Chapter 8
Methodology of Pollution Ecology: Problems and Perspectives

8.1 Importance of Observational Studies

Charles J. Krebs (1989) started his famous book *Ecological Methodology* with the following statement: 'Ecologists collect data and, as in other fields in biology, the data they collect are to be used for testing hypotheses'. This is true, but where do the new hypotheses originate?

The roots of many theories, including the Newton's law of universal gravitation and Darwin's evolutionary theory, emerge from systematic or occasional observations. Accumulation of data on rainfall acidity, measurements of which for a long time were driven by scientific curiosity, in the 1950s and 1960s allowed scientists to determine the origin of acidity and recognise the damaging effects of acid rain. The importance of observational studies (sometimes termed mensurative experiments) still remains high, especially for environmental sciences, which often face novel problems associated with the rapid development of our civilisation.

Environmental scientists have often been blamed for preferring a narrative approach to hypothesis testing. For example, only 20% of papers published in the *Journal of Applied Ecology* in 1999 explicitly stated clearly testable hypotheses (Ormerod et al. 1999). However, in our opinion, this situation reflects a shortage of relevant hypotheses rather than the incompetence of environmental scientists or their reluctance to use the hypothesis-testing approach. Our survey of several dozens of publications summarising the knowledge on pollution effects on biota (Smith 1974; Miller & McBridge 1975; Kozlowski 1980; Auerbach 1981; Alstad et al. 1982; Newman & Schreiber 1984; Odum 1985; Rapport et al. 1985; Schindler 1987; Sigal & Suter 1987; Bååth 1989; Freedman 1989; Riemer & Whittaker 1989; Treshow & Anderson 1989; Barker & Tingey 1992; Heliövaara & Väisänen 1993; Clements & Newman 2002) turned up quite a few testable predictions.

The advancement of pollution ecology requires further accumulation of observational data on changes in landscapes, ecosystems, communities, populations and individual organisms occurring around industrial polluters. This information is necessary for exploratory analyses leading to the generation of specific hypotheses, which can be tested using field and laboratory experiments.

8.2   Interpretation of Experimental Results

A substantial part of our knowledge of responses of organisms, populations, and communities to pollution originates from experimental studies, mostly conducted in artificial (laboratory) environments. For example, 42 of 50 relevant recent (2008) publications (25 from *Environmental Pollution* and another 25 from *Water, Air, and Soil Pollution*) referred to laboratory experiments, compared to 11 papers that reported field manipulations; only four papers combined these approaches. Six papers reported use of micro- or mesocosms, while the remaining majority of experimental systems were obviously oversimplified in terms of both abiotic and biotic environments.

More generally, ecosystem-level and community-level field experiments with industrial pollutants remain as rare as they were 2 decades ago (Schindler 1987), and most of these experiments are conducted in aquatic systems. The limited number of large-scale manipulations with terrestrial biota is considered a significant shortcoming of ecotoxicology (Clements & Newman 2002). Harvesting the results of 'unintentional pollution experiments' (Lee 1998) established by industries long ago (Section 1.4) can partially overcome this problem.

Laboratory studies, by eliminating a substantial part of natural variability, are likely to produce biased results, in particular due to (a) investigation of only a few 'model' species, with preferences for short-living and easy-to-handle organisms, (b) unrealistic environments, including the use of closed chambers, unnatural growth media for plants, the absence of mutualists, e.g., mycorrhizal fungi, competitors and benefactors, (c) unrealistic forms in which pollutants are applied to organisms, (d) unrealistic demographic structures of experimental populations, e.g., preferential use of seedlings to explore responses of woody plants, and (e) the short duration of the experiments relative to decades or even centuries of ecosystem exposure to industrial pollutants (Patterson & Olson 1983; Stenström 1991; Sandermann et al. 1997; Saxe et al. 1998; Weltje 1998; Ahonen-Jonnarth et al. 2000; Koster et al. 2006). Importantly, the outcomes of field and laboratory studies addressing pollution effects on biota have not, to our knowledge, been compared systematically (except for some specific ecotoxicological tests: Hose et al. 2006), and therefore the biases introduced by experimental methodology remain insufficiently known.

The importance of indirect effects, which are usually neglected in manipulative studies with industrial pollutants, can be demonstrated by a seemingly paradoxical increase in plant performance, repeatedly observed in the heavily polluted sites near the nickel-copper smelter in Monchegorsk. For example, in industrial barrens, some dwarf shrubs grow and reproduce better than in unpolluted forests, especially when they are sheltered by mountain birch trees (Zvereva & Kozlov 2004, 2005). The leaf size and shoot growth of boreal willow, *Salix borealis* increased with an increase in pollution (Zvereva et al. 1997a). Similarly, birch seedlings planted in metal-contaminated bare ground after 3 years of exposure were taller, had longer leaves and had a higher survival rate than those in the unpolluted forest (Eränen & Kozlov 2009). In our opinion, these 'positive' effects observed in heavily polluted

environments resulted from the absence of competing vegetation that declined decades ago. In general, the effects of pollution on competitors or benefactors of the organism under study may be so strong that they counterbalance direct toxic effects of pollutants.

Additional uncertainties in interpretation of experimental results are introduced by using organisms from populations that have not been exposed to pollutants prior the experiments. One of the examples of an obvious discrepancy between observational and manipulative studies, presumably resulting from this approach, concerns the effects of industrial pollutants on the growth of herbaceous plants. Experimental studies, conducted both in fully controlled environments and under field conditions, usually report the adverse effects of different pollutants on herbaceous plants (Brun et al. 2003; Tuma 2003; Hassan 2004; Rämö et al. 2006; Ryser & Sauder 2006). Consistently, many ecotoxicological tests are based on measurements of plant size or biomass (Rajput & Agrawal 2005; An 2006; Everhart et al. 2006; Rombke et al. 2006; Rooney et al. 2006). At the same time, herbaceous plants naturally occurring near big polluters only rarely differed from plants collected from unpolluted environments in terms of growth characteristics (Kozlov & Zvereva 2007b; Figs. 4.11, 4.12, 4.17 and 4.18). We hypothesise that micro-evolution, often leading to the development of pollution tolerance (Bradshaw & McNeilly 1981; Shaw 1990; Macnair 1997; Barnes et al. 1999; Medina et al. 2007), is a plausible explanation of the discrepancy between the results of controlled experiments and field-collected data (Kozlov & Zvereva 2007b).

We conclude that short-term experiments with non-adapted organisms in oversimplified laboratory environments are likely to overestimate the adverse effects of industrial pollutants. Results of these experiments can be used for estimation of the relative toxicity of different substances, but they are of limited value for both explaining and predicting effects observed in polluted environments. Even more importantly, community responses cannot be inferred from the results of single-species experiments. Thus, both dedicated observational studies and field experiments remain of critical importance for the development of pollution ecology, in particular as a tool to validate the results of laboratory tests.

8.3   The Amount of Reliable Information

It seems that many ecologists consider exploration of the effects of industrial pollution on biota outdated and believe that this research field is nearly exhausted. Both impressions are far from correct.

First and most importantly, inputs of industrial pollutants, primarily sulphur dioxide, into the atmosphere remain extremely high and continue to contribute to local air pollution, smog, acid rain, dry deposition, and global climate change. Although the global SO_2 emissions at the beginning of the 2000s were reduced by 22% relative to the peak value observed at the end of the 1980s, still they remain at the level of the mid-1960s (Stern 2006). Sulphur dioxide was recently identified as the air pollutant

of the highest national concern for India (Agrawal 2005). Thus, there is no doubt that sulphur dioxide and many other pollutants, especially metals and fluorine accumulated in soils, continue to affect terrestrial biota. Even under the most optimistic emission scenarios, these effects will remain an issue of importance for at least several decades (Barcan 2002a; Fowler et al. 1999; Karnosky et al. 2003a).

Second, even the acute effects of pollution are not documented properly. In spite of the high number of publications, the amount of reliable data published in a form suitable for meta-analyses is surprisingly low (Fig. 1.2). Moreover, researchers tend to focus on the impacts imposed by 'pollution superstars', such as large non-ferrous smelters, while the effects caused by smaller industries (emitting 1,000 to 10,000 t SO_2 annually) remain almost unexplored. This shortage of information decreases reliability of estimations of dose-response relationships at lower levels of pollution, i.e., in concentration ranges which are most important for predicting pollution effects at regional scales.

Third, researchers tended to explore pollution effects on species of the highest economic importance, primarily forest-forming conifers and agricultural crops. However, incorporation of pollution effects into biogeochemical models, such as LPJ-GUESS (General Ecosystem Simulator: Smith et al. 2001; Sitch et al. 2003), requires parameterisation of growth responses to pollution (including changes in competitive abilities as well) for all cohorts of plant functional types included. Absence of this information hampers the modelling of regional and global effects of industrial pollution on vegetation structure. Importantly, we are likely to underestimate the importance of both 'low' (i.e., not exceeding the critical loads) depositions of pollutants and 'minor' differences in pollution impacts on plant species or functional groups. For example, the introduction of a small biotic disturbance (insect herbivory) into the LPJ-GUESS model demonstrated that relatively minor damage to birch (annual removal of 1–10% of foliage) changes predictions of future forest composition (Wolf et al. 2008).

Thus, there is acute need for reliable quantitative information concerning the responses of different groups of biota to industrial pollution. This information can be immediately utilised in both building phenomenological models and adjusting existing ecosystem simulators in order to improve our predictions of regional to global pollution effects on terrestrial ecosystems.

8.4 Quality of Information

8.4.1 Design of Impact Studies

For a long time, researchers exploring pollution effects on biota were advised to pay special attention to both experimental design and statistical analysis, because they 'are as important as the choice of monitoring parameters and techniques' (Sigal & Suter 1987). However, both recent meta-analyses (Ruotsalainen & Kozlov 2006; Zvereva et al. 2008; Zvereva & Kozlov 2009; Roitto et al. 2009) and narrative

review of studies addressing forest health (Percy & Ferretti 2004) clearly demonstrated that the majority of primary studies suffer from a number of methodological shortcomings.

Different approaches have been suggested to monitor chronic, local environmental impacts (Stewart-Oaten & Bence 2001). The most well known include intervention analysis, which compares time series before and after an onset of impact at the affected site, so-called BACI design (Before-After, Control-Impact comparisons), and impact versus reference sites. The latter approach estimates error variation among sites, while intervention analysis and BACI estimate error variation over time (Stewart-Oaten & Bence 2001). All approaches require the selection of adequate temporal and spatial scales (Hewitt et al. 2001), and the results of comparisons should always be interpreted with caution, especially in terms of the causality of the observed differences (Section 9.2.3).

Except for dendrochronological analysis, comparison between observations conducted before and after the beginning of impact is only rarely used in studies of biotic effects caused by industrial pollutants. This is mostly due to an absence of adequately collected information about the state of the impacted ecosystems before the onset of pollution. As a rule, the error term in testing for the significance of the effect is obtained from comparisons among study sites. Therefore the quality of information, obtained by comparison between polluted (treatment) and unpolluted (control) habitats depends critically on the number of study sites. For the following analysis, we used a database of approximately 2,000 primary studies that describe responses of terrestrial biota to industrial pollution and the fit criteria listed in Section 5.1.2.

Generally, the sampling design of published studies was poorly replicated. The median number of study sites in a random sample of 1,000 publications was five (Fig. 8.1). Even more importantly, 35% of publications were based on two or
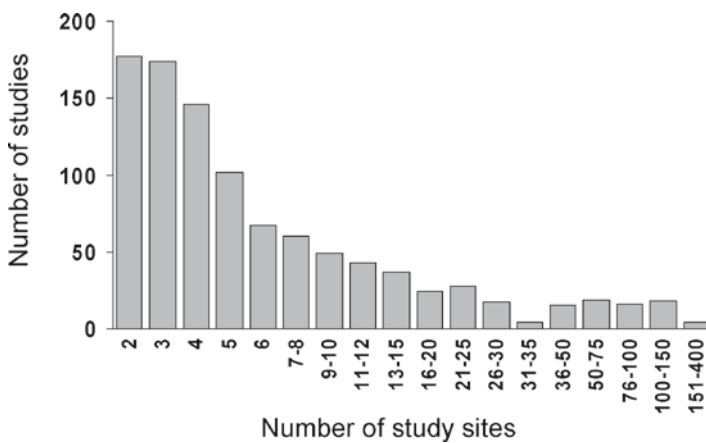

Fig. 8.1 Frequencies of studies based on different numbers of study sites in a random sample of 1,000 publications reporting biotic responses to industrial pollution. For criteria used to select these publications consult Section 5.1.2

three study sites, which means that either 'treatment' or 'control' or both were not replicated. The use of statistical analysis in this situation yields the pseudoreplication problem (Hurlbert 1984; Kozlov 2003, 2007; Kozlov & Hurlbert 2006). Because of repeated attempts to defend the use of non-replicated experimental design (Tatarnikov 2005; Veličković 2007b), we briefly explain the problems arising from comparisons between one polluted and one unpolluted study site.

A statement on the similarity or dissimilarity of two groups of objects is valid only when between-group differences are compared with within-group variation. Very importantly, the level at which the variation is measured within a group is critical. It is also obvious that assessment of within-group variation is only possible when the group consists of more than one object. When there is only a single experimental unit under each treatment, and within-group variation is calculated from measurements made on multiple samples or evaluation units within a single experimental unit (for definitions, consult Kozlov & Hurlbert 2006), then simple pseudoreplication is committed (Hurlbert 1984; Kozlov & Hurlbert 2006). Variation within study sites represents another level, one step down in the hierarchical analysis. This variation can only be used to test for differences between sites, but such tests cannot provide a statistical ground for attributing the differences to pollution (Kozlov 2007).

Earlier meta-analyses (Zvereva & Kozlov 2006a, 2009) demonstrated that effect sizes calculated on the basis of pseudoreplicated studies may be both higher and the same as those based on properly replicated studies. Pseudoreplicated studies allow only calculation of Hedge's d effect size, which increases with both decrease in within-group variation and increase in sample size (Rosenberg et al. 2000). The higher effect sizes obtained by pseudoreplicated studies on insect and plant responses to pollution result from both (a) the generally higher number of samples collected within a site, relative to the number of study sites in properly replicated studies, and (b) generally lower within-site variation relative to between-site variation. Research bias associated with the selection of two 'typical' study sites, one polluted and one unpolluted (control), is likely to further overestimate the effect (Zvereva & Kozlov 2009). Thus, when exclusion of pseudoreplicated studies from meta-analysis results in too low a number of suitable studies (Ruotsalainen & Kozlov 2006; Zvereva & Kozlov 2006a, 2009), the outcomes of these studies should be contrasted with the results based on properly replicated studies to account for the effects of methodology.

To conclude this part of the data quality assessment, we strongly recommend that the impact versus reference sites design includes at least two polluted and two control plots. This design allows correct testing of H0 (no differences between polluted sites and unpolluted sites) by using ANOVA based on site-specific mean values of the character under study. Furthermore, this design allows use of the data in meta-analyses employing both Hedge's d and correlation coefficients as measures of the effect size. However, although four study sites allow the correct use of statistical methods, both the accuracy of the effect estimate and the power of the analysis based on four study sites are rather low (Section 8.4.3).

Another insufficiently explored potential source of problems is the spatial arrangement of study sites. First, this information is reported much more seldom

than the number of study sites: it was missing in 493 of 1,000 publications randomly selected from our database. The majority of studies that reported sampling design in sufficient details either selected each study site in a different direction from the polluter (38%) or positioned study sites along a single gradient (transect) starting from the polluter (37%). The latter design is potentially dangerous due to the possibility to erroneously attribute effects caused by another environmental variable to the effects of pollution. For example, the selection of all study sites to the South of the nickel-copper smelter in Monchegorsk, with control sites located up to 200 km from the polluter (Norin & Yarmishko 1990; Kabirov 1993; Zhirov et al. 1993; Evdokimova 2000; Lukina et al. 2005), does not allow discrimination of the effects of pollution from climatic (latitudinal) variation. Choosing study sites in several directions from the polluter (Figs. 2.2–2.19) is likely to minimise impact of confounding variables on the outcome of data analyses.

8.4.2 Signal to Noise Ratio

Most of studies addressing pollution impacts on biota were initiated when effects were already evident, and papers describing severely degraded ecosystems around major smelters (Gordon & Gorham 1963; Wood & Nash 1976; Freedman & Hutchinson 1980b) are most frequently used to illustrate consequences of pollution impact on biota. However, it would be risky to rely on these studies alone to predict ecosystem responses to contamination (Schindler 1987).

Wolterbeek et al. (1996) criticised the literature on biomonitoring for focussing on the detected changes in monitored parameters (signal) and neglecting the information about the range of natural variation (noise). The latter information is indeed essential to test whether the monitored character began to vary outside its normal range, thus indicating that the ecosystem is perturbed or stressed (Hurlbert 1984; Schindler 1987; Hewitt et al. 2001). The problem is that we usually do not know the normal range for any variable, at least for time periods greater than a few years (Schindler 1987). The only exception is dendrochronology, routinely accounting for natural variation in tree growth over decades or even centuries. Even exploration of insect population dynamics is commonly (Turchin 1990; Berryman 1994) based on time series shorter than 30–40 generations (which are considered ideal for detecting factors that influence population dynamics: Royama 1992), because longer-term data are rare (Hunter & Price 1998). Cyclic fluctuations in many population characteristics always pose a risk of erroneous interpretation of the results of short-term studies: different characteristics of populations from polluted and unpolluted habitats may result from asynchronous density changes in spatially isolated populations rather than from pollution impact on population dynamics (Zvereva et al. 2002; Zvereva & Kozlov 2006b, 2009).

In our study we detected significant between-site variation in 73.8% of 782 statistical tests (Chapters 3–7); however, only 20.4% of 1,446 individual correlation coefficients demonstrated significant relationships between the explored characteristics and pollution load. The proportion of significant correlations increased with

the proportion of significant tests of between-site variation (r_s = 0.66, N = 19 characters, P = 0.0019), hinting that the contribution of detectable pollution-related effects to the overall spatial variation in all explored characteristics was more or less the same.

Of course, the proportion of significant tests is a very rough measure of the frequency of detectable effects associated with pollution, because this proportion depends on both the accepted level of the significance and on the statistical power of the tests (Section 8.4.3). Still, we were able to detect pollution 'signal' in only about one fourth of situations when the between-plot variation was significant. Furthermore, the detected signal was rather small: on average, only about 15% of variability among our study sites, which always included both most and least polluted habitats (Figs. 2.2–2.19), was explained by pollution load. This conclusion further stresses the need to properly account for variation in the characteristics under study in order to partition pollution effects on biota from variation caused by other factors.

There is no doubt that regional increases in pollutant concentrations are generally below the levels of environmental contamination observed near industrial polluters. Since the localised and relatively strong environmental contamination generally causes biotic effects of low magnitude, accurate identification of the consequences of regional pollution may require more effort than commonly thought. Moreover, the low signal to noise ratio increases the probability of erroneous attribution of natural spatial or temporal variation to pollution. Properly replicated sampling design, selection of adequate temporal and spatial scales, careful use of statistics and exploration of causal relationships behind the observed effects are essential to overcome this problem.

8.4.3   Power of Correlation Analyses

Data collected around the industrial polluters are most frequently analysed by calculating correlations between the measured parameters and either distance from the polluter or concentration of some pollutant. In a random sample of 1,000 publications from our collection, correlation analysis was used in 32% of studies properly reporting the use of statistical methods. Together with publications that employed regression analysis (15%), they comprise nearly a half of all studies.

Analysis of 1,446 individual correlation coefficients calculated from our data (Chapters 3–7) demonstrated that the average absolute value of correlation was relatively small (mean ± S.E.: 0.395 ± 0.006), corresponding to the effect size (z-transformed correlation coefficient) of 0.42. According to classification by Cohen (1988), our sample included 24.9% of small effects (≤0.20), 24.3% of medium effects (0.20–0.40) and 50.8% of large effects (>0.40).

Assuming that the effect size of 0.4 is sufficiently representative for correlations between the measures of pollution load and biotic variables, we conclude that the statistical power (α = 0.05, one-tailed test) of each individual correlation analysis

based on ten study sites was on average only 34%. In other words, two thirds of the individual analyses committed the Type II statistical error, i.e., acceptance of the null hypothesis (variation of measured variable is independent from variation in pollution load), while in fact H1 (variation of measured variable is associated with variation in pollution load) was true.

The average statistical power of 40–47% (to detect a medium effect size) of tests published in behavioural journals was considered 'distressingly low' by Jennions and Møller (2003). The situation in pollution ecology is even worse. The median number of study sites in a random sample of papers that have used correlation analysis to detect the biotic effects of pollution was seven. Thus, on average only 25% of the published tests had a chance to detect an effect size of 0.4. Power of these published tests was only sufficient (80%) to detect very large effects ($|r| \geq 0.83$, corresponding to $z_r = 1.19$), which are extremely infrequent (5% of our sample).

The most straightforward solution to this problem is to increase sample sizes. To increase the statistical power of correlation analysis to the recommended level of 80% (to detect an effect size of 0.4), the number of study sites should be at least 36. It may be argued (Jennions & Møller 2003) that logistic, ethical, conservation, and financial constraints make this impossible: only 7.2% of 1,000 publications from our database reported data sampling from 38 or more study sites (Fig. 8.1). Moreover, since many researchers tend to design their studies by analogy with previously published work, it seems unlikely that the number of study sites in field studies addressing pollution effects on biota will increase rapidly.

The low statistical power of tests based on small samples demonstrates the need to use meta-analysis for generalization of the accumulated data. Although this need was recognised long ago (Armentano & Bennett 1992), little was done to fully interpret the literature and ascertain the likelihood of trends common to ecosystems and pollutant regimes (Ruotsalainen & Kozlov 2006; Zvereva et al. 2008; Zvereva & Kozlov 2009; Roitto et al. 2009). The experience of medical sciences demonstrated that meta-analysis of numerous small-scale studies (only few of which are likely to detect significant trends) may provide a more cost-effective way of assessing the value of a treatment than investing in a few large-scale studies (Song et al. 2000).

8.4.4   Correlation with Pollution or Correlation with Distance?

It has been frequently argued that comparisons of the results from the literature are complicated by varying methods and objectives (Armentano & Bennett 1992; Glasziou & Sanders 2002). One of specific sources of variation in methodology of pollution-oriented studies is the choice of explanatory variables for correlation analyses. Authors of primary studies correlate their results with a variety of parameters, including distance from the polluter and concentrations of different pollutants in ambient air, snow, soils, or different organisms. Some authors constructs so-called 'toxicity indices', usually computed as linear combinations of absolute or relative concentrations of several pollutants (Rühling & Tyler 1973; Vorobeichik et al.

1994; Mumtaz et al. 1997; Mowat & Bundy 2002; Simonsen et al. 2008). Consequently, the validity of a meta-analysis that combines studies based on correlations with different explanatory variables can potentially be questioned.

Averaging the absolute values of Pearson's linear coefficients across the entire study (Chapters 3–7) demonstrated that correlations with both log-transformed distance from the polluter and foliar concentration of one of the pollutants yielded apparently the same overall conclusions (z-transformed values: $F_{1,\ 1444} = 0.00$, $P = 0.95$). This is not surprising, since concentrations of all pollutants in all media (ambient air, soils, plant and animal tissues) decrease proportionally with increasing distance from the smelter (Freedman & Hutchinson 1980a; Barkan 1993; Ruohomäki et al. 1996; Kozlov 2005a). Thus, the selection of the measure of the pollution load depends primarily on the study goals. If the study aims to demonstrate the effect, then distance from the polluter may be the best proxy of pollution load. Even in medical studies, population exposure to pollution is often estimated on the basis of distance to pollution source (Gottlieb et al. 1982; Biggeri et al. 1996; Monge-Corella et al. 2008). Of course there are numerous exceptions related to local orography and meteorology (Vorobeichik et al. 1994; Kozlov et al. 1995), which may be critically important for case studies, but are of relatively little value for meta-analyses. Moreover, exploration of dose–response relationships requires the use of pollutant concentration (usually log-transformed) instead of distance.

Emissions of any polluter consist of dozens of substances, many of which are toxic. Since it is impossible to attribute the effects observed in the course of field studies to any of the individual pollutants, the best solution is to correlate the results with one of the 'main' pollutants, which may serve as an indicator of pollution load. Concentrations of metals, fluorine, or sulphur dioxide have been the most frequently used indicators in published primary studies. Importantly, estimates of pollutant concentrations in both ambient air and plant foliage show substantial temporal variation (Vorobeichik et al. 1994; Kozlov et al. 1995; Kozlov 2005a). Therefore, we recommend using the concentrations of individual pollutants (but not the toxicity indices) in media that accumulate pollutants during months (snow) or decades (soils and litter).

8.4.5   Gradient Approach or the Planned Contrast?

Keeping in mind that the median number of study sites in a random sample of 1,000 published primary studies was five (Section 8.4.1), we checked whether a meta-analysis based on the contrast between the two most polluted and the two least polluted study sites (Hedge's d effect size) yielded the same conclusion as a meta-analysis based on z-transformed correlation coefficients calculated from ten study sites.

Three methods of data analysis (correlation with distance, correlation with pollution, and contrast between two most and two least polluted sites) after appropriate transformations produced similar effect sizes in all individual analyses (reported in Chapters 3–7) and in a pooled data set ($F_{2,\ 1824} = 1.36$, $P = 0.26$). However, this

result should not be interpreted as an excuse for using low number of study sites: although the average effect size was not affected, variance of each individual effect based on the contrast between the two most and two least polluted sites was on average seven times higher than for effect sizes based on ten study sites, thus affecting the significance of the overall effect and hampering identification of the parameters explaining variation in effect sizes.

8.4.6 Importance of Supplementary Information

Information on the polluters, as well as on their environmental impacts, was poorly reported in most of the primary studies used in meta-analyses of the effects caused by industrial pollution (Ruotsalainen & Kozlov 2006; Zvereva et al. 2008; Zvereva & Kozlov 2009; Roitto et al. 2009). As a result, the values of many explanatory variables necessary for our analyses had to be found in additional publications, on the Internet and by personal contacts with authors, regional authorities or company representatives. This was the most difficult and time-consuming part of data collection; still, in some cases we were unable to obtain the data of critical importance, which affected the number of studies involved in individual analyses.

Integrating our experiences over the course of this work, we suggest the *minimum minimorum* list of characteristics that need to be reported in each study addressing impact of industrial polluters on biota in order to allow its efficient use in subsequent comparative studies an\d meta-analyses.

Emission source
    General information
        Name of the polluter
        Type of the polluter (e.g., copper smelter, coal-burning power plant)
    Geographical information
        Country, administrative region within the country
        Position relative to large settlement(s)
        Latitude, longitude, and altitude above sea level
    Historical information
        The year of establishment (beginning of pollution impact)
    Chemical information
        List of most important pollutants
        Amount of emissions during the study year(s)
Impact zone
    Geographical information
        Presence of other point polluters with similar or larger environmental impacts
        within the impact zone of the polluter under study
    Landscape information
        Biome
        Landscape characteristics (plain, hilly, river valley, rocky mountain slopes)

Ecological information
    Pre-industrial vegetation (type, physiognomy)
    Soil type and basic characteristics (including pre-industrial pH of topsoil)
Historical information
    Brief history of environmental impact
    References to most important studies reporting pollution loads and environmental effects
Chemical information
    Spatial pattern of pollutant distribution (maximum concentrations, size and shape of the contaminated area)

8.5   Research and Publication Biases

Bias is a term used to describe a tendency or preference towards a particular perspective, ideology or result, especially when the tendency interferes with the ability to be impartial, unprejudiced, or objective. Investigation of biases and prevention of their impacts on general conclusions is of specific importance for meta-analyses (Begg 1994; Palmer 1999; Murtaugh 2002; Leimu & Koricheva 2004; Delgado-Rodríguez 2006; Formann 2008).

The accuracy of conclusions made on the basis of meta-analyses may suffer from the research bias - the tendency to perform experiments on organisms or under conditions in which one has a reasonable expectation of detecting statistically significant effects (Gurevitch & Hedges 1999). For example, studies of herbivore populations are likely to start in the year and on the site when herbivore damage is apparent (Zvereva & Kozlov 2009). Similarly, studies of pollution impact on vegetation are generally initiated following the appearance of visible damage (Anderson 1966; Linzon 1966; Murtha 1972; Tikkanen & Niemelä 1995) and are often confined to sites where the damage is most prominent. Some of these impact studies were even lacking controls (Wong 1978; Banásová et al. 1987). On the other hand, the decline of forests near point polluters resulted in the exclusion of barren areas from forestry-oriented projects, reports of which (Wotton & Hogan 1981; Scale 1982) therefore provided no information on the consequences of the most severe impacts on vegetation. The non-random selection of a pair of 'typical' study sites (one polluted vs. one control site) tends to overestimate the effect through a bias towards intuitive results (Zvereva & Kozlov 2009). Last but not least, the majority of pollution studies originated from Europe and, to a lesser extent, from North America, while other regions remain almost unexplored (Ruotsalainen & Kozlov 2006; Zvereva et al. 2008; Zvereva & Kozlov 2009; Roitto et al. 2009). Especially critical for the generalization of the results is the absence of information on subtropical and tropical regions, housing the largest part of terrestrial biodiversity.

If the probability of publication depends on factors other than the quality of the research, then we may face the problem of publication bias. Studies that disagree with the prevailing trend will not be published, especially when they are based on relatively small samples (Light & Pillemer 1984; Begg 1994), or their publication

will be delayed (Møller & Jennions 2001), or they will appear in less visible journals (Leimu & Koricheva 2004). Preferential publication of studies with significant outcomes that confirm the general paradigm may result in a correlation between effect size and sample size. In our meta-analyses of the published results, effect sizes were generally independent of sample size (Ruotsalainen & Kozlov 2006; Zvereva et al. 2008; Zvereva & Kozlov 2009), except for studies on plant fluctuating asymmetry (Section 5.4.3). We also found bias in reporting the data when quantitative information was provided only for species fitting the research hypothesis (density increase), while species that did not show density changes were only briefly mentioned (Selikhovkin 1986; Shelukho 2002).

Due to the sensitivity of the research topic, some studies may remain unpublished due to efforts by industry to protect itself from unwanted examination of its impact on the environment and surrounding population (discussed by Moffatt et al. 2000). The best known example is the governmental policy of the former USSR, where publication of 'negative' information was prohibited for decades (Komarov 1978). We are aware of several recent examples, based on financial rather than legal restrictions. The experience of epidemiological research showed that opposite situations may also well exist, when negative or inconclusive evidence is likely to be discounted (Balshem 1991; Moffatt et al. 2000). This source of dissemination bias deserves special investigation.

The studies included in ISI databases reported stronger negative effects of pollution on diversity of vascular plants (Zvereva et al. 2008) and abundance of herbivorous insects (Zvereva & Kozlov 2009) but weaker effects on plant growth and reproduction (Roitto et al. 2009) than other publications. The discovery of this bias emphasises the need to account for studies published in less visible data sources (book chapters, conference proceedings, journals published in national languages, and 'grey' literature) in order to obtain less biased estimates of effect size in meta-analytical research syntheses (Murtaugh 2002).

The temporal trend in the magnitude of the reported effects is a general phenomenon in ecology (Jennions & Møller 2002; Leimu & Koricheva 2004). We detected a decrease in the estimates of pollution effects on herbivore population density from 1965–1989 to 1990–2008 (Zvereva & Kozlov 2009), which we tentatively attribute to the improvement in research methodology. The hypothesis on the increase in herbivory under pollution impact came through developmental stages that are typical for any ecological hypothesis (Leimu & Koricheva 2004): from supportive evidence of the newly formulated hypothesis to the accumulation of disconfirming evidence. This accumulation of non-supporting evidence during the period of most intensive studies in the 1990s resulted in a shift from highly significant positive effects to non-significant effects and caused doubts about the generality of this phenomenon. Meta-analysis of published data contributes to the third stage in the development of the hypothesis - the restriction of its scope by discovering sources of variation in herbivore responses to pollution (Zvereva & Kozlov 2009).

Comparison of original data (Chapters 3–7) with the outcomes of earlier studies (Table 8.1) demonstrated that published data generally overestimate (by a factor of 5 on average) the magnitude of the effects of industrial pollution on terrestrial biota.

Table 8.1 Comparison between mean effects sizes based on published studies and on the original data

Character	Published data			Original data		
	r	N	Reference	r	N	Reference
Leaf/needle size	−0.48	204	Roitto et al. 2009	−0.11	88	Fig. 4.7
Shoot length	−0.47	164	Roitto et al. 2009	−0.14	111	Fig. 4.13
Radial increment	−0.59	40	Roitto et al. 2009	−0.35	10	Fig. 4.19
Plant fluctuating asymmetry	0.31	25	Section 5.4.3, Table 5.1	0.00	61	Fig. 5.5
Plant diversity	−0.64	45	Zvereva et al. 2008	−0.33	12	Fig. 6.9
Density of insect herbivores	0.44	142	Zvereva & Kozlov 2009	−0.06	45	Fig. 7.1
Plant damage by insect herbivores	0.64	17	Zvereva & Kozlov 2009	−0.23	59	Fig. 7.4

All effect sizes were converted to correlation coefficients.

Furthermore, for both density of insect herbivores and plant damage by these insects, the original data indicated either no effect or a slight decrease with pollution, while published data reported strong increase with pollution (Table 8.1). Of course, it can be argued that we collected our data in the 2000s, when emissions from industrial enterprises were generally lower than in previous decades. However, we think that an overall decline of emissions explains only a minor part of the detected differences between the published and original data. A substantial part of the effects observed near the point polluters is due to a large pool of pollutants (metals and fluorine-containing substances) accumulated in soils (Haidouti et al. 1993; Lyubashevsky et al. 1996; Giller et al. 1998; Nahmani & Lavelle 2002; Kozlov & Zvereva 2007a). Natural leaching of these pollutants will continue for decades or centuries before they approach pre-industrial levels (Tyler 1978; Barcan 2002a), and only the very first signs of vegetation recovery were observed around some of the polluters explored by our team (Eeva & Lehikoinen 2000; Chernenkova et al. 2001). Therefore, we tend to attribute the differences detected between the effect sizes calculated from published and original data (Table 8.1) not to emission decline but to the biases discussed above.

Since we revealed a marked diversity of pollution effects on terrestrial biota, we strongly encourage researchers and editors to publish results that are unexpected or seem strange. The occasional observation of truly surprising phenomena is the norm in ecology, not the exception (Doak et al. 2008). Elimination of these results at the pre-publication stage (decision not to submit the manuscript) or by reviewers (frequently due to disagreement with the prevailing paradigm) is likely to bias our estimates of overall effects, sometimes leading to the dominance of incorrect or exaggerated opinions (partially discussed in Section 9.2). Even more importantly, these 'outliers' are critically important for the exploration of the sources of variation in responses of organisms and ecosystems to pollution.

8.6   Summary

Further accumulation of reliable observational data remains of critical importance for the development of pollution ecology. Results of short-term experiments with non-adapted organisms in over-simplified laboratory environments should be inter- preted cautiously, since they are likely to overestimate adverse effects. Proper replication of sampling at all hierarchical levels, selection of adequate temporal and spatial scales, and careful use of statistics are the key factors assuring the quality of information on the effects of pollution on terrestrial biota. This information, sum- marised by meta-analyses or other appropriate procedures, is necessary for explora- tory research followed by the generation of specific hypotheses, which can be tested using field and laboratory experiments. In research syntheses, the utmost care should be taken to recognise biases affecting the outcomes of individual studies and mitigate their impacts on our knowledge.