spent on monitoring versus analysis of results. More-complex analyses generally allow more powerful inference from lower quality data, or from wider but less intensive monitoring strategies. However, there is a limit to the sampling strategy issues that analyses can compensate for, and it is generally better to minimise such issues where resources allow. For example, the NeoTropical Migratory Bird Conservation Program defined an effective monitoring scheme as one that has 90% chance of detecting a 50% decline in a species' abundance over 25 years; however, the uncertainty due to the analysis method employed alone was larger than the absolute change in population size effect that was to be detected[51].

Box 3: Effect of varying the number and timing of sampling events

Here, we consider an example where the overall linear trend in soil quality was a declining one, but there were also large temporal fluctuations in soil quality (e.g. driven by some unknown nutrient cycle). Assuming only three sampling events occurred (figure a), then the likelihood of accurately measuring the underlying linear trend was low and strongly influenced by the timing of sampling events (e.g. sampling scenario B detected a declining trend, while sampling scenario C measured an increasing one). If five sampling events were implemented (figure b), then the likelihood of accurately measuring the true trend was high, irrespective of when the sampling events occurred. Thus, implementing a higher level of sampling effort (in this case five rather three sampling events) more effectively discriminated a real trend from a false one.



ARGOS SOIL MONITORING DESIGN

Why monitor soils?

Agro-ecosystems are both providers and consumers of ecosystem services[52]. Soil structure and fertility, for example, provide essential ecosystem services to agro-ecosystems[53]. While agricultural land can help regulate soil quality, it can also be the source of adverse impacts, e.g. nutrient runoff and sedimentation of waterways[52]. Thus, it is critical that we identify appropriate agricultural management practices for realising the benefits of ecosystem services and reducing disservices from agricultural activities.

Soil quality is not only fundamental to sustaining production and livelihoods[20]; it is also required to maintain diverse and abundant ecological communities on farms. The concept of soil quality includes soil properties and processes that determine the ability of soil to function effectively as an ecosystem component[54]. Soil quality may be broadly defined to include capacities for water retention, carbon sequestration, plant productivity, waste remediation, and other functions. Intensively managed agro-ecosystems are sustainable in the long term only if the outputs of all components produced are balanced by appropriate inputs[55]. Such inputs (e.g. fertiliser) are often costly, make up a significant component of the energy footprint for food production[56] and increase the risk of environmental impacts both on and off the farm (e.g. nutrient runoff).

The ARGOS soil monitoring scheme for New Zealand's kiwifruit sector was initiated in 2004, as part of a broader research programme examining the environmental, social and economic sustainability of New Zealand's farming systems[20]. The scheme was originally designed to address four key objectives:

1. Establish baseline information on soil quality in relation to farming systems and locations as well as other habitats and countries.
2. Determine drivers of variation in soil quality to provide the necessary information required to underpin management and conservation.
3. Identify a subset of soil quality measures that can be used as indicators for monitoring the impact of land use change.
4. See how these soil quality measures can be integrated with economic and social indicators to understand drivers of change.

What to monitor?

The subset of target indicators selected to measure different aspects of soil quality (Table 1) was identified following an extensive literature review[20]. The choice of indicators for monitoring soil quality on kiwifruit orchards was strongly influenced by the need to: (a) cover biological, physical and chemical aspects of soil quality; (b) ensure the indicators, wherever possible, were comparable to historical information for New Zealand soils; and (c) encourage growers and consultants to use low-tech but reliable and meaningful soil quality indicators throughout their operations.

Having identified the target indicators, the next priority was to identify the types of variables that need to be measured (Box 1). A 'snapshot' approach was used to assess soil quality in relation to different management systems and locations as well as other habitats and countries (e.g. Objectives 1 and 2). This involved quantifying the current 'state' of soils at certain time-points that were comparable (e.g. mean indicator values among management systems[32,57]). In contrast, research aiming to monitor the impact of land use change (Objective 3) will require measurement of dynamic variables that quantify temporal changes (trend) in soil quality variables. Integrating soil quality measures with related economic and environmental indicators to understand drivers of change (Objective 4) will require a combination of state and dynamic measures.

The scheme was designed to focus on monitoring the main production areas within each sector to facilitate comparisons of soil quality: (a) among three agricultural sectors (kiwifruit, sheep-beef and dairy); (b) among the two or three management systems monitored within each sector; and (c) over time[20]. Inferences drawn from the monitoring scheme are, therefore, limited to the production areas of the orchard. Another reason for focusing on the dominant landforms within each sector was to reduce the risk of spatial variance in soil quality masking any temporal trends, thus reducing the scheme's statistical power to detect long-term trends in soil quality.

How to monitor?

A nested sampling design (Box 4) was used to monitor soils on 36 kiwifruit orchards, consistent with the overarching ARGOS study design[20]. The orchards were grouped into 12 geographic clusters of three orchards each. Within each cluster were three different management systems: an 'organic' system growing the 'Green' kiwifruit variety (Hayward, *Actinidia deliciosa*), an integrated management system with the same 'Green' kiwifruit variety, and an integrated management system growing the 'Gold' variety (Hort 16A, *A. chinensis*). Clusters were distributed to cover the study area in a stratified random design[58] (i.e. to maximise geographical spread for high-level inference while retaining the ability to compare between farm systems by having one farm of each sort within each cluster). The orchard was the primary study unit because it is the key 'site of action', i.e. for decision-making by owners[59]. On each orchard, three kiwifruit blocks (management units) were randomly selected and three permanent soil monitoring sites (SMS) were randomly placed within each block. Sampling occurred at two locations within each SMS: within and between vine rows (the dominant landforms). Sampling occurred on three occasions at intervals of 2–3 years in the winter (before fertiliser was applied). A suite of chemical, biological and physical measures (Table 1) in the top 15 cm of soil were done at each SMS, using a combination of field and laboratory techniques.

Box 4: ARGOS nested sampling design for monitoring soil quality[20,58,59]

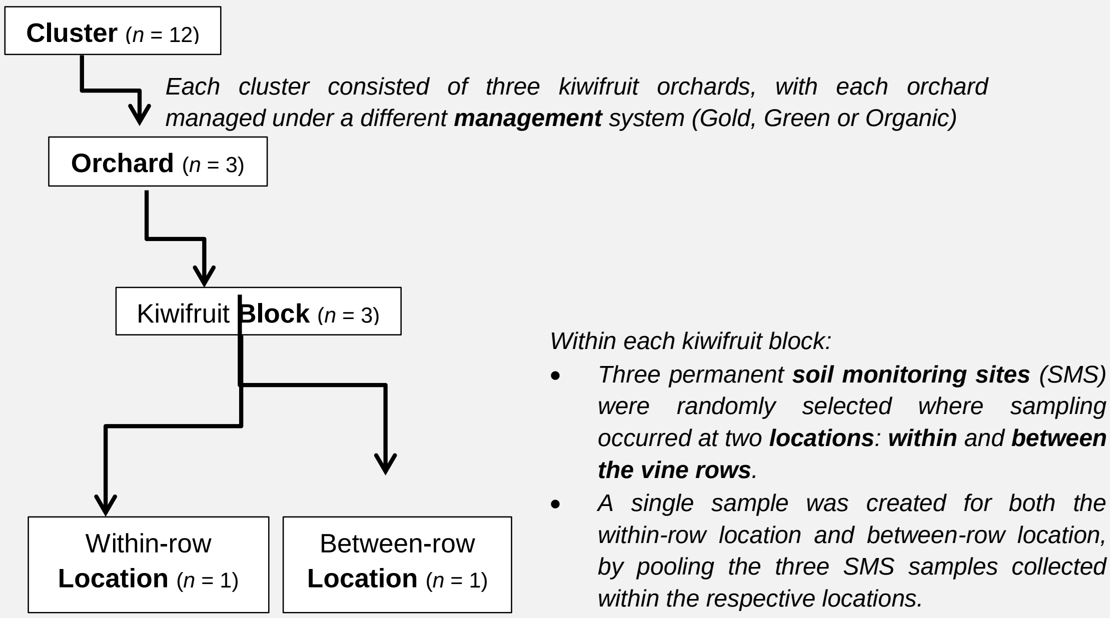


Table 1: Summary of indicators used to assess different aspects of soil quality in kiwifruit orchards

Aspect	Indicator	Code	Units	Value of measure [20]	Survey
Structure	Bulk density	Total BS %	g cm^{-3}	Soil compaction, physical environment for roots and soil organisms	Ongoing [32,57,60]
	Aggregation		categorical	Soil compaction, physical environment for roots and soil organisms	Ongoing [32,57,60]
	Porosity		categorical	Soil compaction, physical environment for roots and soil organisms	Ongoing [32,57,60]
Chemistry	pH	pH	pH	Acidity or alkalinity of soil (influences availability of nutrients)	Ongoing [32,57,60]
	Cation exchange capacity	CEC	$cmol^{+}$ kg^{-1}	Capacity of soil to hold cations	Ongoing [32,57,60]
	Exchangeable calcium	Ca	$cmol^{+}$ kg^{-1}	Major nutrient for plant growth	Ongoing [32,57,60]
	Exchangeable magnesium	Mg	$cmol^{+}$ kg^{-1}	Major nutrient for plant growth	Ongoing [32,57,60]
	Exchangeable potassium	K	$cmol^{+}$ kg^{-1}	Major nutrient for plant growth	Ongoing [32,57,60]
	Phosphorus retention	P retention	ASC%	Amount of clay minerals present that immobilise phosphorus	Ongoing [32,57,60]
	Olsen phosphorus	Olsen P	µg L^{-1}	Phosphorus readily available	Ongoing [32,57,60]
	Resin phosphorus	Resin P	µg g^{-1}		Ongoing [32,57,60]
	Sulphate-sulphur	Sulphate-S	µg g^{-1}		Ongoing [32,57,60]
	Organic sulphur	Organic S	µg g^{-1}		Ongoing [32,57,60]
	Total nitrogen	N%	%	Total soil nitrogen	Ongoing [32,57,60]
	Potentially mineralisable nitrogen	AMN	µg g^{-1}	Surrogate measure for soil microbial biomass	Ongoing [32,57,60]
	Total carbon	C%	%	Organic matter content	Ongoing [32,57,60]
	Ratio carbon to nitrogen	C/N	Ratio	Relative measure of soil fertility	Ongoing [32,57,60]
Biology	Microbial N		µg g^{-1}	Measure of total amount of N present in living microbes in soil	Ongoing [32,57,60]
	Microbial C		µg g^{-1}	Measure of total amount of C present in living microbes in the soil	Ongoing [32,57,60]
	Basal respiration			Measure of soil microbial activity	Ongoing [32,57,60]
	Earthworm abundance		No. m^{-2}	Incorporate and break down organic matter to make nutrients available. Improve soil structure.	Ongoing [32,57,60]
	Nematode total		No. m^{-2}		Case study [61,62,63]
	Nematode groups		No. m^{-2}		Case study [61–63]

EVALUATING THE ARGOS SOIL MONITORING DESIGN

This report represents an initial step towards identifying the optimal survey design for monitoring long-term trends in soil quality and health on New Zealand's kiwifruit orchards. More specifically, it evaluates whether the existing ARGOS soil survey design (Box 4) will be able to detect detrimental changes in soil quality indicators at the industry level in the future.

First, meaningful reference points and timelines (Box 1) that could be used to raise a 'red-alert' alarm if soil quality was declining at the industry level were identified from the literature. ARGOS data were then used to simulate red-alert trends in soil quality in the kiwifruit sector. Finally, a power analysis (Box 2[64]) was used to test the likelihood of detecting the 'red-alert' trend at the industry level in the future. The effect of varying the interval between sampling events was tested (2, 5 and 10 years).

Raising the 'red-alert' alarm: identifying meaningful thresholds

The scale, design and extent of monitoring required will depend on the specific purpose of the monitoring scheme. Thus, the scheme's objectives need to be clearly defined from the outset (see 'Why monitor?' in Box 1). Here, we assume that the primary purpose of the ARGOS soil monitoring scheme is to warn the kiwifruit sector that 'red-alert' trends in soil quality are occurring at the industry level. To address this objective, six target indicators considered most relevant and sensitive for monitoring soil quality in the kiwifruit sector were selected. Red-alert trend values were then calculated for each target indicator (Box 5[65]).

Box 5: Calculating red-alert trend values for target indicators

Target indicators were used to evaluate the ARGOS survey design's ability (power) to detect 'red-alert' trends in soil quality. These target indicators were selected in consultation with ARGOS soil experts (J. Benge, The AgriBusiness Group; P. Carey, Land Research Services) (see Table 1 for a complete list and detailed descriptions of ARGOS soil indicators).

Target indicator	Baseline values (μ_i)	Red-alert thresholds (R_i)	Red-alert trends (β_{Ri})
AMN*	70	20	−2.00
Bulk density	0.77	1.4	0.025
Carbon	5.66	2.5	−0.13
Nitrogen	0.48	0.7	0.0088
Olsen phosphorus	58	100	1.68
pH	6.5	5	−0.06

*Available mineralisable nitrogen

For each target indicator:

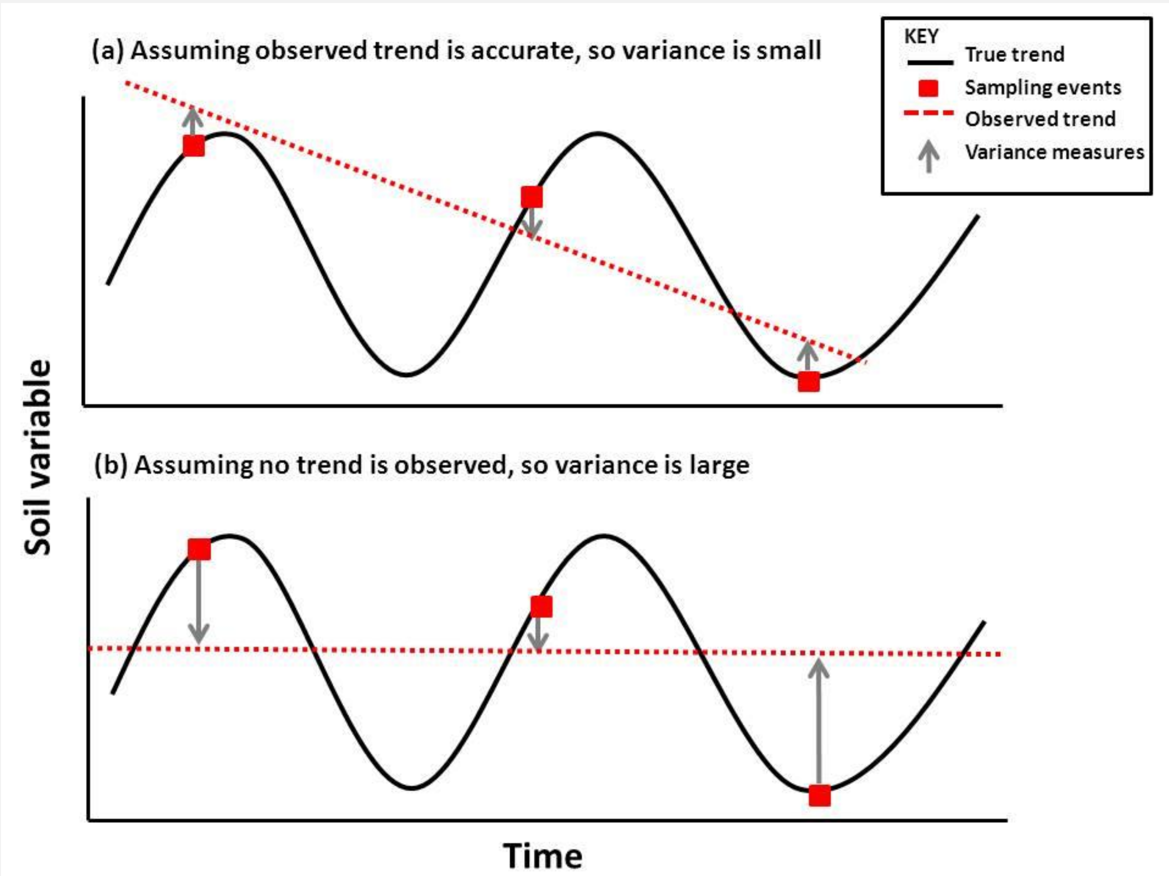
- **Baseline values** (μ_i) were mean parameter values for existing ARGOS data[57] (collected during three surveys over a 6-year period).
- **Red-alert thresholds** (R_i) were defined using soil quality targets from the literature[65].
- **Red-alert trends** (β_{Ri}) were assumed to be equal to the indicator's baseline value (μ_i) becoming a 'red-alert' threshold value (R_i) over a 25-year period. The red-alert trend was the annual change, assuming a linear trend, given by the equation: $\beta_{Ri} = (R_i - \mu_i)/25$.

Measuring power to detect a red-alert trend

We set the extreme bounds on the power of the design using different variance scenarios estimated from the data. We assumed (1) the common variation around the red-alert trend (i.e. variation between years, across all orchards) was either small or large and (2) the trend was either consistent or variable among orchards. To estimate the common variation around the red-alert trend we fitted two variance-estimation models to the ARGOS soil monitoring data. In the first model (trend model) we fitted a linear trend to the data over time to account for temporal variation in the soil nutrients. The resulting unexplained variation was used as our variance estimate for the small-variance scenario. However, the ARGOS soil monitoring scheme only completed three surveys over a 6-year period and it is possible that the observed trend could be an artefact of the timing of our sampling events (see Box 6). Therefore we also considered a variance scenario where the observed variation over time was random fluctuation around a constant mean rather than a real trend (no-trend model). The estimate of this random fluctuation was used in the simulations for our large-variance scenario.

Box 6: Measuring temporal variance around a linear trend

Here, we consider an example where there were large temporal fluctuations in soil quality (e.g. driven by some unknown nutrient cycle) but the overall linear trend in soil quality was unchanged (see 'true trend' in figure). Assuming only three sampling events occurred, simply by chance we could detect a false declining trend ('see observed trend' in figure a), where the unexplained variance measured is small (see arrows in figure a). With more sampling events we could be more certain that any trend that we observed was real (see Box 3). A more conservative approach is to assume that there is no overall trend (see 'observed trend' in figure b) and that any variation around an assumed constant mean value of the soil variable is random. The estimate of the unexplained variance is large (see arrows in figure b). If there is no trend (as in figure a), then fitting a trend underestimates the true random variation in the data. If the observed trend is real, then the second scenario (figure b) overestimates the random variation in the data.



We also considered scenarios in which the trend was either consistent or variable among orchards so that we could explore the effects on power of variability among orchards in the trends in soil quality indicators, due to factors such as soil type and/or other land management practices or systems on orchards. Estimates of the orchard-to-orchard variation in the trend were obtained by fitting a variance-estimation model to the ARGOS soil monitoring data that included an overall trend over time (the small-variation scenario above), but also allowing that trend to vary at random between orchards. The variance in the observed orchard-to-orchard trends was used for our simulations of orchard-to-orchard variation in the red-alert trend. We used a Bayesian modelling approach to quantify these spatial and temporal measures of variance in the kiwifruit sector (Appendices 1 and 2).

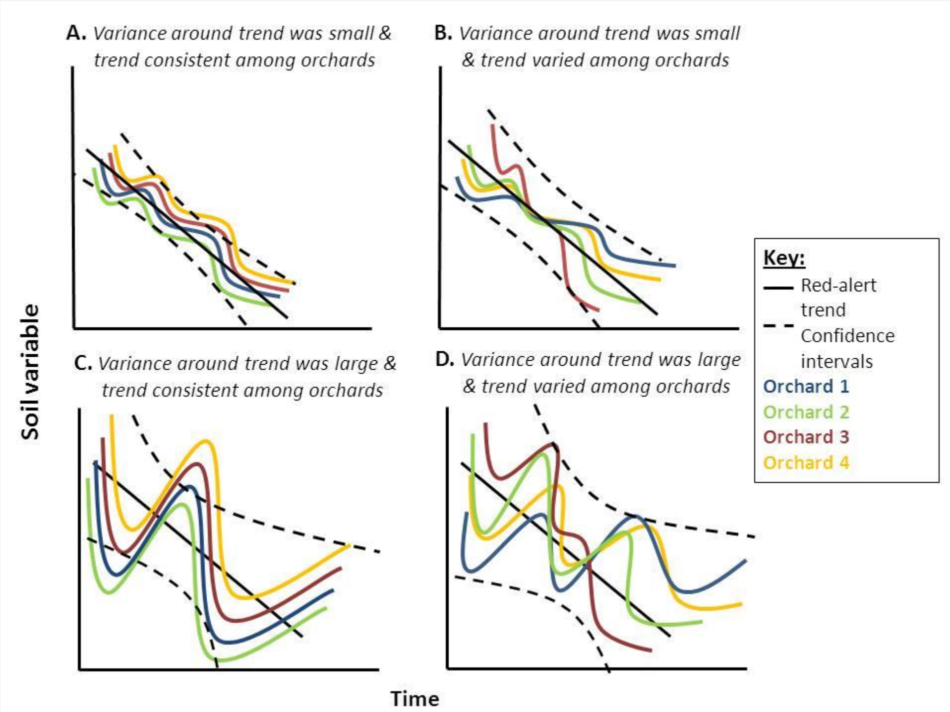
Red-alert trends in soil quality (at the industry level) were simulated for each of the six focal indicators in the kiwifruit sector over a 25-year period under four different scenarios (Box 7) each defined by one of the following assumptions:

A. Variance (fluctuation) around the red-alert trend was small[i] and the trend was consistent among orchards (i.e. all orchards follow the same trend).
B. Variance around the red-alert trend was small[i] and the trend varied among orchards (i.e. each orchard followed its own trend trajectory but on average they followed the red-alert trend).
C. Variance around the red-alert trend was large[ii] and the trend was consistent among orchards.
D. Variance around the red-alert trend was large[ii] and the trend varied among orchards.

The likelihood of detecting the simulated ‘red-alert’ trends was then tested for each of the target indicators (Box 5; Appendices 1 and 2). Field sampling was assumed to follow the existing ARGOS survey design (i.e. the nested survey design outlined in Box 4). To assess the effect of changing the frequency and timing of sampling events (field surveys), we varied the number of sampling events and the interval between them (2, 5 and 10 years).

[i] This was calculated using the error sample variance from the corresponding ‘trend’ variance-estimation model (see Appendices 1 and 2).
[ii] This was calculated using the error sample variance plus variance due to time (i.e. time was specified as a random effect) from the ‘no trend’ variance-estimation model (see Appendices 1 and 2).

Box 7: Examples of red-alert decline trends simulated assuming four different temporal variance scenarios



Power depends on red-alert trend characteristics – results of our analyses

Assuming the variance around the real red-alert trend for soil indicators is small and the trend consistent or variable among orchards (Scenarios A and B in Box 7), the current ARGOS survey design should be able to detect trends as large as a red-alert accurately 95% of the time (see specifications for individual target indicators in Box 5). Assuming that sampling occurred every 2 years under these scenarios, there was a 95% chance that the red-alert would be detected within 4 years (or two sample events) for all variables except Olsen P, where three or five sample events were required for a consistent or variable trend among orchards, respectively (Table 2; Figs 1–6). Sampling at greater intervals (5 and 10 years) under scenarios A and B produced similar power to sampling at 2-year intervals; however, it took longer to detect the red-alert trend due to the longer sampling intervals.

The power to detect a red-alert trend was considerably reduced for most soil indicators when the trend was allowed to vary among the orchards and the variation around the simulated trend was large (D in Box 7); if sampling occurred every 2 years under this scenario, the current ARGOS survey design would be able accurately detect a red-alert trend within 10–25 years for all variables, except available mineralisable nitrogen and pH, where it should be feasible within a 4-year and 10-year period, respectively (Table 2; Figs 1–6). Power at greater sampling intervals (5 and 10 years) was comparable to sampling at 2-year intervals except under scenario D when it was reduced at the higher intervals for carbon and pH.

Overall, the power was lowest (<70%) for all indicators when the trend was consistent among orchards but the variance around it was large (C in Box 7), with the ARGOS survey design unable to accurately detect a red-alert trend within a 25-year period, even if sampling occurred every 2 years (Table 2; Figs 1–6).

Table 2: Minimum duration (years) of a monitoring scheme (using the current ARGOS survey design) required to be 80% or 60% sure of detecting a decline at least as fast as a red-alert trend (Box 5) using three different sampling intervals (2, 5 or 10 years) for six target soil indicators actually declining at the red-alert rate, assuming four different temporal variance scenarios (Box 7).

Power	Red-alert trend scenario	AMN*			Bulk density			Carbon			Nitrogen			Olsen P			pH		
		2-yr	5-yr	10-yr	2-yr	5-yr	10-yr	2-yr	5-yr	10-yr	2-yr	5-yr	10-yr	2-yr	5-yr	10-yr	2-yr	5-yr	10-yr
80%	A. Stable, consistent among orchards	2	5	10	2	5	10	2	5	10	2	5	10	4	5	10	2	5	10
	B. Stable, variable among orchards	2	5	10	2	5	10	2	5	10	2	5	10	8	10	10	2	5	10
	C. Unstable, consistent among orchards	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	D. Unstable, variable among orchards	4	5	10	24	25	>25	20	>25	>25	24	>25	>25	>25	>25	>25	10	15	20
60%	A. Stable, consistent among orchards	2	5	10	2	5	10	2	5	10	2	5	10	2	5	10	2	5	10
	B. Stable, variable among orchards	2	5	10	2	5	10	2	5	10	2	5	10	4	5	10	2	5	10
	C. Unstable, consistent among orchards	2	5	10	c. 6	10	>25	>25	>25	>25	>25	>25	>25	8	10	10	2	5	10
	D. Unstable, variable among orchards	2	5	10	2	5	10	6	5	10	6	5	10	12	15	20	2	5	10

*Available mineralisable nitrogen

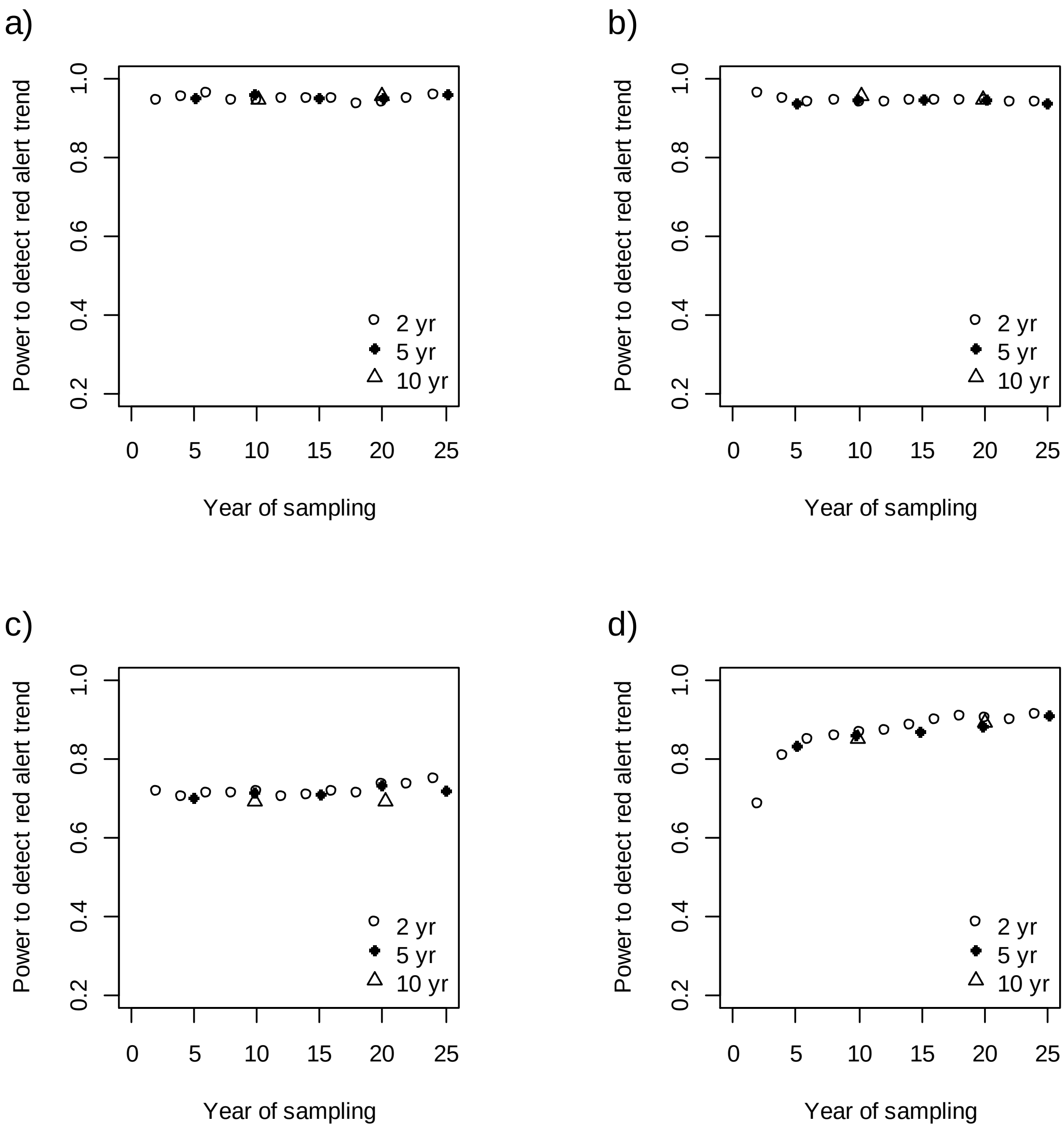


Figure 1: Power for detecting a red-alert trend in available mineralisable nitrogen in relation to the year of sampling and varying the interval between sampling events (2, 5 or 10 years) under four scenarios (Box 7) that assumed either (a) variance around the red-alert trend was small and the trend consistent among orchards, (b) variance around the red-alert trend was small and the trend varied among orchards, (c) variance around the red-alert trend was large and the trend was consistent among orchards, or (d) variance around the red-alert trend was large and the trend varied among orchards.

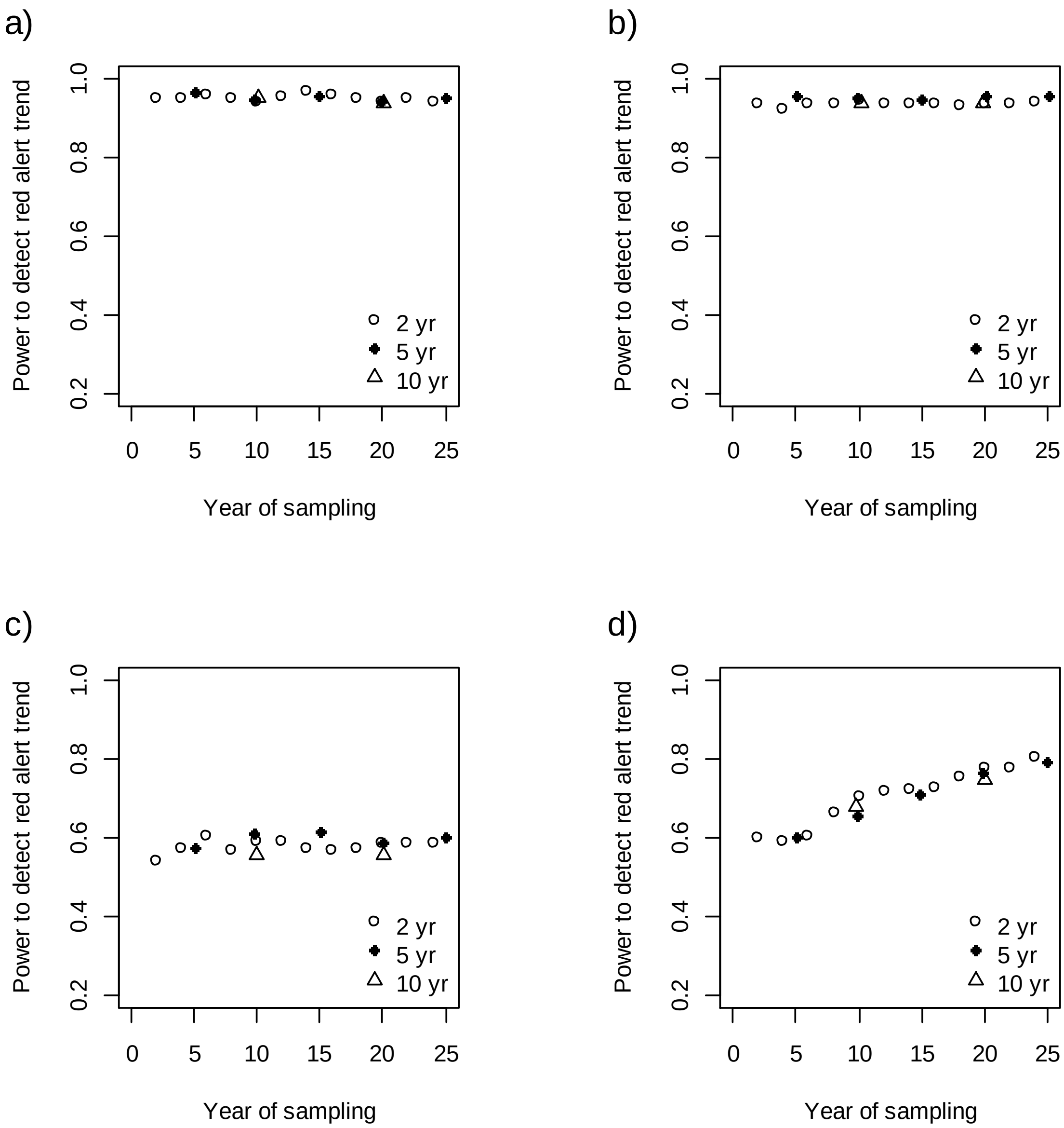


Figure 2: Power for detecting a red-alert trend in bulk density in relation to the year of sampling and varying the interval between sampling events (2, 5 or 10 years) under four scenarios (Box 7) that assumed either (a) variance around the red-alert trend was small and the trend consistent among orchards, (b) variance around the red-alert trend was small and the trend varied among orchards, (c) variance around the red-alert trend was large and the trend was consistent among orchards, or (d) variance around the red-alert trend was large and the trend varied among orchards.

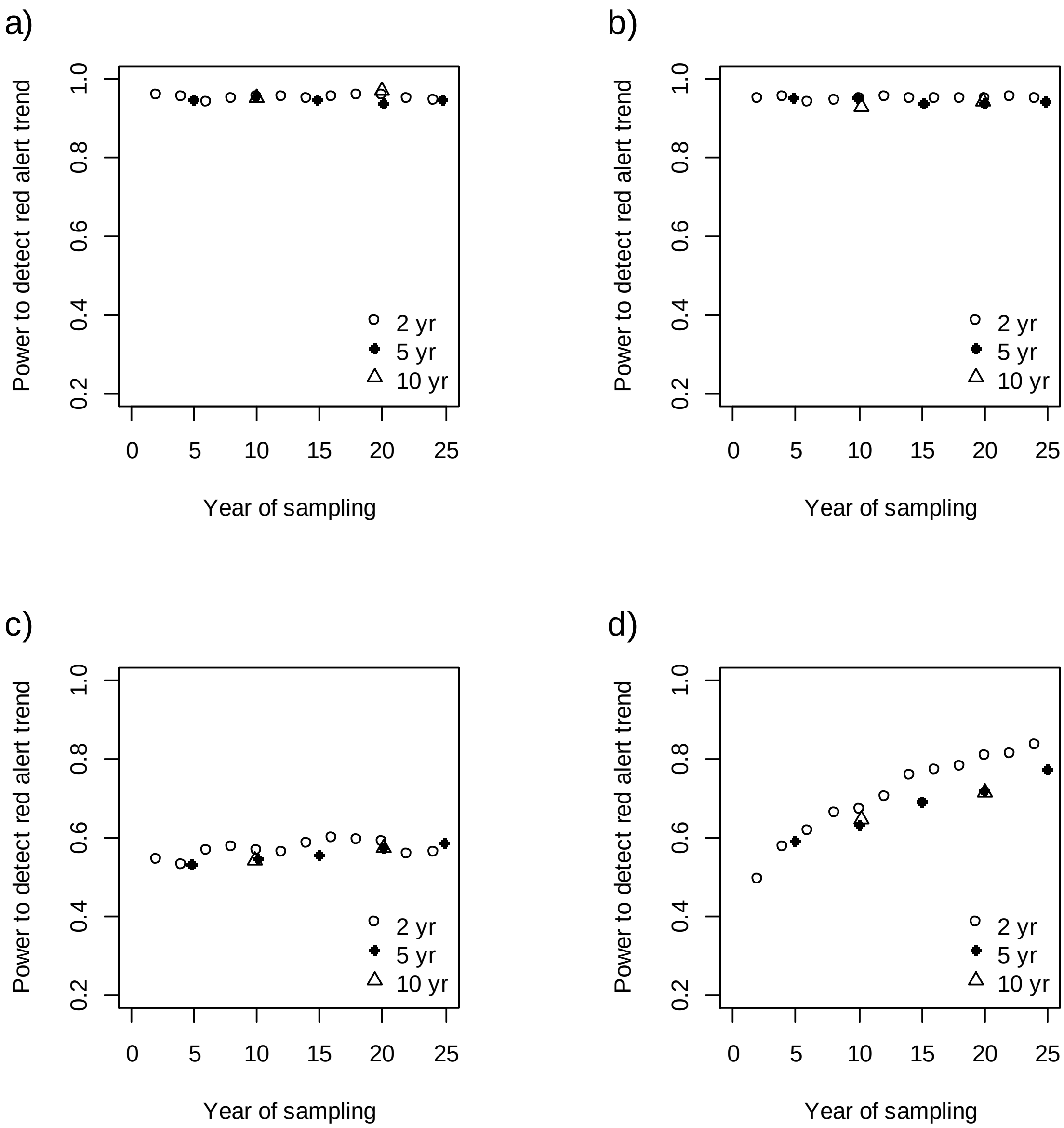


Figure 3: Power for detecting a red-alert trend in carbon in relation to the year of sampling and varying the interval between sampling events (2, 5 or 10 years) under four scenarios (Box 7) that assumed either (a) variance around the red-alert trend was small and the trend consistent among orchards, (b) variance around the red-alert trend was small and the trend varied among orchards, (c) variance around the red-alert trend was large and the trend was consistent among orchards, or (d) variance around the red-alert trend was large and the trend varied among orchards.

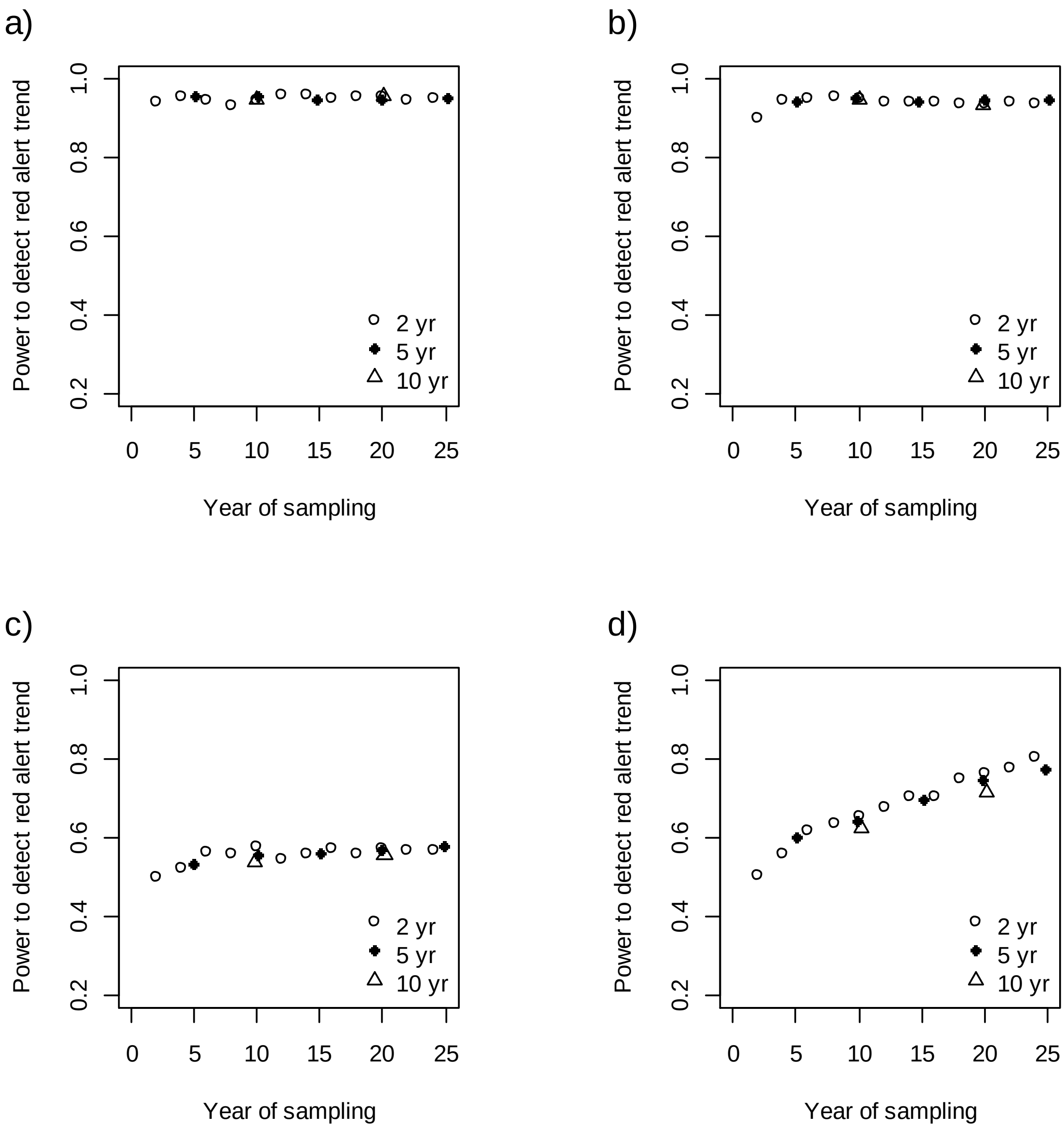


Figure 4: Power for detecting a red-alert trend in nitrogen in relation to the year of sampling and varying the interval between sampling events (2, 5 or 10 years) under four scenarios (Box 7) that assumed either (a) variance around the red-alert trend was small and the trend consistent among orchards, (b) variance around the red-alert trend was small and the trend varied among orchards, (c) variance around the red-alert trend was large and the trend was consistent among orchards, or (d) variance around the red-alert trend was large and the trend varied among orchards.

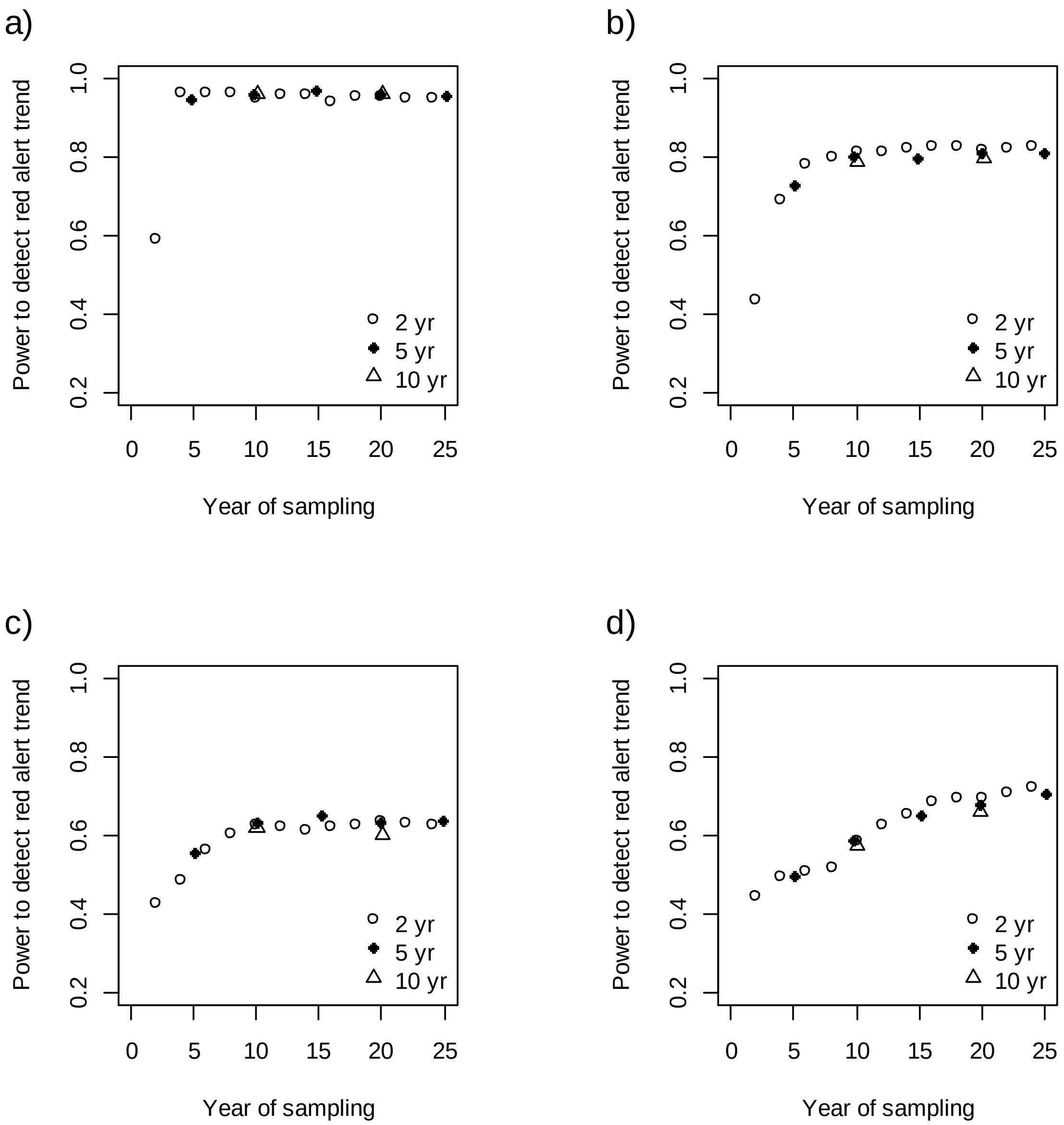


Figure 5: Power for detecting a red-alert trend in Olsen phosphorus in relation to the year of sampling and varying the interval between sampling events (2, 5 or 10 years) under four scenarios (Box 7) that assumed either (a) variance around the red-alert trend was small and the trend consistent among orchards, (b) variance around the red-alert trend was small and the trend varied among orchards, (c) variance around the red-alert trend was large and the trend was consistent among orchards, or (d) variance around the red-alert trend was large and the trend varied among orchards.

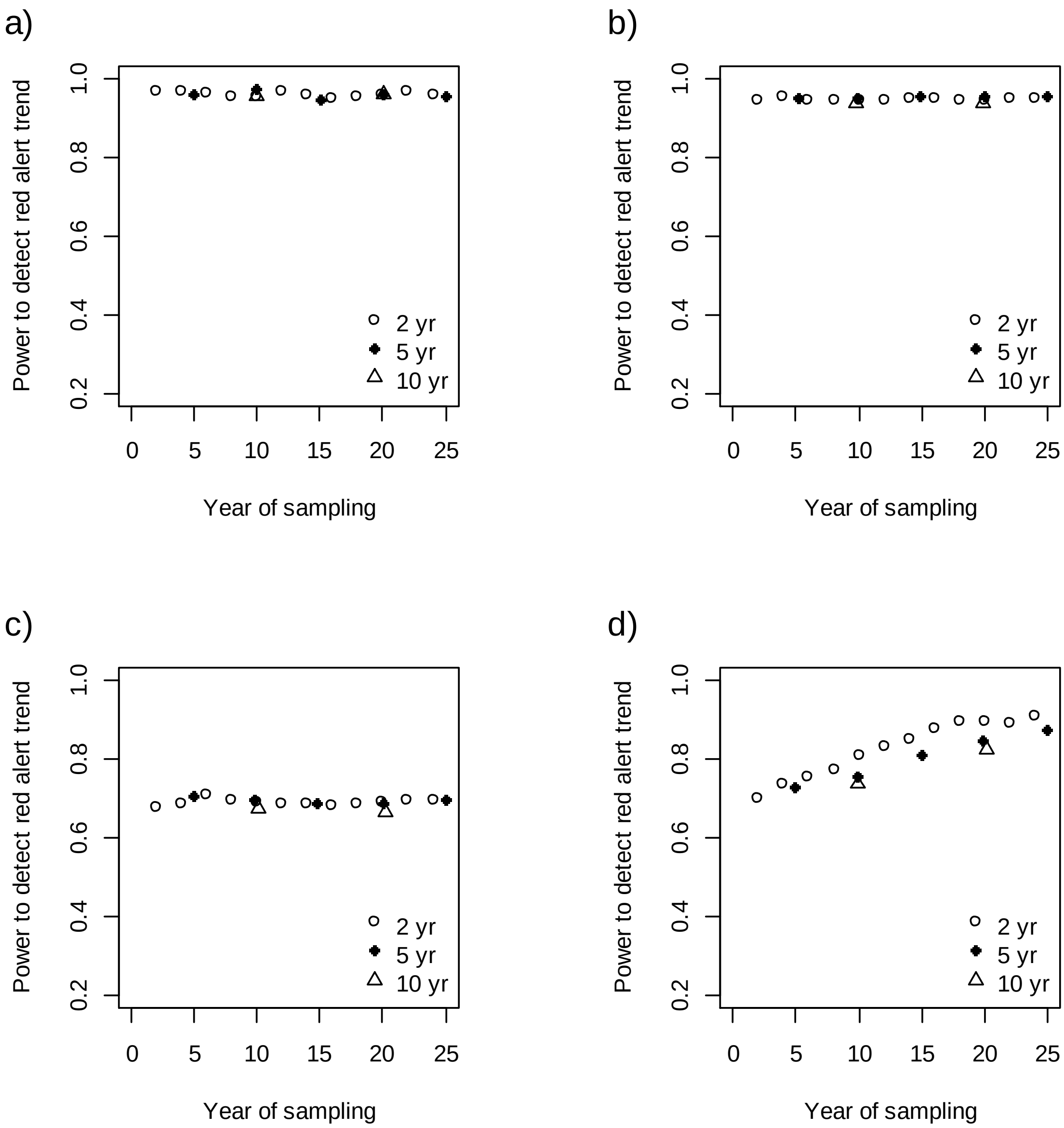


Figure 6: Power for detecting a red-alert trend in pH in relation to the year of sampling and varying the interval between sampling events (2, 5 or 10 years) under four scenarios (Box 7) that assumed either (a) variance around the red-alert trend was small and the trend consistent among orchards, (b) variance around the red-alert trend was small and the trend varied among orchards, (c) variance around the red-alert trend was large and the trend was consistent among orchards, or (d) variance around the red-alert trend was large and the trend varied among orchards.

CONSIDERING ALTERNATIVE MONITORING DESIGNS

Here we investigate the effect of modifying the ARGOS soil monitoring design, with the aim of identifying an optimal design for accurately detecting an industry-level red-alert trend in soil quality, while minimising the risk of a false alarm. More specifically, we examine the effect of varying the total sampling effort employed per sampling event for two monitoring designs (see Tables 3 and 4):

1. **ARGOS design**: assumed the ARGOS design was maintained but the total sampling effort employed per sampling event was either equivalent to, two-thirds of, or one-half of the current ARGOS design.

2. **New-sites design**: assumed the ARGOS design was implemented but a new (unique) set of randomly-selected orchards sampled at each sampling event; the total sampling effort employed per sampling event was either equivalent to, two-thirds of, or one-half of the current ARGOS design.

These monitoring designs were tested using simulated datasets that were derived from the scenario where variance around the trend was considered small but trends varied among orchards (Scenario B in Box 7). Sampling was assumed to occur at 2-yearly intervals.

Trade-offs in alternative survey designs

Assuming the ARGOS design was implemented but the sampling effort was halved (n = 18 orchards), the power to detect the red-alert trends was high (>0.8) for four of the six soil indicators after 2 years and for nitrogen after 4 years (Fig. 7). Acceptable power (>0.8) was never attained over a 25-year sampling period for Olsen P. These results suggest that the existing sampling effort is the minimum effort required to monitor all of the soil variables in order to detect changes in the soil nutrient trends equal to the red-alert levels described in Box 5.

Orchard turnover (new-sites design) strongly decreased the short-term predicted power of the design to detect red-alert trends for all nutrients (Fig. 7). This pattern highlights the gains in statistical power that are made by repeatedly sampling the same orchards when there is variation between orchards in their soil nutrient trends. Note that the levels of variation between orchards used in these simulations is realistic, having been estimated from the field data collected in the ARGOS study, Interestingly, over the longer term, predicted power was as high or higher (after 14 years for Olsen P) when new orchards were sampled at each sampling event. This suggests that the standard ARGOS design does not sample enough orchards to adequately estimate the orchard-to-orchard variation in Olsen P.

Table 3: Schematic specifications[66] for two monitoring designs considered for evaluating soil quality indicators on orchards (see Table 4 for more information).

SAMPLING DESIGN	ORCHARD SET	TIME PERIOD (YEARS) 0	2	4	6	8	10	12	14	16	18	20	22	24
ARGOS DESIGN	1	X	X	X	X	X	X	X	X	X	X	X	X	X
NEW-SITES DESIGN	1	X												
	2		X											
	3			X										
	4				X									
	5					X								
	6						X							
	7							X						
	8								X					
	9									X				
	10										X			
	11											X		
	12												X	
	13													X

Table 4: Alternative monitoring designs considered for measuring red-alert trends in soil quality at the industry level in the kiwifruit sector, where datasets were simulated assuming the variance around the trend was small but trends varied among orchards (Scenario B in Box 7). 'Orchard turnover' indicates the frequency that the set of orchards subject to sampling was changed (see examples in Table 3). 'Sampling effort' was either equivalent to the original ARGOS survey design (*n* = 36 orchards; see Box 3 for detailed description) or two-thirds (*n* = 24 orchards) or half (*n* = 18 orchards) the ARGOS design.

Design	Scenario	Orchard turnover	Sampling effort	80% power to detect red-alert trend						60% power to detect red-alert trend					
				AMN*	Bulk density	Carbon	Nitrogen	Olsen P	pH	AMN	Bulk density	Carbon	Nitrogen	Olsen P	pH
ARGOS	A1	No change	36 orchards	2	2	2	2	8	2	2	2	2	2	4	2
	A2	No change	24 orchards	2	2	2	2	>25	2	2	2	2	2	6	2
	A3	No change	18 orchards	2	2	2	4	>25	2	2	2	2	2	>25	2
New-sites	N1	Each sampling event	36 orchards	8	8	10	12	14	4	6	6	8	10	10	4
	N2	Each sampling event	24 orchards	8	8	12	14	16	4	6	6	10	12	12	4
	N3	Each sampling event	18 orchards	10	10	14	16	20	6	8	8	10	14	14	4

*Available mineralisable nitrogen

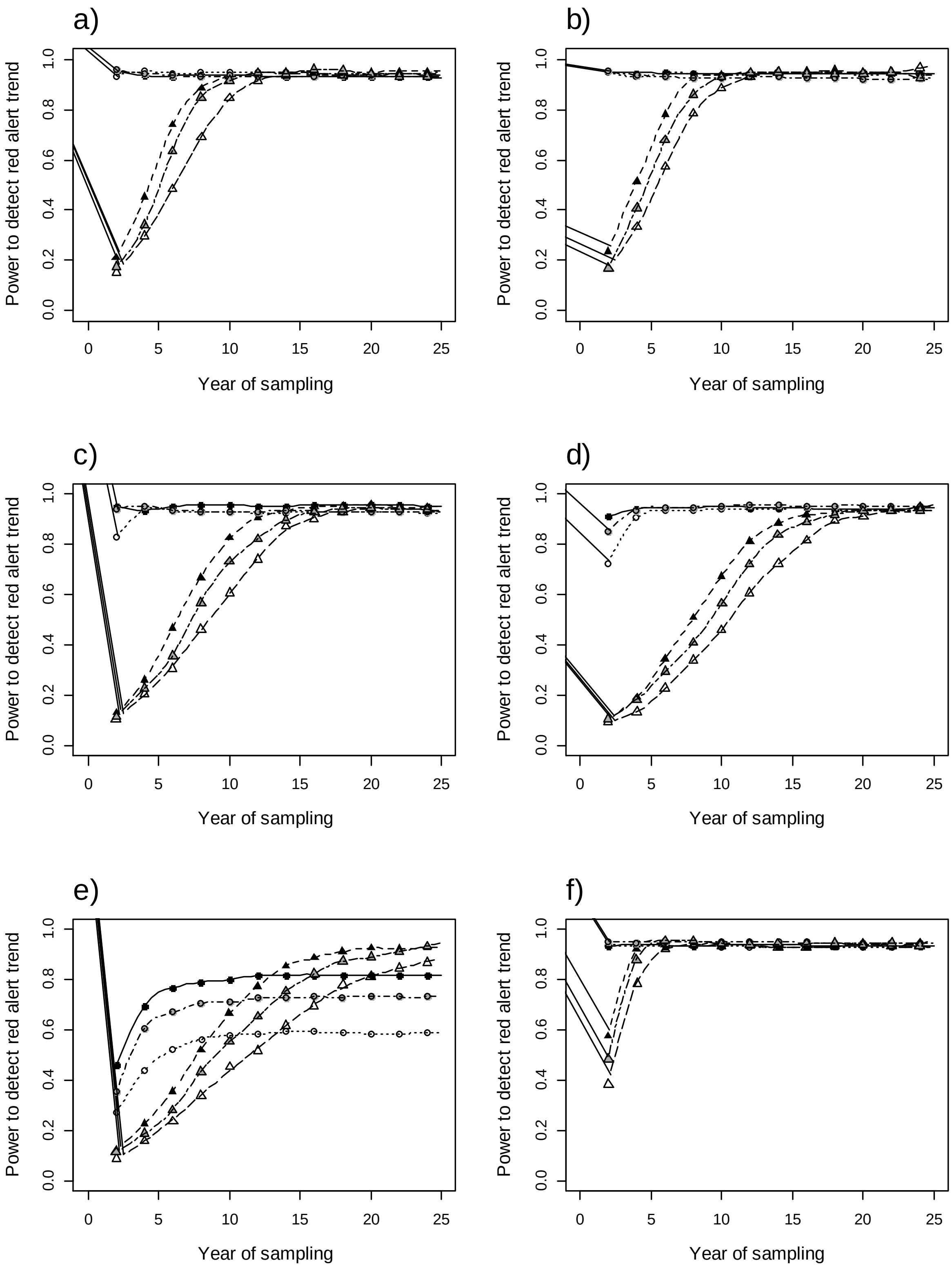


Figure 7: Power for detecting a red-alert trend in (a) available mineralisable nitrogen, (b) bulk density, (c) carbon, (d) nitrogen, (e) Olsen P and (f) pH in relation to the sampling design scenario and the year of sampling (samples every 2 years). (Datasets were simulated assuming the variance around the trend was small but trends varied among orchards; Scenario B in Box 7.) Key: ● existing ARGOS design; two-thirds ARGOS; ○ half ARGOS; ▲ existing ARGOS with orchard turnover; two-thirds ARGOS with orchard turnover; △ half ARGOS with orchard turnover.

FUTURE CHALLENGES

Thinking beyond the 'red-alert' alarm

Overall, our analysis shows that the power to detect a red-alert trend in soil quality on orchards is dependent on the variance characteristics of the real red-alert trend. We recommend, therefore, applying the precautionary principle (i.e. industry assumes the high-variance scenarios reflect the real trends) and maintaining the current frequency of soil surveys within the kiwifruit sector, following the current ARGOS survey design, for at least another 6-year period. This approach would have the following clear advantages:

1. Providing more robust measures of fluctuations in soil quality over time (as the current measures of inter-annual variation are potentially crude, with data only available from three sampling events), to allow the industry to more effectively discriminate a real trend from a false one (e.g. Box 2).
2. Facilitating the timely identification of trends in soil indicators so that remedial action can be implemented before large changes have taken place, which would otherwise be costly and difficult to reverse.
3. Testing the feasibility of detecting more subtle changes in soil quality at the industry level and at finer spatial scales. These could inform more powerful model building to predict farm outputs or examine consequences of different choices in soil nutrient management rather than just treating it as a monitoring framework:
    a. To measure differences in trends among different management panels to determine, for example, if organic growers are running down their soil quality more than integrated management growers or vice versa.
    b. To measure trends within orchards to determine, for example, (i) the optimum soil quality measures for maximising production/dry matter/profit, or (ii) target zones for soil nutrient management.

Assuming that our best-case scenario (where variance around the trend is assumed to be small and the trend consistent among orchards) is a true reflection of real trends in soil quality in New Zealand's kiwifruit orchards, it looks likely that the ARGOS sampling design will have high power to detect a red-alert trend at the industry level for all indicators. If this is the case, the industry will be well positioned to safeguard soil quality, an important ecosystem service for kiwifruit production. At the same time, this result highlights the potential for detecting much more subtle changes in soil quality to inform management at much finer spatial scales (e.g. comparing trends between different management systems or responses to targeted management actions within orchards).

Informing the *New Zealand Sustainability Dashboard* design

The development of a sustainability assessment and reporting tool, the *New Zealand Sustainability Dashboard*, was recently initiated for multiple primary industry sectors within New Zealand; this initiative is being led by ARGOS. It will combine internationally recognised frameworks and their key generic sustainability performance indicators (KPIs), with a subset of complementary KPIs developed specifically for New Zealand and participating sectors. Farmers' performance will be scored across economic, social and environmental dimensions of food and fibre production. This information will not only allow overseas consumers to benchmark and verify the sustainability credentials of goods exported from New Zealand, but will also enable New Zealand industries and farmers to self-regulate their performance. Testing the accuracy and statistical reliability of the candidate KPIs is an important aspect of the *Sustainability Dashboard* design. This will include:

- Establishing clear monitoring objectives, with meaningful threshold points for KPIs to enable a robust assessment of the status and trends of KPIs. Here, for example, a red-alert threshold would indicate breaching regulatory limits or codes of practice, an amber one would highlight that performance is adequate but with scope for improvement, and a green signal would show best practice.

- Testing the power of the monitoring design to report on status and trend of individual KPIs at different temporal and spatial scales (i.e. examining the feasibility of meeting multiple reporting requirements); this will require adapting the power analysis approach presented in this report to accommodate other environmental KPIs, as well as economic, social and production KPIs. This is important because the scale, design and intensity of sampling and the types of data collected will vary among the KPIs. For example, detecting changes in categorical data (e.g. information collected using an ordinal scale to classify performance: excellent, good, poor) will potentially be less sensitive to change relative to indicators measured on a continuous scale (e.g. soil pH). These analyses will be used to prescribe the optimal scale, design and intensity of sampling required to cost-effectively detect deleterious trends or critical thresholds in performance indicators scored by producers themselves.

- Determining the minimum set of cost-effective methods required for reliable sustainability assessment; this will require an integration of KPIs to determine the optimal monitoring design to support multi-functional reporting.

- Developing processes for independent KPI audits to ensure the sustainability assessments are robust, thus allowing overseas consumers and regulatory bodies to verify the reliability of the information provided.

APPENDIX 1: Power simulations

We partitioned the variance in the soils data into different components (time, cluster, orchard, block, etc.) that were then used to simulate new soils data incorporating a 'red-alert' trend. We generated these variance estimates by fitting a separate linear mixed-effects model to the data ('variance-estimation model') for each variable, using Bayesian techniques. Bayesian model fitting assumes that all parameters in the model are random variables with their own distributions. The objective of the model-fitting process was to simulate these distributions, such that summary statistics (mean, standard error, etc.) for each parameter could be calculated. We then simulated new data incorporating a red-alert trend ('RATr-simulation model') using the variance components estimated in the variance-estimation models. Using this framework to simulate new data for power analyses ensures that the uncertainties in the estimates of the variance components are fully reflected in the simulated data. Finally, we fitted statistical models ('power-estimation models') to the simulated data to determine the probability of successfully detecting a red-alert trend given that it is present.

Variance-estimation models

We took two approaches to account for time-dependent variation in the soil variables when estimating the variance components for the power simulations. In the first approach we modelled changes in the soil variables (*Y*) over time as a linear trend:

$$Y = Intercept + \ Location * Management * Year + Cluster/Orchard/Block/Site + Orchard{:}Year + Error \quad (1)$$

This mixed-effects model incorporated sample location (2 levels: between rows and within rows), management (conventional 'Green', organic 'Green', 'Gold'), an annual trend across years and all 2- and 3-way interactions as fixed effects (Location × Management × Year). We included 'cluster', 'orchard' nested within 'cluster', 'block' nested within 'orchard', and 'site' nested within 'block' as random effects (Cluster/Orchard/Block/Site). We also fitted a separate annual trend for each orchard (Orchard:Year interaction). This represented our least conservative scenario; it assumed any linear relationships detected were real and the systematic variation due to them was removed (i.e. was not reflected in the simulated data). In this approach, any time-dependent variation that could not be modelled by a linear trend went into the error variance (and was later used to simulate variation in the new datasets).

In the second approach, we modelled year of sampling as the categorical random effect 'time' (i.e. a separate intercept term for each year, rather than a slope as in the previous model):

$$Y = Intercept + \ Location * Management + Cluster/Orchard/Block/Site + Time + Error \quad (2)$$

This approach recognised that the data came from only three separate time periods, and that it was possible that a trend could be observed in data from a sample of a population that fluctuates around a constant mean. If this was the case then describing that fluctuation with a trend line would lead to an underestimate of the variance components used in the power simulations, and an overestimate of the power. Using this second approach, the data that we simulated for the power calculations included all time-dependent sources of variation as time-specific noise.

Specific details of the models and model fitting are given in Appendix 2.

Model fit of variance-estimation models

All models provided very good fits to the data as measured by Bayesian r^2 (Table S2).

Table S2. Summary of variance estimation model fits (using Bayesian r^2) [67]

Soil variable	Model	Bayesian r^2
AMN*	Trend	0.811
	No trend	0.749
Bulk density	Trend	0.879
	No trend	0.863
Carbon	Trend	0.860
	No trend	0.839
Nitrogen	Trend	0.864
	No trend	0.854
Olsen P	Trend	0.829
	No trend	0.753
pH	Trend	0.821
	No trend	0.741

*Available mineralisable nitrogen

The observed soil variable patterns were similar between the trend and no-trend variance-estimation models (Figs S1 and S2 respectively). Levels of available mineralisable nitrogen, carbon and nitrogen were lower within rows than between rows. Olsen P was higher within rows that between rows. Only bulk density and pH showed differences between and within rows that were dependent on management system. pH was slightly higher within rows in organic ‘Green’ orchards and lower within rows in ‘Gold’ orchards compared with conventional ‘Green’ orchards. Bulk density within rows was slightly lower in organic ‘Green’ orchards than in the other management types. Carbon and nitrogen were both higher, and bulk density lower, in organic ‘Green’ and ‘Gold’ orchards than conventional ‘Green’ orchards. Available mineralisable nitrogen was also higher in organic ‘Green’ orchards than under the other management systems. For the trend model, positive trends were observed in carbon and nitrogen over time. For nitrogen between rows, the trend-over-time's 95% credible intervals (0.0097–0.0161) indicate that nitrogen is increasing faster than the red-alert trend (0.0088, specified in Box 5). Overall, bulk density declined over time.

The key variance estimates defining the variance scenarios are shown in Table S2. For bulk density and pH, the median variance estimates for the 'small-variance' (σ^2_{trend}) and 'large-variance' ($\sigma^2_{notrend}$ + σ^2_{time}) scenarios are identical. For available mineralisable nitrogen, the median variance components in the 'small-variance' scenario are marginally smaller than for the 'large-variance' scenario. For the remainder of the soil variables (carbon, nitrogen and Olsen P), the median variance estimates in the 'large-variance' scenario are 1.6 to 3.5 times larger than the 'small-variance' scenario. The variance estimate for orchard-to-orchard variation in the slope of the trend is relatively small by comparison with the main variance components (range 0–3.8% of the main component in the 'small-variance' scenarios).

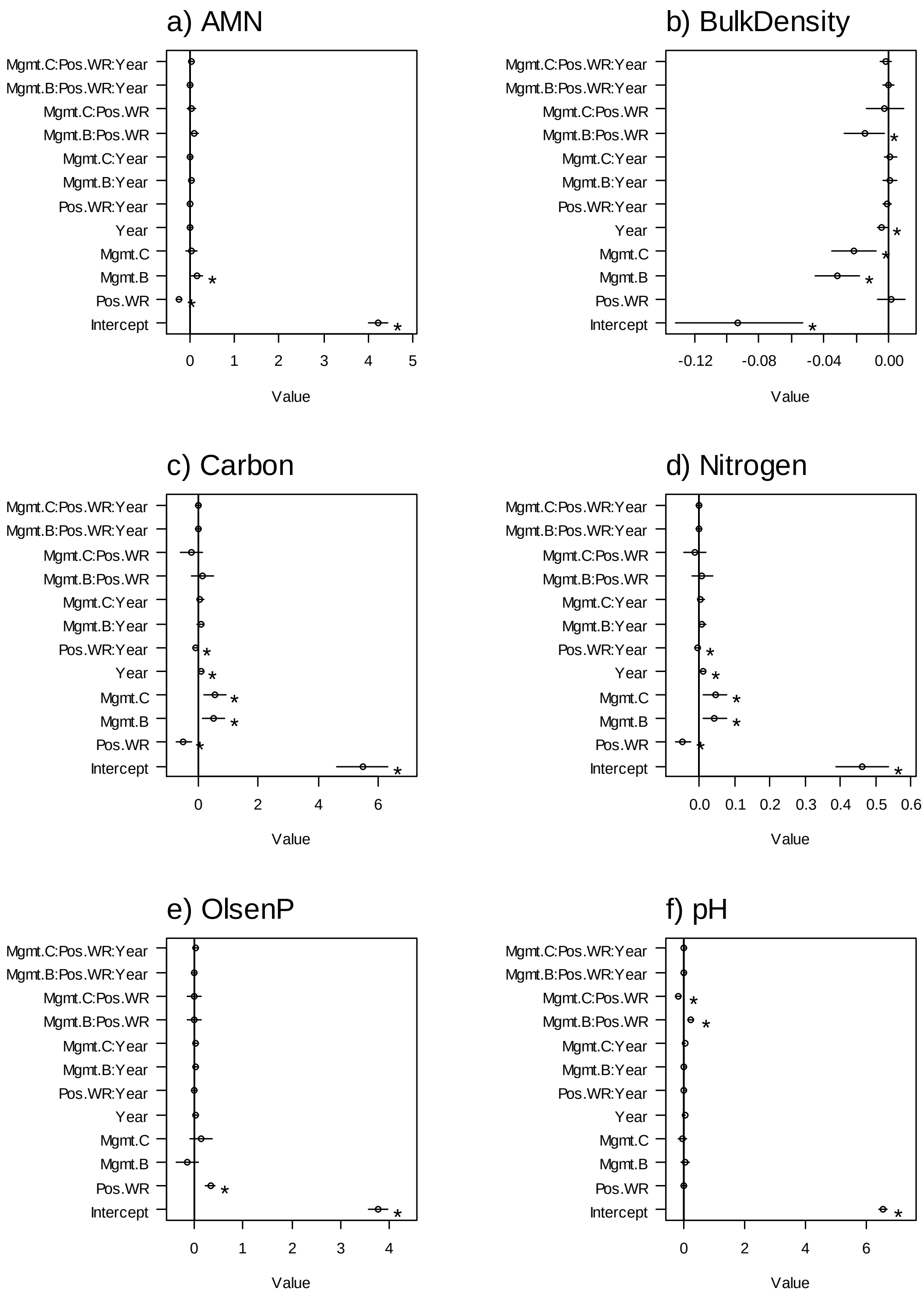


Figure S1: Median parameter values (○) and 95% credible intervals (horizontal lines) of the fixed effects from the variance-estimation trend models for each of the six soil variables. * indicate 95% CI that do not overlap zero (i.e. are 'significant'). Parameter values of factors (location and management) give the effects relative to a baseline level. For location within a block, the baseline is between-row samples. For management, the base level is A (conventional 'Green' kiwifruit). Key: Pos.WR (sampling location is within rows); Mgmt.B (organic 'Green' kiwifruit block); Mgmt.C ('Gold' kiwifruit block); Year describes the linear trend over time. ':' indicates an interaction effect between variables (e.g. Pos.WR:Year is the difference in trend between the average year effect and the year effect within rows). AMN is available mineralisable nitrogen.

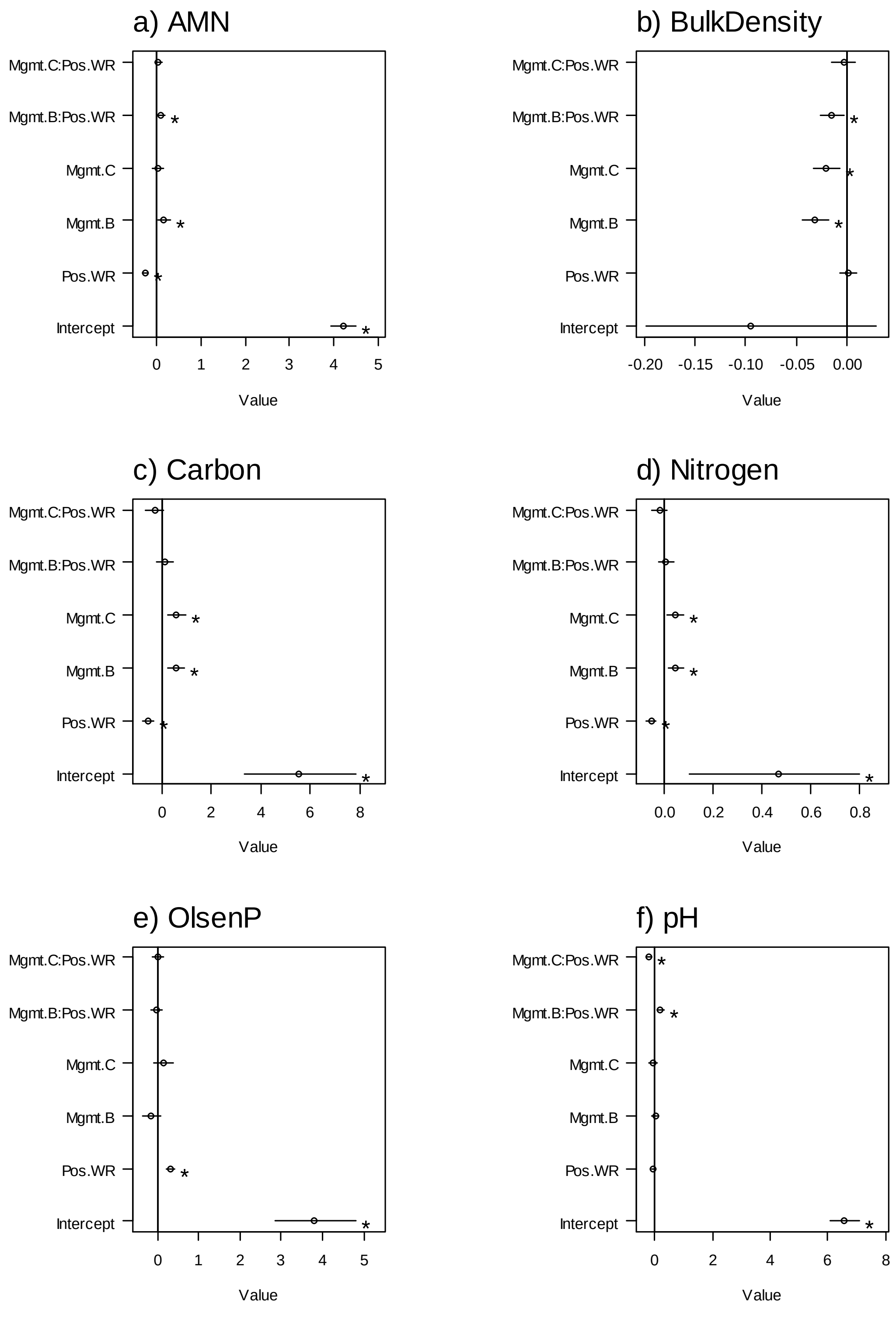


Figure S2: Median parameter values (○) and 95% credible intervals (horizontal lines) of the fixed effects from the variance-estimation no-trend models for each of the six soil variables. * indicate 95% CI that do not overlap zero (i.e. are 'significant'). Parameter values of factors (location and management) give the effects relative to a baseline level. For location within a block, the baseline is between-row samples. For management, the base level is A (conventional 'Green' kiwifruit). Key: Pos.WR (sampling location is within rows); Mgmt.B (organic 'Green' kiwifruit block); Mgmt.C ('Gold' kiwifruit block). ':' indicates an interaction effect between variables (e.g. Mgmt.B:Pos.WR describes how the within-row effect in organic 'Green' kiwifruit blocks differs from the overall within-row effect). AMN is available mineralisable nitrogen.

Table S2: Median variance component estimates from the variance estimation models (95% credible intervals).

Soil variable	No trend model					Trend model			
	$\sigma_{notrend}$	$\sigma^2_{notrend}$	σ_{year}	σ^2_{year}	$\sigma^2_{notrend} + \sigma^2_{year}$	σ_{trend}	σ^2_{trend}	$\sigma_{orchtrend}$	$\sigma^2_{orchtrend}$
AMN	0.173 (0.161–0.187)	0.03	0.029 (0.001–0.829)	0.001	0.031	0.199 (0.186–0.215)	0.040	0.035 (0.025–0.049)	0.001
Bulk density	0.022 (0.021–0.024)	0	0.024 (0.007–0.443)	0.001	0.001	0.024 (0.022–0.026)	0.001	0.004 (0.003–0.006)	0
Carbon	0.563 (0.513–0.718)	0.317	0.670 (0.197–6.251)	0.449	0.766	0.588 (0.549–0.634)	0.346	0.101 (0.072–0.140)	0.01
Nitrogen	0.047 (0.044–0.051)	0.002	0.069 (0.020–1.420)	0.005	0.007	0.049 (0.046–0.053)	0.002	0.009 (0.007–0.012)	0
Olsen P	0.198 (0.184–0.214)	0.039	0.227 (0.057–3.380)	0.051	0.090	0.238 (0.222–0.256)	0.056	0.050 (0.038–0.068)	0.002
pH	0.135 (0.125–0.145)	0.018	0.092 (0.011–2.247)	0.008	0.026	0.162 (0.151–0.174)	0.026	0.037 (0.028–0.050)	0.001

*Available mineralisable nitrogen

RATr-simulation models

To estimate the power of the different sampling designs, 1000 datasets were simulated for each design using different RATr-simulation models. The RATr-simulation models were mathematical expressions specifically used to create the simulated data. The values of the parameters in the RATr-simulation models were obtained from the two variance-estimation models (trend and no trend) fitted to the soil nutrient data, and incorporated the red-alert trend. All RATr-simulation models followed the general form:

$$\mathrm{Y} = \mathrm{Location} * \mathrm{Management} + \mathrm{Red\ alert\ trend} + \mathrm{Cluster/Orchard/Block/Site} + \mathrm{Var} + \mathrm{Error} \qquad (3)$$

with the 'Var' component changing between models (Table S3). RATr-simulation models A to D were used to evaluate the existing ARGOS design (i.e. the number of clusters, properties, and blocks was as per the existing sampling scheme), under various assumptions about the nature of the variance in the data. We generated model E and used it to evaluate the alternative designs where we varied the sampling intensity and the replacement of orchards during each sampling event.

Table S3. RATr-simulation models and the variance-estimation models used to estimate the variance components for the simulated datasets

RATr-simulation model	Variance-estimation model	Between-orchard variance in red-alert trend†	Variable component ('Var' in equation 3)
Evaluating the existing design			
A	Trend	No	–
B	Trend	Yes	Orchard-specific trend
C	No trend*	No	Time
D	No trend*	Yes	Orchard-specific trend + Time
Evaluating alternative designs‡			
E	Trend	Yes	Orchard-specific trend

* Simulated variance included the time-variance component estimated in the no-trend model.
† Variance estimated from orchard-specific slope random effect in the trend model.
‡ Designs included full, two-thirds and half ARGOS sampling effort and repeated sampling of the same orchards compared with selecting new orchards at each sampling period.

All RATr-simulation models assumed the fitted main effects and interaction for location within blocks and management (location × management). In addition, we used the random-effects-variance components to simulate sites, nested within blocks nested within orchards nested within clusters (Cluster/Orchard/Block/Site). We assumed the error variance as estimated by the variance-estimation models. RATr-simulation models C and D also included random variation due to the time of sampling, estimated from variance-estimation model equation 2. We simulated the red-alert trend in two ways. For RATr-simulation models A and C, we assumed a fixed ‘red-alert’ trend across all of the simulations for each orchard. For RATr-simulation models B, D and E, where the simulated red-alert trend was orchard-specific, we assumed that the trend was drawn from a normal distribution, with mean equal to the ‘red-alert’ trend and standard deviation estimated from the orchard-specific trend (Orchard:Year term) random effect from variance-estimation model equation 1. For evaluating the existing ARGOs design, we simulated data for a number of sampling scenarios in which the interval between samples and the length of time that sampling had been occurring were changed. For evaluating the alternative designs we varied the sampling effort (full, two-thirds, and half the ARGOS effort corresponding to 12, 8 and 6 clusters of 3 orchards respectively) and whether the same orchards were sampled each time or different orchards.

To calculate the power, we fitted random-effects models (power-estimation models) to each simulated dataset, using maximum likelihood. For scenarios in which the same orchards were repeatedly sampled we fitted a power-estimation model equivalent to variance-estimation model 1 to the data (RATr-simulation models A–D and E with repeated measures, Table S3). The same power-estimation model, but without a specific orchard-level random effect, was fitted to scenarios in which the same orchards were not repeatedly sampled (RATr-simulation model E with new orchards for each sample, Table S1). The power for each scenario was estimated as the proportion of the simulations in which a significant trend was detected and that this trend was not significantly smaller than a red-alert trend (both at statistical significance level $\alpha = 0.05$).

APPENDIX 2: Description of the Bayesian variance-estimation models

The trend variance-estimation model

We modelled the i^{th} soil observation, at the m^{th} site in the n^{th} position nested within block l, within the k^{th} orchard, within the j^{th} cluster, under the p^{th} management type ($Y_{ijklmnp}$) according to the following equation:

$Y_{ijklmnp} \sim N(\mu_{ijklmnp}, \sigma^2)$

where σ^2 is the error variance

$$\mu_{ijklmnp} = \beta_0 + \beta_{1n} + \beta_{2p} + \beta_{3np} + x_i(\beta_4 + \beta_{5n} + \beta_{6p} + \beta_{7np} + \alpha_{5k}) + \alpha_{1j} + \alpha_{2k} + \alpha_{3kl} + \alpha_{4klm} \quad (4)$$

The model terms are as follows: β_0 is the intercept term; β_{1n} is the effect of the position within block relative to the between-rows position (within rows or between rows); β_{2p} is the effect of p^{th} management type ('Green', 'Gold' or organic 'Green') relative to the gold type; β_{3np} describes how position n varies according to management type p; β_4 is the overall mean trend with year (x_i); β_{5n} describes how the trend varies depending on position within block; β_{6p} describes how the trend varies depending on management type; β_{7np} describes how the position-dependent trend varies with management type; α_{1j} is the effect of cluster j; α_{2k} is the effect of orchard k; α_{3lk} is the effect of block l nested with orchard k; α_{4klm} is the effect of site m nested within block l and orchard k; α_{5k}. describes how the average trend varies with orchard k.

We assumed uninformative Bayesian priors for all parameters. Specifically the β were assumed $\sim N(0, 10^6)$, the α_v assumed $\sim N(0, \sigma_v)$ with each $\sigma_v \sim U(0, 100)$. The prior distribution on σ was also assumed $\sim U(0, 100)$.

The 'no trend' variance-estimation model

The no-trend variance-estimation model differed from the trend model by excluding all of the year terms described in the trend model and modelling the q^{th} time as an orchard-dependent categorical effect. Each soil variable ($Y_{ijklmnpq}$) was modelled according to the following equation:

$Y_{ijklmnpq} \sim N(\mu_{ijklmnpq}, \sigma^2)$

$$\mu_{ijklmnp} = \beta_0 + \beta_{1n} + \beta_{2p} + \beta_{3np} + \alpha_{1j} + \alpha_{2k} + \alpha_{3kl} + \alpha_{4klm} + \alpha_{5q} + \alpha_{6kq} \tag{5}$$

The model terms are as described for the trend model, with the addition of α_{5q} as the categorical effect of year *q,* and α_{6kq} the categorical effect of year *q* dependent on orchard *k*. We assumed uninformative priors as per the trend model.

Model fitting

All variance-estimating models were fitted using jags v3.1[68], accessed from R (R Development Core Team 2012 http://www,R-project.org) using the package ‘dclone’[69]. The models were run for 10 million iterations or until convergence. Model convergence was assessed using BGR plots and visual inspection of simulation traces.

[1] Yoccoz NG, Nichols JD, Boulinier T 2001. Monitoring of biological diversity in space and time. Trends in Ecology and Evolution 16: 446–453.

[2] Wilson JD, Evans AD, Grice PV 2010. Bird conservation and agriculture: a pivotal moment? Ibis 152: 176−179.

[3] Krebs JR, Wilson JD, Bradbury RB, Siriwardena GM 1999. The second Silent Spring? Nature 400: 611–612.

[4] Donald PF, Green RE, Heath MF 2001. Agricultural intensification and the collapse of Europe's farmland bird populations. Proceedings of the Royal Society London Series B 268: 25–29.

[5] Benton TG, Bryant DM, Cole L, Crick HQP 2002. Linking agricultural practice to insect and bird populations: a historical study over three decades. Journal of Applied Ecology 39: 673–687.

[6] Butchart SHM, Walpole M, Collen B, van Strien A, Scharlemann JPW, Almond REA, Baillie JEM, Bomhard B, Brown C, Bruno J, Carpenter KE, Carr GM, Chanson J, Chenery AM, Csirke J, Davidson NC, Dentener F, Foster M, Galli A, Galloway JN, Genovesi P, Gregory RD, Hockings M, Kapos V, Lamarque J-F, Leverington F, Loh J, McGeoch MA, McRae L, Minasyan A, Hernández Morcillo M, Oldfield TEE, Pauly D, Quader S, Revenga C, Sauer JR, Skolnik B, Spear D, Stanwell-Smith D, Stuart SN, Symes A, Tierney M, Tyrrell TD, Vié J-C, Watson R 2010. Global biodiversity: indicators of recent declines. Science 328: 1164–1168.

[7] Gregory RD, Noble DG, Custance J 2004. The state of play of farmland birds: population trends and conservation status of lowland farmland birds in the United Kingdom. Ibis 146 (Suppl. 2): 1–3.

[8] Golden JS, Dooley KJ, Anderies JM, Thompson BH, Gereffi G, Pratson L 2010. Sustainable product indexing: navigating the challenge of ecolabeling. Ecology and Society, 15, 8. [online]. URL: http://www.ecologyandsociety.org/vol15/iss3/art8.

[9] Siriwardena GM, Baillie SR, Buckland ST, Fewster RM, Marchant JH, Wilson JD 1998. Trends in the abundance of farmland birds: a quantitative comparison of smoothed Common Birds Census indices. Journal of Applied Ecology 35: 24−43.

[10] Chamberlain DE, Fuller RJ, Bunce RGH, Duckworth JW, Shrubb M 2000. Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. Journal of Applied Ecology 37: 771–788.

[11] Fewster RM, Buckland ST, Siriwardena GM, Baillie SR, Wilson JD 2000. Analysis of population trends for farmland birds using generalized additive models. Ecology 81: 1970–1984.

[12] Peach WJ, Lovett LJ, Wotton SR, Jeff C 2001. Countryside stewardship delivers cirl buntings (*Emberiza cirlus*) in Devon, UK. Biological Conservation 101: 361–373.

[13] Hole DG, Whittingham MJ, Bradbury RB, Anderson GQA, Lee PLM, Wilson JD, Krebs JR 2002. Agriculture: Widespread local house-sparrow extinctions. Nature 418: 931–932.

[14] PCE (Parliamentary Commissioner for the Environment) 2004. Growing for good. Intensive farming, sustainability and New Zealand's environment. Wellington, Parliamentary Commissioner for the Environment.

[15] MacLeod CJ, Moller H. 2006. Intensification and diversification of New Zealand agriculture since 1960: an evaluation of current indicators of land use change. Agriculture, Ecosystems and Environment 115: 201–218.

[16] Meurk CD, Swaffield SR 2000. A landscape ecological framework for indigenous regeneration in rural New Zealand-Aotearoa. Landscape & Urban Planning 50: 129–144.

[17] Norton DA, Miller CJ 2000. Some issues and options for the conservation of native biodiversity in rural New Zealand. Ecological Management and Restoration 1: 26–34.

[18] Perley C, Moller H, Hamilton WJ, Hutcheson J 2001. Towards safeguarding New Zealand's agricultural biodiversity: research gaps, priorities and potential case studies. Ecosystems Consultants Report 23, prepared for the Ministry for Agriculture and Forestry. 230 p. [Online at http://www.maf.govt.nz]

[19] PCE (Parliamentary Commissioner for the Environment) 2010. How clean is New Zealand? Measuring and reporting on the health of our environment. Wellington, Parliamentary Commissioner for the Environment.

[20] Moller H, Wearing A, Pearson A, Perley C, Steven D, Blackwell G, Reid J, Johnson M. 2005. Environmental monitoring and research for improved resilience on ARGOS farms. ARGOS Working Paper No. 6. 136 p.

[21] Moller H, MacLeod CJ, Haggerty J, Rosin C, Blackwell G, Perley C, Meadows S, Weller F, Gradwohl M 2008. Intensification of New Zealand agriculture; implications for biodiversity. New Zealand Journal of Agricultural Research 51: 253–263.

[22] MacLeod CJ, Blackwell G, Moller H, Innes J, Powlesland R 2008. The forgotten 60%: bird ecology and management in New Zealand's agricultural landscape. New Zealand Journal of Ecology 32: 240–255.

[23] MacLeod CJ, Tompkins DM, Drew KW, Pyke N 2011. Does farm-level habitat composition predict pest bird numbers and distribution? Wildlife Research 38: 464–474.

[24]MacLeod CJ, Blackwell G, Weller F, Moller H 2012a. Designing a bird monitoring scheme for New Zealand's agricultural sectors. New Zealand Journal Ecology 36: 312–323.

[25] HortResearch 2007. Fresh Facts. New Zealand Horticulture 2007. The Horticulture & Food Research Institute of New Zealand Ltd, Auckland, New Zealand. http://www.hortresearch.co.nz/files/aboutus/factsandfigs/ff2007.pdf

[26] Campbell H, Fairweather J, Steven D 1997. Recent developments in organic food production in New Zealand: part 2, Kiwifruit in the Bay of Plenty. Studies in Rural Sustainability No. 2, Department of Anthropology, Otago University.

[27] Walker JTS 2005. Development of GAP programs in New Zealand's fruit industries October 25th 2005. In: Proceedings of International Seminar on Technology Development for Good Agricultural Practice in Asia and Oceania. October 25–26 2005, Tsukuba, Japan. Pp. 15–28.

[28] Steven D, Benge J 2007. Sprays on New Zealand kiwifruit – use patterns and outcomes. Proceedings of the Sixth International Symposium on kiwifruit. Volume 2. Acta Horticulturae 735: 711–717.

[29] Moller H, Wearing A, Perley C, Roxin C, Blackwell G, Campbell H, Hunt L, Fairweather J, Manhire J, Benge J, Emanuelsson M, Steven D 2007. Biodiversity on kiwifruit orchards: the importance of shelterbelts. Proceedings of the Sixth International Symposium on kiwifruit. Volume 2. Acta Horticulturae 735: 609–618.

[30] Todd TH, Malone LA, McArdle BH, Benge J, Poulton J, Thorpe S, Beggs JR 2011. Invertebrate community richness in New Zealand kiwifruit orchards under organic or integrated pest management. Agriculture, Ecosystems and Environment 141: 32–38.

[31] MacLeod CJ, Blackwell G, Benge J 2012. Reduced pesticide toxicity and increased woody vegetation cover account for enhanced native bird densities in organic orchards. Journal of Applied Ecology 49: 652–660.

[32] Carey PL, Benge JR, Haynes RJ 2009. Comparison of soil quality and nutrient budgets between organic and conventional kiwifruit orchards. Agriculture, Ecosystems and Environment 132: 7–15.

[33] Bengtsson J, Ahnström J, Weibell A-C 2005. The effects of organic agriculture on biodiversity and abundance: a meta-analysis. Journal of Applied Ecology 42: 261–269.

[34] Hole DG, Perkins AJ, Wilson JD, Alexander IH, Grice PV, Evans AD 2005. Does organic farming benefit biodiversity? Biological Conservation 122: 113–130.

[35] Fuller RJ, Norton LR, Feber RE, Johnson PJ, Chamberlain DE, Joys AC, Mathews F, Stuart RC, Townsend DW, Firbank LG 2005. Benefits of organic farming to biodiversity vary among taxa. Biology Letters 1: 431–434.

[36] Gomiero T, Pimental D, Paoletti MG 2011. Environmental impact of different agricultural management practices: Conventional vs. Organic agriculture. Critical Reviews in Plant Sciences 30: 95–124.

[37] Reganold JP, Glover JS, Andrews PK, Hinman HR 2001. Sustainability of three apple production systems. Nature 410: 926–930.

[38] Lotter DW 2003. Organic agriculture. Journal of Sustainable Agriculture 21: 59–128.

[39] Gregory RD 2000. Development of breeding bird monitoring in the United Kingdom and adopting its principles elsewhere. Ringing 122: 35–44.

[40] Voříšek P, Klvaňová A, Wotton S, Gregory RD 2008. A best practice guide for wild bird monitoring schemes. First edition, CSO/RSPB, Prague.

[41] Sergeant CJ, Moynahan BJ, Johnson WF 2012. Practical advice for implementing long-term ecosystem monitoring. Journal of Applied Ecology doi: 10.1111/j.1365-2664.2012.02149.x

[42] Thompson WL 2002. Towards reliable bird surveys: accounting for individuals present but not detected. The Auk 119: 18–25.

[43] Buckland ST, Marsden SJ, Green RE 2008. Estimating bird abundance: making methods work. Bird Conservation International 18: S91–S108.

[44] Guillera-Arroita G, Ridout MS, Morgan JT 2010. Design occupancy studies with imperfect detection. Methods in Ecology and Evolution 1: 131–139.

[45] Field SA, O'Connor PJ, Tyre AJ, Possingham HP 2007. Making monitoring meaningful. Austral Ecology 32: 485–491.

[46] Guillera-Arroita G, Lahoz-Monfort JJ 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection. Methods in Ecology and Evolution 3: 860–869.

[47] Field SA, Tyre AJ, Possingham HP 2005. Optimising allocation of monitoring effort under economic and observational constraints. Journal of Wildlife Management 69: 473–482.

[48] Schalk G, Dewar HJ, Cadman MD 2002. Recommendations for assessing trends in forest bird populations based on the experience of the Ontario Forest Bird Monitoring Program. Journal of Field Ornithology 73: 340–350.

[49] Legg CJ, Nagy L 2006. Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management 78: 194–199.

[50] Sims M, Elston DA, Harris MP, Wanless S 2007. Incorporating variance uncertainty into a power analysis of monitoring designs. Journal of Agricultural, Biological and Environmental Statistics 12: 236–249.

[51] Thomas L, Martin K 1996. The importance of analysis method for breeding bird survey population trend estimates. Conservation Biology 10: 479–490.

[52] Power AG 2010. Ecosystem services and agriculture: trade-offs and synergies. Philosophical Transactions Royal Society of London B 365: 2959–2971. doi: 10.1098/rstb.2010.0143

[53] Zhang W, Ricketts TH, Kremen C, Carney K, Swinton S M 2007. Ecosystem services and dis-services to agriculture. Ecological Economics 64: 253–260. doi:10.1016/j.ecolecon.2007.02.024.

[54] Schoenholtz SH, Miegroet HV, Burger JA 2000. A review of chemical and physical properties as indicators of forest soil quality: challenges and opportunities. Forest Ecology and Management 138: 335–356.

[55] Pretty J, Sutherland WJ, Ashby J, Auburn J, Baulcombe D, Bell M, Bentley J, Bickersteth S, Brown K, Burke J, Campbell H, Chen K, Crowley E, Crute I, Dobbelaere D, Edwards-Jones G, Funes-Monzote F, Godfray HCJ, Griffon M, Gypmantisiri P, Hadda L, Halavatau S, Herren H, Holderness M, Izac AM, Jones M, Koohafkan P, Lal R, Lang T, McNeely J, Mueller A, Nisbett N, Noble A, Pingali P, Pinto Y, Rabbinge R, Ravindranath NH, Rola A, Roling N, Sage C, Settle W, Sha JM, Shiming L, Simons T, Smith P, Strzepeck K, Swaine H, Terry E, Tomich TP, Toulmin C, Trigo E, Twomlow S, Vis JK, Wilson J, Pilgrim S 2010. The top 100 questions of importance to the future of global agriculture. International Journal of Agricultural Sustainability 8: 219–236.

[56] Norton S, Lucock D, Moller H, Manhire J 2010. Energy return on investment for dairy and sheep/beef farms under conventional, integrated or organic management. Proceedings of the New Zealand Grassland Association 72: 145–150.

[57] Benge J, Carey P 2011. Analysis of soil quality in New Zealand's kiwifruit orchards and industry implications. Report for ZESPRI International Ltd (Project EC1143).

[58] Campbell H, Fairweather J, Manhire J, Saunders C, Moller H, Reid J, Benge J, Blackwell G, Emanuelsson M, Greer G, Hunt L, Lucock D, Rosin C, Carey P, Norton D, MacLeod CJ 2012. The Agriculture Research Group on Sustainability Programme: A longitudinal and transdisciplinary study of agricultural sustainability in New Zealand. ARGOS Research Report No. 12/01. 145 p.

[59] Benge J, Manhire J, Pearson AJ, Reid J, Moller H 2007. Differences in soil quality between and within organic and integrated management kiwifruit orchards in New Zealand. Acta Horticulturae 753: 599–608.

[60] Richards S 2005. Evaluation of soil fauna as potential indicators of soil quality in kiwifruit orchard systems. Dunedin, University of Otago.

[61] Richards S, Benge J, Moller H 2006. Soil nematodes in kiwifruit orchards. ARGOS Research Note 15.

[62] Richards S, Hewson K, Moller H, Wharton D, Campbell H, Benge J, Manhire J 2007. Soil biota as indicators of soil quality in organic and integrated management kiwifruit orchards in New Zealand. Acta Horticulturae 753: 627–632.

[63] Darnhofer I, Fairweather J, Moller H 2010. Assessing a farm's sustainability: Insights from resilience thinking. International Journal of Agricultural Sustainability 8: 186–198.

[64] Hoenig JM, Heisey DM 2001. The abuse of power: The pervasive fallacy of power calculations for data analysis. American Statistician 55: 19–24.

[65] Sparling G, Lilburne L, Vojvodic-Vukovic M 2008. Provisional targets for soil quality indicators in New Zealand. Landcare Research Science Series 34. Lincoln, Landcare Research.

[66] Urquhart NS, Paulsen SG, Larsen DP 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications 8: 246–257.

[67] Gelman A, Pardoe I 2006. Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. Technometrics 48: 241–251.

[68] Plummer M 2011 Jags Version 3.1 manual. http://ftp.freebsd.org/pub/FreeBSD/ports/distfiles/mcmc jags/jags_user_manual.pdf (accessed 10 April 2012).

[69] Solymos P 2010. dclone: Data cloning in R. The R Journal 2(2): 29–37. http://journal.r-project.org/archive/2010-2/RJournal_2010-2_Solymos.pdf. (accessed 10 April 2012).