# Annex 9

## Power analysis of CS data

This power analysis of CS data analyses species richness and Grime R scores in Countryside Survey X plots (randomly located within the CS 1km square) and arable field margin plots (located within the crop field, but adjacent to the boundary). To investigate the power CS has to detect signs of change within GM and non-GM areas, only the plots located within the crop of interest in the CS2007 survey are used in this analysis. In CS2007 there were a total of 41 plots containing maize covering 21 unique squares and a total of 45 plots containing potato from 22 unique squares. Few CS squares contain any sugar beat and therefore it was deemed unsuitable for Countryside Survey to comment on this. We therefore focus the remainder of this analysis on maize and potato. We start by considering power to detect changes in species richness.

### Spatial Analysis

The first analysis is purely a spatial one, where the difference between GM to non-GM plots is examined within the same survey – CS2007. Under the different uptake and change scenarios, a proportion, equal to the uptake scenario, of CS plots containing the crop of interest were randomly selected and the observed species richness score associated with these plots changed by a factor equal to the change scenarios. The resulting data was then modelled using the model below and the significance of the indicator term corresponding to GM areas was stored. This was then repeated 100 times to obtain a percentage of times we observe a significant effect of the GM indicator term.

The statistical model used in the analysis is a log-linear model with Poisson error distribution (as species richness is a count variable) and a random effect accounting for differing levels of variation between CS squares to within CS squares. The model is given by:

$$\ln(S_{C,i,j}) = \mu_C + \alpha G_{C,i,j} + V_{C,i} + \varepsilon_{C,i,j}$$

where $S_{C,i,j}$ is the species richness in plot $j$ within square $i$ containing crop $C$, $\mu_C$ is the mean species richness for plots containing crop $C$, $G_{C,i,j}$ is an indicator variable taking value of 1 if plot $j$ containing crop $C$ in square $i$ is in GM and 0 otherwise, $\alpha$ is the affect that GM has on species richness, $V_{C,i}$ is a random effect for the $i$th square for crop $C$ and $\varepsilon_{C,i,j}$ is a random effect for the $j$th plot in square $i$ containing crop $C$.

Note that no temporal component is included here as we are not modelling change, merely a difference between two strata – GM and non-GM.

The results of the power analysis to detect the effect of a difference in species richness between GM and non-GM plots are shown in Table 1.

**Table 1: Power to detect the effect of GM on species richness given the proportion of plots that adopted GM and the difference we expect between GM and non-GM. Also shown is the mean number of species lost per plot in GM given the expected effect (% difference) and the original mean species richness as estimated from CS2007.**

| MAIZE | Mean no. of species lost given non-GM mean of 13.54 | Uptake | | | |
|---|---|---|---|---|---|
| Difference | | 20% | 40% | 60% | 80% |
| -75% | 10 | 100 | 100 | 100 | 99 |
| -50% | 7 | 64 | 77 | 81 | 64 |
| -40% | 5 | 38 | 44 | 53 | 37 |
| -30% | 4 | 19 | 33 | 26 | 22 |
| -20% | 3 | 13 | 13 | 17 | 10 |

| POTATO | Mean no. of species lost given non-GM mean of 10.17 | Uptake | | | |
|---|---|---|---|---|---|
| Difference | | 20% | 40% | 60% | 80% |
| -75% | 8 | 100 | 100 | 100 | 100 |
| -50% | 5 | 81 | 92 | 97 | 82 |
| -40% | 4 | 48 | 68 | 67 | 50 |
| -30% | 3 | 27 | 37 | 38 | 32 |
| -20% | 2 | 12 | 21 | 11 | 14 |

For example, the highlighted figure of 77 means that if 40% of the 41 CS plots containing maize in 2007 were to adopt GM and the effect that GM crops had was for 50% less species on average, then within a single survey we would detect a statistically significant difference in the means of the two groups (GM to non-GM) 77% of the time.

Temporal analysis

The previous analysis looked at the difference between the two groups in one survey. We now focus our attention on detecting significant differences between species richness change between two consecutive surveys. For this we use data on plot level species richness from CS1998 and CS2007, which henceforth we refer to as survey 1 and survey 2 respectively for generality. The plots used here are the same 41 maize and 45 potato plots used in the spatial analysis with richness in the two surveys noted. Over the between survey period of 9 years, we have various possibilities on how uptake and change may have occurred. If our only scenario is that there has been an x% uptake over the period then this could have occurred at any time over the period for each individual plot. We therefore consider three possibilities of uptake over the period:

- All uptake occurred 1 year after survey 1

- All uptake occurred 1 year prior to survey 2
- Uptake occurred equally each year over the between survey period.

These were considered as they covered the two possible extremes of uptake over the period and the most even uptake scenario. For scenarios of change, we assumed that there would be an effect the year after GM uptake, but then there would be some recovery for 5 years following this. We take changes of 50%, 25%, 10% and 5% with respective recovery rates of 5%, 2%, 1% and 0.5% per year as our scenarios of change. So for example if uptake occurred in year 1, in year 2 we see a 50% decline in species richness, in year 3 we see a 5% increase on the year 2 species richness, in year 4 we see a 5% increase on the year 3 species richness and so on to year 8, where recovery fails and the year on year species richness remains constant.

The statistical model fitted for this temporal analysis is

$$\ln(S_{2,C,i}) = \mu_C + \alpha G_{C,i} + \ln(S_{1,C,i}) + V_{C,i} + \varepsilon_{C,i,j}$$

Where $S_{2,C,i}$ is the species richness for plot $i$ in crop $C$ in survey 2, $\mu_C$ is the mean species richness change over the period for plots containing crop $C$, $G_{C,i,j}$ is an indicator variable taking value of 1 if plot $j$ containing crop $C$ in square $i$ is in GM and 0 otherwise, $\alpha$ is the affect that GM has on species richness change, $V_{C,i}$ is a random effect for the $i$th square for crop $C$ and $\varepsilon_{C,i,j}$ is a random effect for the $j$th plot in square $i$ containing crop $C$.

Plots were randomly sampled and changed according to the uptake and change scenarios to obtain new values of species richness in survey 2 to compare to species richness values in survey 1. Again the significance of the GM term in the model was stored. This was repeated 100 times under each scenario for both crop types to achieve the power of detecting a GM signal in the data. Full results are displayed in table 2 displayed below.

Table 2: **Power to detect effect of GM on species richness change between two survey periods given the proportion of plots for each crop that adopted GM, at what point over the period uptake occurred, the difference we expect between GM and non-GM and the recovery rate. Also shown is the minimum mean number of species lost per plot in GM given the expected effect (% difference) over the whole period and the original mean species richness as estimated from CS1998. The minimum occurs when all uptake occurs the year after survey 1 and hence there is time for the plot to recover.**

## Maize

| Change after uptake | Recovery rate | Min. change over period | Min. mean no. of species lost given initial non-GM mean of 10.50 | UPTAKE | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | 20% | | | 40% | | | 60% | | | 80% | | |
| | | | | Year after survey1 | Year prior survey 2 | Even over period | Year after survey1 | Year prior survey 2 | Even over period | Year after survey1 | Year prior survey 2 | Even over period | Year after survey1 | Year prior survey 2 | Even over period |
| -50% | 5% | -36% | 4 | 19 | 29 | 19 | 34 | 62 | 52 | 28 | 59 | 53 | 20 | 47 | 54 |
| -25% | 2% | -17% | 2 | 4 | 11 | 7 | 6 | 14 | 14 | 11 | 19 | 16 | 12 | 14 | 19 |
| -10% | 1% | -5% | 1 | 1 | 5 | 6 | 6 | 5 | 6 | 8 | 5 | 7 | 5 | 7 | 7 |
| -5% | 0.5% | -3% | 0 | 5 | 3 | 5 | 5 | 4 | 6 | 6 | 7 | 5 | 4 | 6 | 4 |

## Potato

| Change after uptake | Recovery rate | Min. change over period | Min. mean no. of species lost given initial non-GM mean of 10.50 | 20% | | | 40% | | | 60% | | | 80% | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | Year after survey1 | Year prior survey 2 | Even over period | Year after survey1 | Year prior survey 2 | Even over period | Year after survey1 | Year prior survey 2 | Even over period | Year after survey1 | Year prior survey 2 | Even over period |
| -50% | 5% | -36% | 3 | 24 | 41 | 26 | 38 | 68 | 50 | 41 | 73 | 58 | 37 | 62 | 70 |
| -25% | 2% | -17% | 1 | 5 | 19 | 8 | 16 | 17 | 10 | 10 | 19 | 20 | 13 | 19 | 19 |
| -10% | 1% | -5% | 0 | 6 | 3 | 6 | 6 | 4 | 0 | 5 | 9 | 8 | 11 | 7 | 7 |
| -5% | 0.5% | -3% | 0 | 9 | 4 | 6 | 8 | 4 | 6 | 1 | 8 | 3 | 6 | 8 | 1 |

For example, the highlighted figure of 59% in yellow shows that if there was a 60% uptake in GM that happened the year prior to CS survey and the affect of GM was a 50% decline in species richness following the year of uptake with a 5% recovery rate for the following 5 years, we would be able to detect a significant effect of GM on the change over the period 59% of the time.

## Power to detect change in Grime R score: Ruderals

Having run the spatial and temporal analysis for the countryside survey species richness data, we repeated both analyses using the total Grime R score within each of the same plots containing maize and potato. The grime R score is a measure of the Ruderals in each plot. Ruderals are species that prosper in areas of high disturbance and low levels of stress. It therefore provides another measure of vegetation diversity within arable plots. The same statistical models were used as with the species richness data except we used a gamma error distribution as Grime R score is not a count variable like species richness. The same inferential procedure of sampling plots was applied to perform 100 tests and obtain the estimated power.

**Table 3: Power to detect effect of GM on Grime R score given the proportion of plots that adopted GM and the difference we expect between GM and non-GM. Also shown is the mean difference in Grime R score per plot in GM given the expected effect (% difference) and the original mean Grime R score as estimated from CS2007.**

| MAIZE | Decline in Grime R score given non-GM mean of 39.32 | Uptake | | | |
|---|---|---|---|---|---|
| Change | | 20% | 40% | 60% | 80% |
| -75% | 29.49 | 100 | 100 | 100 | 100 |
| -50% | 19.66 | 100 | 100 | 100 | 100 |
| -40% | 15.73 | 92 | 99 | 99 | 93 |
| -30% | 11.80 | 62 | 86 | 84 | 63 |
| -20% | 7.86 | 36 | 39 | 42 | 33 |

| POTATO | Decline in Grime R score given non-GM mean of 52.18 | Uptake | | | |
|---|---|---|---|---|---|
| Change | | 20% | 40% | 60% | 80% |
| -75% | 39.14 | 100 | 100 | 100 | 100 |
| -50% | 26.09 | 100 | 100 | 100 | 100 |
| -40% | 20.87 | 93 | 99 | 100 | 96 |
| -30% | 15.65 | 64 | 87 | 86 | 72 |
| -20% | 10.44 | 25 | 42 | 50 | 34 |

Table 4: **Power to detect effect of GM Grime R score change between two survey periods given the proportion of plots for each crop that adopted GM, at what point over the period uptake occurred, the difference we expect between GM and non-GM and the recovery rate. Also shown is the minimum mean decline in Grime R score per plot in GM given the expected effect (% difference) over the whole period and the original mean species richness as estimated from CS1998. The minimum occurs when all uptake occurs the year after survey 1 and hence there is time for the plot to recover.**

# Maize

| Change after uptake | Recovery rate | Min. change over period | Min mean decline in Grime R given initial non-GM mean of 34.45 | UPTAKE | | | | | | | | | | | |
| | | | | 20% | | | 40% | | | 60% | | | 80% | | |
| | | | | Year after survey1 | Year prior to survey 2 | Even over period | Year after survey1 | Year prior to survey 2 | Even over period | Year after survey1 | Year prior to survey 2 | Even over period | Year after survey1 | Year prior to survey 2 | Even over period |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -50% | 5% | -36% | 12 | 39 | 65 | 56 | 44 | 81 | 61 | 50 | 81 | 72 | 52 | 68 | 74 |
| -25% | 2% | -17% | 6 | 11 | 19 | 20 | 18 | 21 | 19 | 15 | 29 | 24 | 18 | 22 | 32 |
| -10% | 1% | -5% | 2 | 6 | 7 | 9 | 5 | 10 | 4 | 4 | 6 | 8 | 10 | 6 | 9 |
| -5% | 0.5% | -3% | 1 | 6 | 6 | 4 | 2 | 6 | 8 | 5 | 4 | 3 | 3 | 6 | 9 |

# Potato

| Change after uptake | Recovery rate | Min. change over period | Min mean decline in Grime R given initial non-GM mean of 34.10 | UPTAKE | | | | | | | | | | | |
| | | | | 20% | | | 40% | | | 60% | | | 80% | | |
| | | | | Year after survey1 | Year prior to survey 2 | Even over period | Year after survey1 | Year prior to survey 2 | Even over period | Year after survey1 | Year prior to survey 2 | Even over period | Year after survey1 | Year prior to survey 2 | Even over period |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -50% | 5% | -36% | 12 | 33 | 73 | 43 | 62 | 91 | 72 | 56 | 90 | 86 | 49 | 78 | 88 |
| -25% | 2% | -17% | 6 | 14 | 15 | 18 | 19 | 28 | 17 | 16 | 29 | 27 | 17 | 25 | 26 |
| -10% | 1% | -5% | 2 | 5 | 6 | 5 | 3 | 11 | 7 | 6 | 6 | 10 | 7 | 7 | 9 |
| -5% | 0.5% | -3% | 1 | 1 | 5 | 4 | 7 | 2 | 3 | 8 | 6 | 3 | 6 | 5 | 5 |

Square Level Spatial Analysis

The previous two analyses have used CS data at the plot level and the corresponding indicator of GM uptake for each particular plot. Due to confidentiality restrictions on field level GM schemes, we decided to repeat the spatial analysis of species richness but at a square rather than plot level and with two levels of information on GM uptake: known occurrence within the CS 1km square and known number of CS plots within square with GM uptake. The statistical model is largely unchanged, except the modelled data now corresponds to mean species richness in square $k$ containing crop $C$, represented by $S_{C,k}$ Specifically the model is given by,

$$\ln(S_{C,k}) = \mu_C + \alpha G_{C,k} + \varepsilon_{C,k}$$

and as we are modelling mean species richness within a CS square, no square level random effect is needed and we use a gamma error distribution as the response is no longer a count variable. Note that $G_{C,k}$ represents the information we have on GM uptake, which is either an indicator variable taking the value 1 if square $k$ contains any GM occurrence or it represents the number of plots within square $k$ with GM uptake.

Tables similar to those produced earlier are repeated for this square level analysis under the two GM information scenarios and are displayed below in tables 5a and 5b. As this analysis uses the same plots as the results in table 1, the two can be directly compared to see the effect on power the resolution of available GM information has.

Table 5: **Power to detect the effect of GM at a 1km square level on mean species richness given the proportion of plots that adopted GM and the difference we expect between GM and non-GM. This is displayed for when only occurrence of GM within the 1km square is known and for when the number of plots in GM is known.**

a) **MAIZE**

| Known occurrence of GM within 1km$^2$ | Uptake | | | |
|---|---|---|---|---|
| Change | 20% | 40% | 60% | 80% |
| -75% | 20 | 54 | 71 | NA |
| -50% | 11 | 18 | 39 | NA |
| -40% | 5 | 26 | 26 | NA |
| -30% | 2 | 10 | 25 | NA |
| -20% | 3 | 6 | 18 | NA |

| Known no. of CS plots in GM | Uptake | | | |
|---|---|---|---|---|
| Change | 20% | 40% | 60% | 80% |
| -75% | 20 | 66 | 91 | 77 |
| -50% | 9 | 22 | 37 | 37 |
| -40% | 5 | 22 | 24 | 29 |
| -30% | 6 | 11 | 17 | 10 |
| -20% | 9 | 8 | 14 | 12 |

b) **POTATO**

| Known occurrence of GM within 1km$^2$ | Uptake | | | |
|---|---|---|---|---|
| Change | 20% | 40% | 60% | 80% |
| -75% | 83 | 96 | 93 | NA |
| -50% | 32 | 71 | 72 | NA |
| -40% | 21 | 47 | 54 | NA |
| -30% | 11 | 33 | 35 | NA |
| -20% | 6 | 15 | 26 | NA |

| Known no. of CS plots in GM | Uptake | | | |
|---|---|---|---|---|
| Change | 20% | 40% | 60% | 80% |
| -75% | 93 | 99 | 99 | 97 |
| -50% | 46 | 69 | 80 | 72 |
| -40% | 33 | 38 | 58 | 50 |
| -30% | 16 | 29 | 31 | 33 |
| -20% | 9 | 16 | 21 | 17 |

Note that when we only have information at a square level of whether or not uptake has occurred and uptake is high (eg 80%), we may find that every square in our sample contains some GM uptake. If this is the case then we cannot perform a test of significance of the GM term in our model. In this instance the table displays an NA.