

# 2014 Global Soil Survey Report

Micah Woods, Larry Stowell, and Wendy Gelernter

12 September 2014

WE'VE BEEN LOOKING FORWARD to writing this first report<sup>1</sup> since we announced the Global Soil Survey<sup>2</sup> (GSS) in August 2013.

This is an exciting citizen science project, and it has been fun to review the data and analyze the results. Thank you to everyone who has taken an interest in this project and has submitted samples. As you read this report, you can compare your sample results to the summary of all the data. If you haven't participated in the survey, please consider doing so,<sup>3</sup> because this is an ongoing project. With a broader range of soil types and grass types in this dataset, the results will be even more useful to turfgrass managers around the world.

We have written this report with three objectives. First, to provide a summary of the first year results to those who have participated in the survey. Second, to describe and demonstrate how we use these data to calculate a nutrient guideline level. Third, to share the results of this project with the public.

## Year one summary: soil samples, grass species, and locations

THE FIRST SAMPLES from this project were analyzed in September 2013. This report includes a summary and analysis of the data from kits returned up to 31 August 2014. In total this is 28 kits, each with 3 samples from good performing turf. That comes to 84 soil samples from areas producing good performing turf at the time the samples were collected.

Although this is an analysis of *soil* nutrient levels, the objective is to identify the soil nutrient levels sufficient to produce good performing turf. Table 1 lists the grass species growing on the soils which were submitted in the first year of the GSS.

Samples submitted in the first year of the GSS come from four countries (Canada, Japan, Thailand, and USA) and one Crown dependency (Isle of Man). Figure 1 shows the approximate locations from which samples were collected.<sup>4</sup>

The samples submitted so far represent a broad range of grass species and climates and soil types – from British Columbia to Prince Edward Island, Seattle to Florida, Tennessee and Texas to Thailand and Tokyo.

## Data summary

TABLE 2 SUMMARIZES the data collected through 31 August 2014. The *n* column is the number of data points for that particular test

Micah Woods, Asian Turfgrass Center, Bangkok, Thailand. Larry Stowell and Wendy Gelernter, PACE Turf, San Diego, USA.

<sup>1</sup> The code to generate this report is hosted at [https://github.com/micahwoods/2014\\_gss\\_report](https://github.com/micahwoods/2014_gss_report). We wrote this as a .Rnw file, used **knitr** to weave this document, and compiled it in L<sup>A</sup>T<sub>E</sub>X using the `tufte-handout` document class.

<sup>2</sup> [http://www.paceturf.org/journal/global\\_soil\\_survey](http://www.paceturf.org/journal/global_soil_survey)

<sup>3</sup> One kit includes three complete soil nutrient analyses. You receive the soil test results and a report; we add the data from the tests to the GSS database. To purchase a kit go to [https://www.paceturf.org/shop/product/global\\_soil\\_survey\\_kit](https://www.paceturf.org/shop/product/global_soil_survey_kit)

---

Grass species
perennial ryegrass
kentucky bluegrass
creeping bentgrass
bent-poa
poa-bent
fescue-bent
fescue-poa
<i>Poa annua</i>
bent-poa-rye
rye-bluegrass-bent
bermudagrass
manilagrass ( <i>Zoysia matrella</i> )
japanese lawngrass ( <i>Z. japonica</i> )
seashore paspalum

---

Table 1: Samples were submitted from soils supporting good performing turf of the listed grass species in the first year of the GSS.

<sup>4</sup> Rather than marking the exact sample location, for privacy purposes we've just marked the map in the state, province, or prefecture from which the samples were collected.

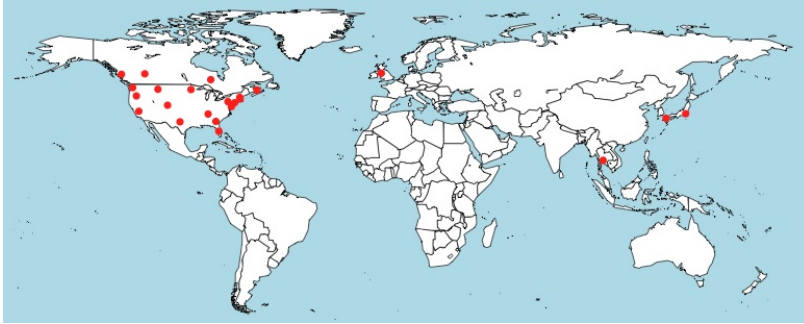


Figure 1: Red circles on this world map indicate the locations from which samples were submitted in year one of the Global Soil Survey.

parameter. Three calcareous samples have been dropped for the calcium analysis. We do not calculate a mean value for pH, because pH is on a logarithmic scale, and we do not calculate a GSS value for pH or for soil organic matter.

Soil parameter	n	Min	Median	Mean	Max	GSS
pH	84	5.2	6.5		8.2	
OM %	84	0.25	2.02	2.4	10.15	
K ppm	84	10	70	79	248	35
P ppm	84	6	76	84	450	22
Ca ppm	81	196	655	902	2844	267
Mg ppm	84	26	92.5	102	516	47
S ppm	84	5	15	17	91	8

Table 2: Summary of Global Soil Survey data from September 2013 through August 2014.

The GSS column shows the GSS MLSN guideline<sup>5</sup> calculated from the GSS dataset. We calculate the “official” MLSN guidelines from a combined dataset of thousands of soil test results.<sup>6</sup> With the GSS data, we are able to do two things. First, we can analyze the GSS data separately, as shown in Table 2, to find the range of soil conditions in which turf is performing well. These GSS data serve as an important dataset which we can use for comparison with the MLSN dataset. Second, we can add the GSS data to the larger MLSN dataset to increase the breadth of grass types and soil types included.

<sup>5</sup> More detail about the MLSN guidelines is in the next section; for a quick introduction see [http://www.paceturf.org/journal/minimum\\_level\\_for\\_sustainable\\_nutrition](http://www.paceturf.org/journal/minimum_level_for_sustainable_nutrition) and <https://www.facebook.com/mlsnturf>.

<sup>6</sup> The MLSN guidelines as of November 2014 are here: <https://scisoc.confex.com/scisoc/2014am/webprogram/Paper86244.html>

### How are the MLSN guidelines calculated?

How do we go about analyzing the soil test data to calculate an MLSN guideline? First, we take a vector of test results for the element of interest. In this case, let’s use the potassium (K) data from the GSS. Here is a vector<sup>7</sup> of all the K data, in ppm using the Mehlich 3 extractant.

[1] 81 121 129 30 119 56 73 107 30 92 82 85 53 49 49 60 117 [18] 116 71 72 92 106 72 48 49 67 60 66 47 52 80 57 214 23 [35] 42 36 53 85 108 74 66 75 112 158 122 49 69 59 66 67 61 [52] 84 53 75 92 36 29 134 248 108 45 32 25 71 57 55 103 10 [69] 19 81 103 27 49 65 41 121 95 82 58 129 49 192 181 196

<sup>7</sup> The numbers inside square brackets, such as [18], are not soil test data. Rather, they are an indicator of the count: “the next number is the first, or the next number is the 18<sup>th</sup>, or the 52<sup>nd</sup>, and so on.”

Those are the same data summarized in Table 2. The minimum value is 10, the mean is 79, and the maximum value is 248. A good way to get a better look at how those data are distributed is to look at them in a histogram (Figure 2). This bins the data, showing how many of the soil test K values are at a certain level.

From Figure 2, one can see that most of the soils have K in the 50 to 100 ppm range, and that there are few (6 samples out of 84) above 150 ppm. This disparity between the actual soil conditions in which good performing turf is being produced around the world, and conventional nutrient guidelines for turfgrass, is one of the main reasons we started the Global Soil Survey. Penn State University, for example, recommends K be maintained on golf course putting greens at 180 to 220 ppm.<sup>8</sup> As the GSS data for K show, turfgrass soils generally don't hold that much K. And it does not appear that such a high level of K is necessary to produce good turf, because all of the samples in the GSS dataset were producing good turf at the time of sample collection, even though 94% of the GSS samples have K less than 180 ppm.

Now we can look even more closely at the K data, trying to find a probability distribution that is close to the actual distribution of the data. We can plot the density over the histogram, as shown in Figure 3. If the K data came from a normal distribution, the normal curve for the GSS K data would approximate the density of the data. But when we plot the normal distribution based on these data (Figure 4), we can see that the normal distribution doesn't fit these data very well.

The log-logistic (Fisk) distribution fits the actual data much better, as shown in Figure 5. Now the line for the Fisk distribution is similar to that of the density.

Alright, so we've identified a probability distribution that somewhat approximates the actual data. When we think about that, that is quite something, because we are working with data that come from so many different soil types, in which so many grasses are growing, from so many different climates, and of course each site is being managed independently with varying types and rates of K fertilizer being applied.

Yet even with all these variables, we are able to identify a model that can describe the data, in terms of how likely it is that a sample from within this collection of soil samples will be above or below a certain level of soil K. This will be useful as we identify a MLSN guideline for K based on this model describing the data.

Now we move on to the step at which we identify the MLSN guideline level. We have a GSS dataset for K with 84 data points, distributed as shown in Figure 2 and approximated by a Fisk distribution as shown in Figure 5. These data can be looked at in another way too, as a cumulative distribution function, which shows for each level of  $x$ , which in this case is the soil test K level, what percentage of the samples are at or below that level. It is, essentially, a step function. We start at soil test level of 0 ppm, and at that level,

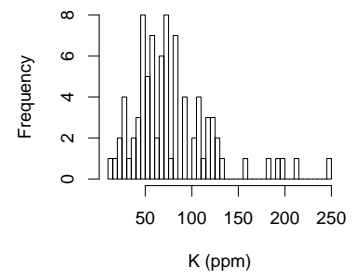


Figure 2: A histogram of the GSS K data.

<sup>8</sup> [http://bit.ly/psu\\_green](http://bit.ly/psu_green)

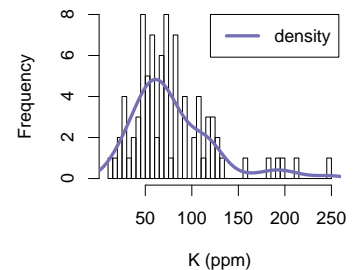


Figure 3: A histogram of the GSS K data with density curve.

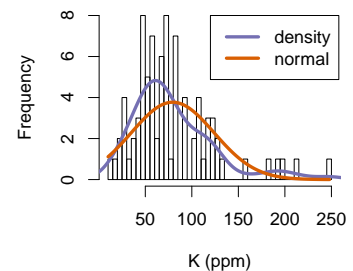


Figure 4: A histogram of the GSS K data with density and normal curves.

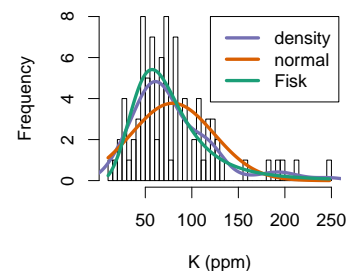


Figure 5: A histogram of the GSS K data with curves showing the density, a normal distribution, and a Fisk distribution.

there are 0 samples in the GSS dataset. Remember, we have a minimum value for K in this dataset of 10.

Then moving up, to 1 ppm, and to 2 ppm, all the way to the maximum of 248 ppm, at each level, we plot on the y-axis just what fraction of the samples that have occurred at or below that level. This is a cumulative distribution function, and it is plotted in Figure 6. What is plotted is the percentage of values at or below the value on the x-axis.

Now we are getting close to where the MLSN guideline comes from! Remember that we found a Fisk distribution is a pretty good fit for these data. A model from that distribution can approximate the distribution of the actual data that we have collected. And we can plot the cumulative distribution function of our model, which is shown in Figure 7. The model is almost right on top of the actual survey data. It's a good fit.

Once we have a model that is a good approximation of the data, we then identify the 0.1 level – that is, the level of  $x$  at which 10% of the data are at or below that level. In the case of K in the survey data so far, the value of  $x$  (K ppm) at which  $y = 0.1$  is 35. As shown in Table 2, the GSS MLSN guideline calculated for the GSS K data is 35.

#### *Why the 0.1 level?*

We have used the 0.1 level to identify the MLSN guideline.<sup>9</sup> The dataset doesn't include any bad samples; all the turf growing in the soils in the dataset was performing well at the time the samples were collected. So one could interpret the results as "any level at or above the minimum level within the dataset is enough to produce good turf." However, we like to be a bit conservative, so what we do is move up from the minimum, assuming that we want to have a bit of a buffer above that absolute minimum value, and the 0.1 level creates that buffer.

Going back to Table 2, you can see that the minimum level for an element is always less than the GSS value. For example, with K the minimum level was 10 ppm, but the guideline is 35. So we first buffer by saying that 10% of the samples, even though the turf growing in those soils is performing well, are for the purposes of our analysis, lower than we want. Then, we develop fertilizer recommendations that are based on keeping the soil levels of nutrients at or above the MLSN guideline. In this way, we can be confident that the guideline is safe. We know it is not the absolute minimum, and furthermore, the fertilizer recommendations are set to ensure the soil will remain at or above the guideline level.

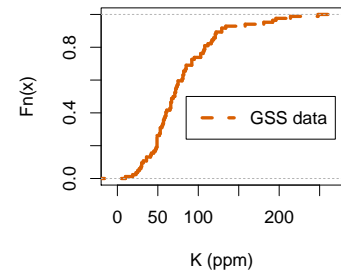


Figure 6: Cumulative distribution of GSS K data.

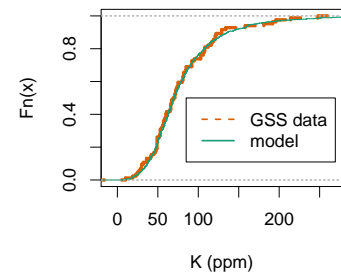


Figure 7: Cumulative distribution of GSS K data and cumulative distribution function of the model from the Fisk distribution fit to the survey data.

<sup>9</sup> And we have used a 0.1 level to calculate a GSS guideline as shown in Table 2.

## *Reasoning behind the Global Soil Survey*

THERE IS SO MUCH DATA out there, embedded, so to say, in the soils of turfgrass sites around the world. Could we put a system in place to collect some of these data, and analyze them, to provide results that would be of value to the entire industry? We envisioned a collaborative project between turfgrass managers and researchers, helping to develop, refine, and verify new nutrient guidelines. That became the Global Soil Survey.

As you have seen, we work with data from the soils of turfgrass sites from all over the world, and we collect more and more data month after month, and year after year. This project is set up to analyze these data, and to move the guidelines closer and closer to the most accurate values. If turf is not performing well, and the cause is a nutrient deficiency, then those soil data with nutrient deficient conditions will not be included in our data. Why? Because turf has to be performing well to be included. As the data change, so do the guidelines.

## *A comparison of MLSN and GSS results*

THE MLSN GUIDELINES have been updated since their first release in 2012 and we will be sharing these new guidelines over the next few months, including at the Crop Science Society of America annual meeting at Long Beach this November.<sup>10</sup>

We wanted to wait to update the MLSN guidelines until we had the first year of GSS data to include in the large dataset. We also wanted to make a specific comparison between the MLSN guideline in the large dataset and the MLSN guideline calculated from the GSS data only. The updated MLSN guidelines in ppm are K 37, P 21, Ca 331, Mg 47, and S 7. Compared to the 2012 version of the guidelines, K and P have gone up, and Ca, Mg, and S have gone down.

What if we calculate the guidelines just from the GSS data, which come from such a broad distribution of soil, grass, and geography? What are the numbers then? In ppm, we get K 35, P 22, Ca 267, Mg 47, and S 8. With the exception of Ca, which is much lower in the GSS dataset, the other elements<sup>11</sup> are remarkably similar. This comparison between the MLSN guidelines and the guidelines calculated only from the GSS data provides some reassurance that the MLSN guidelines are reasonable. As the GSS dataset increases in size, we will continue to make this comparison to ensure that the guidelines represent a reasonable value for each element.

<sup>10</sup> Abstract of our presentation entitled Only What the Turf Needs: Updating the Minimum Levels for Sustainable Nutrition (MLSN) Guidelines: <https://scisoc.confex.com/scisoc/2014am/webprogram/Paper86244.html>.

<sup>11</sup> What about micronutrients? For now, we are generating MLSN guidelines for the macronutrients (K and P) and the secondary nutrients (Ca, Mg, and S). We have some great data from our MLSN dataset and now from the GSS, so we will be looking into micronutrients using this same methodology as a future project.

*Just a bit more detail about the GSS data*

WE'VE SUMMARIZED the data in Table 2, and the data for K was shown in Figure 3 and other plots. Figure 8 summarizes the main data from the Global Soil Survey, excluding K.

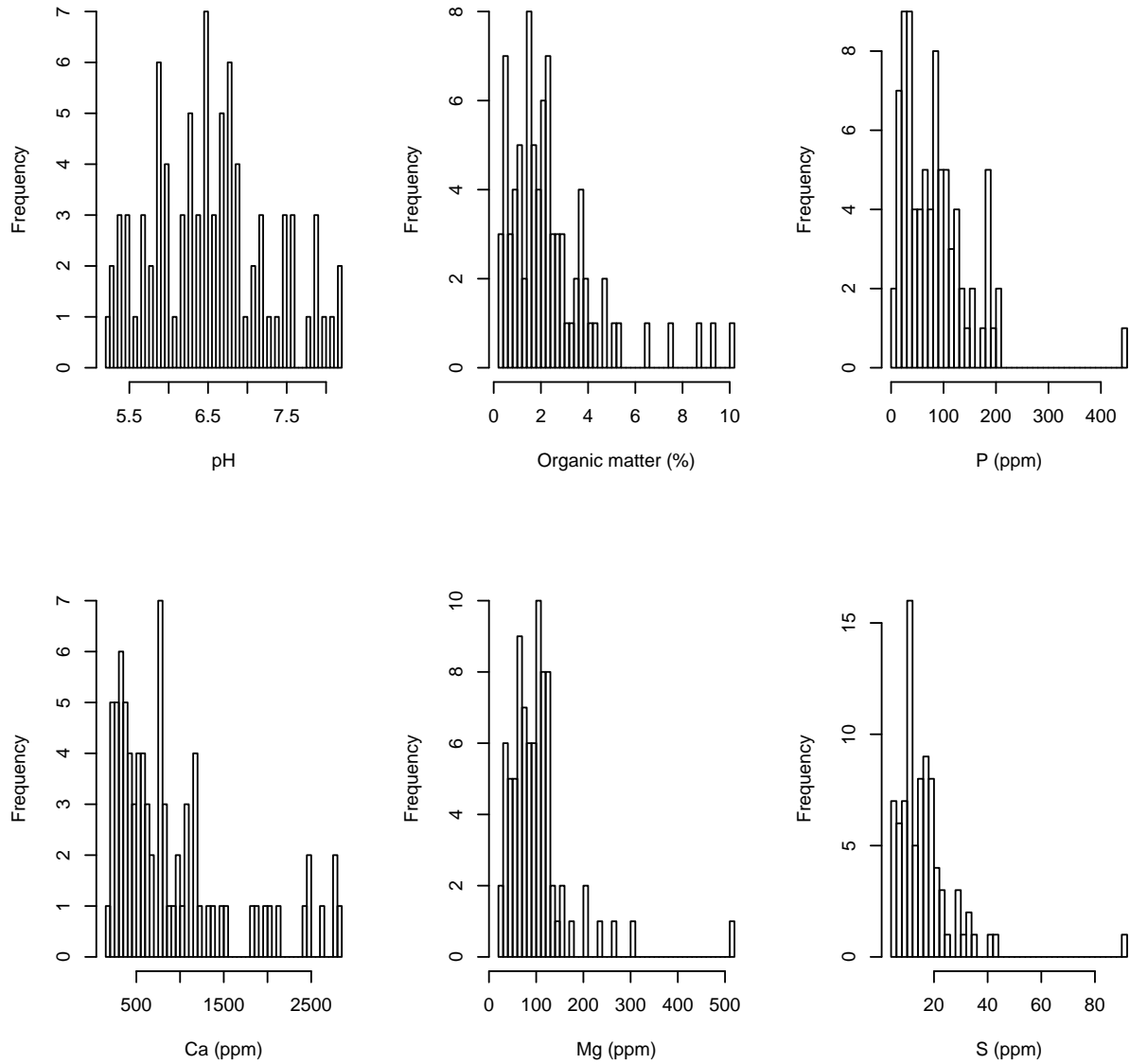


Figure 8: Histograms of Global Soil Survey data on pH, organic matter, and Mehlich 3 extractable P, Ca, Mg, and S.