Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Peak centiles of chlorpyrifos surface-water concentrations in the NAWQA and NASQAN programs

Paul L. Mosquin [a,*], Jeremy Aldworth [a,1], Nicholas N. Poletika [b,2]

[a] RTI International, 3040 East Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709-2194, USA
[b] Environmental Regulatory Sciences, Regulatory Sciences and Regulatory Affairs – Indianapolis, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

ARTICLE INFO

Article history:
Received 15 July 2014
Received in revised form
1 November 2014
Accepted 14 November 2014
Available online 25 November 2014

Keywords:
Chlorpyrifos
Water quality monitoring
Water sampling
Censored data
Survey statistics

ABSTRACT

We provide upper bound estimates for peak centiles of surface water chlorpyrifos concentration readings within spatial, temporal, and land-use domains of the United States Geological Survey (USGS) National Water-Quality Assessment (NAWQA) and National Stream Quality Accounting Network (NASQAN) programs. These datasets have large overall sample sizes but variable sampling frequencies and, for chlorpyrifos, extremely high levels of non-detections. Point and interval estimates are provided for the 90th, 95th, 99th, and the 99.9th centiles, given sufficient sample size. Overall upper bound estimates for the NAWQA program over the period 1992–2011 for the 90th, 95th, 99th, and 99.9th centiles are <0.005, 0.0066, 0.0214, and 0.0852 ug/L, respectively. The estimation method is based on a survey sampling approach, finding centiles of pooled data across aggregates of site-years. Although the population quantity estimated by a pooled data centile is not the easily interpretable average of population site-year centiles, we provide strong support that it bounds this average by a combination of theory, comparison of NAWQA aggregate and direct estimates, and using modeled populations.

1. Introduction

A good knowledge of peak yearly concentrations of pesticides in rivers and streams is important for the development of sound environmental and regulatory policy. If samples are collected on a daily basis, reasonable estimates of a yearly extreme will generally require a relatively large number of days sampled at a given monitoring site. For example, to estimate the 95th centile, it would be desirable to have at least a sample size of 20, given few assumptions. This frequency of sampling is generally cost-prohibitive for multiple-objective monitoring programs of broad spatial coverage, which must balance considerations of spatial (number of sites) and temporal (samples per year) coverage. They also often have the goal of estimating typical or average levels, requiring fewer days sampled for the same precision in comparison to extreme quantities. Nevertheless, these programs collect a large amount of data over many sites and years, and although the data may not support the estimation of peak concentrations at any specific site and year, it is of interest to ask if these

* *Corresponding author*. Tel.: +1 919 597 5134.
E-mail addresses: mosquin@rti.org (P.L. Mosquin), jaldworth@rti.org (J. Aldworth), npoletika@dow.com (N.N. Poletika).
[1] Tel.: +1 919 485 5779.
[2] Tel.: +1 317 337 3476.
http://dx.doi.org/10.1016/j.watres.2014.11.020
0043-1354/

existing datasets of infrequently monitored data can be used to say something useful about extreme values over domains of multiple sites.

In this study we consider a couple of such large national monitoring programs: the United States Geological Survey (USGS) National Water-Quality Assessment Program (NAWQA) and the National Stream Quality Accounting Network (NASQAN) programs. These programs have been continuously collecting data on a wide variety of analytes since 1991 in many of the streams, rivers, and ground-waters of the United States. For chlorpyrifos, there are now well over 20,000 surface water measurements collected at fixed-frequency sampling sites for the period 1991–2012.

Chlorpyrifos is an organophosphate insecticide with moderate toxicity to humans. Its use changed during the 1990's with important withdrawals from the market occurring in February 1997 with withdrawal from the direct application pet care product market, and in December 2001 with complete residential market withdrawal. Residential, commercial, and agricultural applications can find their way to surface water systems by rain- or irrigation-induced runoff and sewerage systems. Surface water chlorpyrifos concentration readings can provide for challenging statistical modeling because the majority of surface-water measurements fall near or below the laboratory limits of detection. Because of this, distributional assumptions are very strong for parametric modeling methods applied to such data.

In this study we use the existing NAWQA and NASQAN data to provide point and interval estimates for peak centiles (90th, 95th, 99th and 99.9th) of chlorpyrifos within spatial, temporal and land-use type domains. Estimation is performed using a survey sampling method, originally developed by Woodruff (1952). The method offers a number of advantages. It is non-parametric in that it uses only assigned survey weights, and so does not assume an underlying distributional model. It also does not use measured values below those of the estimated centile, an important advantage here since chlorpyrifos data is largely composed of non-detections. Finally, provided that data are collected from a probability design it provides (design) consistent estimates and intervals for centiles within each domain of interest. The method was previously applied by Mosquin et al. (2012) for estimation of upper centiles in targeted atrazine monitoring programs. The main challenges in applying the method are: i) the NAWQA and NASQAN data were not collected from a probability sample and ii) the estimation approach uses pooled data from aggregates of sites and so estimates a pooled-data target quantity. This target quantity differs from the ideal target quantity consisting of an average across sites of site-year centiles. We will address both of these challenges in this paper.

2. Materials and methods

2.1. Datasets and analysis domains

Chlorpyrifos surface water readings analyzed in this study were collected under the NAWQA and NASQAN monitoring programs. The NAWQA program has as its goal the long-term monitoring of the conditions of streams and rivers of the United States. The NASQAN program has similar monitoring goals, although it is specific to major river systems of the United States.

The NAWQA program[3] collects surface water samples from distinct river basins and aquifers of the United States, each known as "study units" (see Fig. 1). It has had two complete monitoring cycles, a first cycle from 1991 to 2001, and a second cycle from 2001 to 2012 where certain study units were reduced and combined. During the 1991 to 2001 cycle, monitoring was performed in 51 study units, while in 2001–2012 it was performed in 42 of original 51 units, the remaining 9 being discontinued. Within each cycle, study units were sampled every three years to rotate through all study units in a three year period. For example, during the 2001–2012 cycle 14 of the study units were monitored in each of 2001, 2004, and 2007. Due to budgetary constraints, the number of sites was further reduced from 145 to 84 in 2005, and then increased to 113 in 2007 but with some of these sites sampled less frequently (Crawford, 2006).

The NASQAN monitoring program[4] currently includes stations on thirteen large rivers on the coastal network, twenty rivers on the Mississippi-Atchafalaya River Basin, and five stations on the Ocean Action Plan Monitoring Network. Of these, four are coastal stations and the other is on the Mississippi River.

Given these available data it is important to ask what population might be defined by them. Although neither the study units nor the site-locations were probability sampled, the study units were chosen to be well-dispersed around the United States, representing a large percentage of its fresh-water use (Hirsch et al., 1988). Within study units, fixed station sites were chosen by USGS scientists according to four criteria (Hirsch et al., 1988):

1) at the mouths of major tributaries and at selected points on the main stem that account for a large portion of total basin runoff,
2) upstream and downstream from reservoirs, urban areas, agricultural drains, and other areas that significantly affect water quality,
3) on streams draining large areas that have relatively homogenous land use, and
4) near major public water-supply and other important water uses.

Our analysis population consists of all fixed-frequency daily measurements for the time period of interest for the selected fixed sampling sites. That our analysis population is representative of a population defined by the above criteria is not supported by a probability design. Instead, its representativeness relies on the careful selection and application of the USGS criteria at the time when the study units and site locations were selected.

For this analysis population, if all possible days were sampled at all sites then the full population would have been sampled. Instead, the sample was of days within years for

[3] http://water.usgs.gov/nawqa/.
[4] http://water.usgs.gov/nasqan/.