Geophysical Research Abstracts
Vol. 17, EGU2015-15629, 2015
EGU General Assembly 2015

Testing how voluntary participation requirements in an environmental study affect the planned random sample design outcomes: implications for the predictions of values and their uncertainty.

Louise Ander (1), Murray Lark (1), Pauline Smedley (1), Michael Watts (1), Elliott Hamilton (1), Tony Fletcher (2), Helen Crabbe (2), Rebecca Close (2), Mike Studden (2), and Giovanni Leonardi (2)

(1) Centre for Environmental Geochemistry, British Geological Survey, Keyworth, Nottingham, NG12 5GG, United Kingdom, (2) EnglandCentre for Radiation, Chemicals and Environmental Hazards (CRCE), Public Health England, Chilton, Didcot, Oxfordshire, OX11 0RQ, United Kingdom

Random sampling design is optimal in order to be able to assess outcomes, such as the mean of a given variable across an area. However, this optimal sampling design may be compromised to an unknown extent by unavoidable real-world factors: the extent to which the study design can still be considered random, and the influence this may have on the choice of appropriate statistical data analysis is examined in this work.
We take a study which relied on voluntary participation for the sampling of private water tap chemical composition in England, UK. This study was designed and implemented as a categorical, randomised study. The local geological classes were grouped into 10 types, which were considered to be most important in likely effects on groundwater chemistry (the source of all the tap waters sampled). Locations of the users of private water supplies were made available to the study group from the Local Authority in the area. These were then assigned, based on location, to geological groups 1 to 10 and randomised within each group. However, the permission to collect samples then required active, voluntary participation by householders and thus, unlike many environmental studies, could not always follow the initial sample design.
Impediments to participation ranged from ‘willing but not available’ during the designated sampling period, to a lack of response to requests to sample (assumed to be wholly unwilling or unable to participate). Additionally, a small number of unplanned samples were collected via new participants making themselves known to the sampling teams, during the sampling period. Here we examine the impact this has on the ‘random’ nature of the resulting data distribution, by comparison with the non-participating known supplies. We consider the implications this has on choice of statistical analysis methods to predict values and uncertainty at un-sampled locations.