Statistical Characterization of the Saturated Hydraulic Conductivity at the Plot Scale in Natural Grassy Soils

Flammini A.*, Morbidelli R., Corradini C., Saltalippi C. and Dari J.
[1]*Perugia University, via G. Duranti, Perugia, Italy*
**Corresponding author e-mail id: alessia.flammini@unipg.it*

Abstract: Spatially representative estimates of saturated hydraulic conductivity, K_s, are needed for simulating surface runoff and infiltration at different scales. This quantity is highly variable in space with values that can vary even more than two orders of magnitude at small catchment scale as well as at the plot or hillslope scale. The spatial variation of K_s is linked with both the basic soil properties and soil use and in addition with the presence of macropores and other preferential flow paths that are difficult to quantify. Therefore, the saturated hydraulic conductivity is considered a random variable depending also on random factors. From a theoretical point of view, the most recent areal infiltration approaches available in the literature and incorporated as components of many hydrological models require the assessment of the first two moments of the K_s probability density function. In principle, the estimate of these two parameters would imply the realization of a great number of local K_s measurements. On the other hand, many devices commonly used rely on the attainment of a steady-state flow rate, so that the time taken for each measurement is usually very large. In this context, it is important to determine in a given area the appropriate number of K_s measurements required for estimating the two moments. We previously performed an analysis for the first moment (the K_s mean) that is extended in this study to the K_s coefficient of variation linked with the second moment (the variance). On the basis of a dataset of 69 K_s observations on three grassy plots of a small Austrian catchment an analysis of uncertainty based on the non-parametric bootstrap method has been performed. It has allowed to estimate the 95% confidence interval around the coefficient of variation of K_s for different numbers of observations and different plot sizes. The outcomes have shown that the width of the normalized confidence interval obtained with a specific number of measurements is almost invariant with plot size while decreases with increasing the number of measurements in a specific area. The presented approach defines a methodology useful for determining the minimum number of measurements to be carried out in an area of specific dimensions to have a fixed level of uncertainty in the estimate of K_s coefficient of variation.

Keywords*:* Saturated hydraulic conductivity; Statistical characterization; Areal infiltration.

1. Introduction

The soil saturated hydraulic conductivity, K_s, has a primary role in the evolution of groundwater flow and pollutant transport through the vadose zone, as well as in the partitioning of rainfall into infiltration and overland flow at different spatial scale. Under natural conditions, it is characterized by a huge spatial heterogeneity with values that can vary even more than two orders of magnitude at small catchment scale (Sharma et al., 1987; Loague and Gardner, 1990; Baiemonte et al., 2017), as well as at plot and hillslope scale (Papanicolaou et al., 2015; Morbidelli et al., 2017; Picciafuoco et al., 2019). The spatial variability of K_s is linked with the basic soil properties as soil structure, texture, land cover, and management practices. In addition, a key role is played by the presence of macropores and other preferential flow paths that can influence the infiltration magnitude even more than the characteristics of the soil matrix. The last features cannot be quantified, therefore K_s should be represented involving a random component to be considered by a detailed spatial characterization. Recently proposed theoretical models for estimating expected areal infiltration rate and overland flow in vertically homogeneous and layered soils require the knowledge of the mean and coefficient of variation of the K_s probability density function (Corradini et al., 2011; Flammini et al., 2018). The assessment of these two quantities in principle implies the implementation of many K_s measurements at the point scale, that are typically of long duration because the devices commonly used rely upon the attainment of a steady infiltration rate observed even after a few hours. This indicates the difficulty of planning monitoring campaigns performed in space under the same conditions and with a proper density of measurements. Important insights to make these choices easier can be obtained by experimental investigations directed to define both control factors of the K_s statistical properties in a natural environment and the minimum number of measurements required for estimating the mean and coefficient of variation of K_s representative at the plot scale. Along this line, Picciafuoco et al. (2019) carried out a statistical analysis on 131 double-ring infiltrometer measurements performed in 12 plots of a small Austrian catchment. They found K_s slightly influenced by physical and topographical soil characteristics and strongly affected by land use. The highest values of K_s were obtained in arable fields, where, with respect to the ones of grassland areas, a much larger median and a coefficient of variation of about 75% were derived. In addition Picciafuoco et al. (2019), applying an uncertainty analysis based on the non-parametric bootstrap method (Carpenter and Bithell, 2000), determined the minimum number of K_s measurements necessary for estimating the mean of K_s with a specific accuracy for a given plot size. The objective of this paper is to extend the uncertainty analysis to the coefficient of variation of K_s to complete the investigation concerning the moments of the K_s probability density function to be determined for the application of areal infiltration modeling. The same dataset earlier analyzed by Picciafuoco et al. (2019) is used under the hypothesis of independent observations that allows the application of the non-parametric bootstrap method.

2. Study Area and Methods

The dataset used in this study is that earlier assembled by Picciafuoco et al. (2019) on the Hydrological Open Air Laboratory (HOAL) catchment, located in Petzenkirchen in the western part of Lower Austria, of area 0.66 Km^2 and elevation ranging from 268 to 323 m a.s.l. with a mean slope of 8%. Land cover consisted of arable land (87%), pasture (5%), forested area with grassland and high-stemmed vegetation of low density (6%), and paved surfaces (2%) (Blöschl et al., 2016). Pasture and forested area were collectively denoted as “grassland areas” because the forested area was really characterized by grassland, and not by brushwood, as a result of the patchy type of land use. Organic matter content, clay, silt and sand percentages at multiple depths were available on a 50 x 50 m grid. The topsoil in the catchment consisted (USDA soil classification) of silt loam (75% of the area), silty clay loam (20%), and silt (5%). A high-resolution digital terrain model (DTM) was used to derive elevation (el) and local slope angles (s) across the catchment. Little spatial variation between the measurement locations was observed for the physical and topographical soil characteristics. The maximum Coefficient of Variation (CV) observed for textural composition and organic matter content was about 15%, consistent with the USDA soil classification of the catchment according to which only two main soil types are identified in the basin topsoil.

Different crops were used in the catchment. The harvest of the winter crops occurred in July, allowing the access to the cultivated areas for infiltration measurements until the late-August/September.

The available measurements were performed, by double-ring infiltrometers using a 3x3 m grid, in 131 sites of 12 plots belonging to both grassland areas and areas devoted to agricultural practices. The main statistics associated with the total measurements on grassy and arable plots are summarized in Table 1, where the arithmetic mean of K_s and the corresponding coefficient of variation are shown. In the whole grassland areas the measurements were substantially completed in the period March 15 – April 7, while in the remaining ones were made later in the period between harvest and tillage The measurements in each plot were performed in a narrow period, up to 3 days for the larger plot.

The ANOVA method (Armstrong et al., 2000) was selected by Picciafuoco et al. (2019) for the analysis of variance of the experimental data associated to the different plots to understand whether the variability of K_s across the catchment was linked to specific soil physical characteristics of the measurement locations. An application of ANOVA method requires that two specific conditions on the experimental data subdivided into different groups are satisfied. Specifically, the random elements should be normally distributed and the variance should be the same in all groups, while the means can be variable from group to group. The available K_s dataset did not satisfy these two conditions, while log-transformed values of K_s were found appropriate for the ANOVA method application (see also Snedecor and Cochran, 1980).

The above approach was applied to different group ensembles: (1) ensemble obtained considering the plot locations as sources of variation to understand whether K_s variability is substantially linked to the specific physical and topographical characteristics of the measurement areas; (2) ensemble identified grouping areas with the same land use, considered because the outcomes of step (1) suggested a significant dependency of K_s variability on the plot location; (3) ensembles of plots characterized by the same land use to understand if (a) other plot-specific properties had a significant influence or (b) the observations could be considered as different sets sampled from the same population, even though collected at different locations in the catchment.

The results highlighted that the land cover had a major role, while the plot location influenced K_s only as an effect of the involved differences in the land use.

An uncertainty analysis was also performed using the confidence intervals to determine the minimum number of samples required for estimating the geometric mean of K_s for a selected area with a specific accuracy. In this context, the non-parametric bootstrap method based on the hypothesis of independent observations was used (see also Ahmed et al., 2015) for deriving 95% confidence intervals around the geometric mean of K_s for different numbers of observations. Finally, from the behavior of the confidence intervals with increasing K_s observations the appropriate number of measurements to be performed in the area of interest was taken out.

In this work, a similar procedure has been carried out to define the minimum number of local K_s measurements required for assessing the geometric coefficient of variation of K_s, $CV_g(K_s)$, for a given area, that is computed as:

$$CV_g(K_s) = \left(\frac{\sum_{i=1}^{n}(K_{s,i} - <K_s>)^2}{n-1}\right)^{0.5} \cdot \frac{1}{\overline{K_s}} \tag{1}$$

where n is the number of measurements in the specific plot, $K_{s,i}$ is the ith observation, $<K_s>$ is the plot geometric mean of K_s.

Table 1 – General statistics of the saturated hydraulic conductivity (K_s) values measured with double-ring infiltrometer: Min = minimum value; Max = maximum value; Mean = arithmetic mean; CV = coefficient of variation; Arable land = K_s values measured in arable areas; Grassland = K_s values measured in grassy areas; Total = all K_s values.

Statistics	Arable land	Grassland	Total
Min (mm h^{-1})	2.0	1.0	1.0
Max (mm h^{-1})	130.0	84.0	130.0
Mean (mm h^{-1})	46.9	20.2	25.1
St. dev. (mm h^{-1})	35.5	20.3	25.8
CV (%)	75.6	100.3	102.7

3. Results and Discussion

The aforementioned statistical analysis for the minimum observations to be made for estimating $CV_g(K_s)$ has been performed using the dataset available for grassland plots 2 (see also Figure 1), 11-12 and 7 of area 500 m^2, 200 m^2 and 80 m^2, respectively, where the measurements had the same spatial resolution with a density of about 1 observations per 10 m^2. The choice to limit our investigation to these plots with grassland is linked to the necessity to have extended areas with a significant number of measurements. Data available on arable plots do not satisfy this requirement. The main features associated with the measurements on the selected plots are synthesized in Table 2.

In spite of the limited spatial variation of the physical and topographical soil attributes, saturated hydraulic conductivity varies significantly, with values ranging almost over two orders of magnitude, from a minimum of 1 mm h^{-1} to a maximum of 84 mm h^{-1}.

A site-by-site examination reveals rather similar minima of saturated hydraulic conductivity on the selected plots, while the maxima vary from 48 to 84 mmh^{-1}. The coefficient of variation, $CV(K_s)$ approximately assumes values in the range 0.5-1.1 that fairly well reflects that associated to all the grassy plots of the catchment (Table 1). This suggests that the K_s spatial variability is still traceable in the observation plots individually considered.

In Table 2 also the geometric means, $<K_s>$, and the corresponding coefficients of variation $CV_g(K_s)$ are shown; it can be deduced that the geometric means assume values lower than the arithmetic means while the opposite occur for the coefficient of variation. The use in the uncertainty analysis of $<K_s>$ and $CV_g(K_s)$ as overall representation of the experimental data is supported by the consideration that the K_s dataset follows a lognormal probability density function.

(a)

(b)

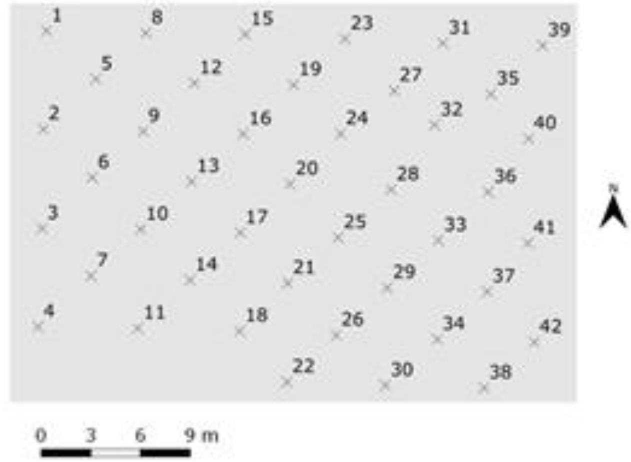

Figure 1: Picture of plot 2, the largest plot located in the proximity of the forested area of the catchment and characterized by a grassy meadow land (a). The correspondent scheme of performed measurements is also shown (b).

Table 2: General statistics the saturated hydraulic conductivity (K_s) values used in this study grouped by plot: n = number of observations in each plot; Min = minimum value; Max = maximum value; Mean = arithmetic mean; CV = coefficient of variation; $<K_s>$ = geometric mean; CV_g = geometric coefficient of variation.

Plot	2	7	11	12
n	40	9	10	10
Min (mm h^{-1})	1.0	3.0	3.0	2.0
Max (mm h^{-1})	78.0	48.0	54.0	84.0
Mean (mm h^{-1})	17.0	30.3	20.2	36.8
CV	1.09	0.50	0.98	0.72
$<K_s>$ (mm h^{-1})	9.1	24.0	11.5	22.7
CV_g	2.22	0.69	1.88	1.34

Representative outcomes of the uncertainty analysis obtained by the non-parametric bootstrap method are shown in the Figures 2-6. Figure 2 shows the 95% confidence interval of $CV_g(K_s)$ for the three different plots. Figure 3 shows explicitly the 95% interval confidence width that decreases with increasing the measurements number tending to an almost constant value and highlights that for a specific number of measurements, n, the plot with smaller area (7) seems to be in a smaller uncertainty. Figure 4 shows the 95% interval confidence of $CV_g(K_s)$ normalized by an average geometric coefficient of variation calculated through the sampling simulation involved in the bootstrap method. As it can be seen, in this case a rather similar behavior in terms of uncertainty is obtained for the three plots, as a result of the normalization procedure that involves different values of the average geometric coefficient of variation. This last outcome referred to $CV_g(K_s)$ is somewhat different from the corresponding one obtained for

<K_s> by Piacciafuoco et al. (2019) who deduced a higher reliability of the geometric mean estimates for smaller areas after the normalization procedure.

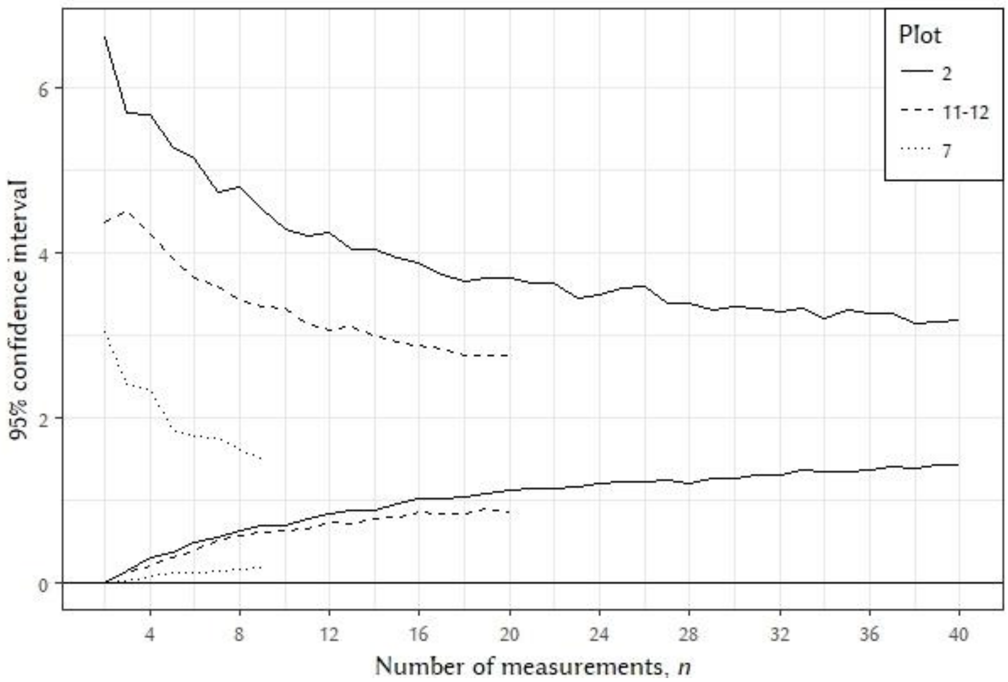

Figure 2: 95% confidence intervals of the K_s geometric coefficient of variation, $CV_g(K_s)$, obtained by the non-parametric bootstrap method for the considered grassy plots. K_s is the saturated hydraulic conductivity.

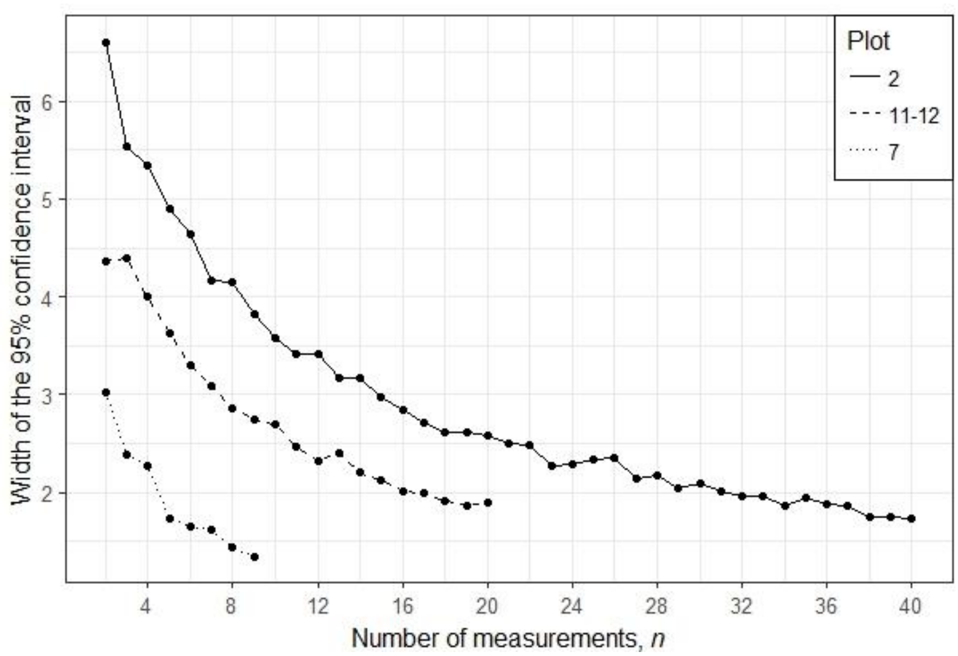

Figure 3: Width of the non-normalized 95% confidence interval of the K_s geometric coefficient of variation, $CV_g(K_s)$, as a function of the number of measurements, n. K_s is the saturated hydraulic conductivity.

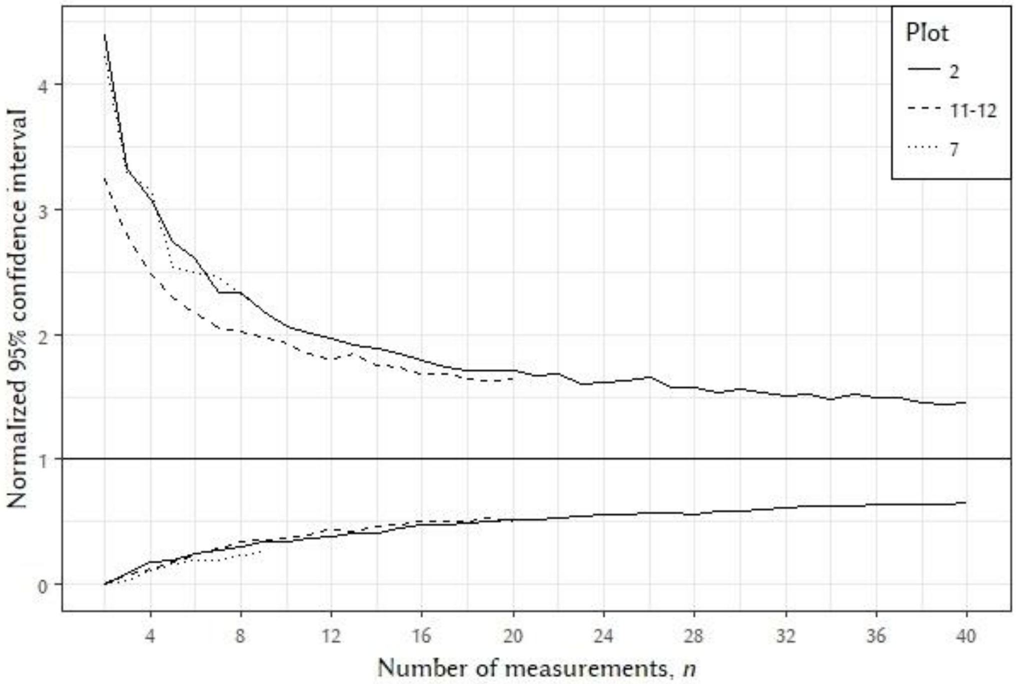

Figure 4: Normalized 95% confidence intervals of the K_s geometric coefficient of variation, $CV_g(K_s)$, obtained by the non-parametric bootstrap method for the considered grassy plots. K_s is the saturated hydraulic conductivity.

Figure 5 shows the reduction of the 95% confidence interval width with increasing the number of measurements for the three plots. This reduction has been obtained as the difference between the width of the non-normalized confidence interval associated with n measurements and the width associated with n-1 measurements. On each plot it is larger for small n while tends asymptotically towards zero for n increasing. This is very clear for the plot 2 interpolated curve. A similar behavior is suggested by the trend of the interpolated curves for the smaller two plots even though the reduced number of measurements does not allow a complete evidence.
Figure 5 provides information on the benefit – in terms of width reduction – gained by performing one extra measurement on each plot. It is high for small n and decreases with plot extent up to a point, specific for each plot, after which becomes negligible. Therefore, the n associated with this point can be considered as the minimum number of measurements necessary to enter the "zone" where the confidence interval is stable. For example, the interpolated curve of plot 2 is almost constant for n greater than 12.
Finally, Figure 6 shows the results of a further uncertainty analysis performed using plot 2 to determine for different plot areas the minimum number of measurements requested to derive $CV_g(K_s)$ with a specific accuracy. This plot has been chosen because the one with the largest dataset. Ten sub-areas of 55, 110, 165, 215, 280, 345, 405, 470, 535 and 595 m^2 have been considered, with larger sub-areas always containing all the smaller ones. In each sub-area, a

number of measurements ranging from 4 to 16 has been drawn, if any, and the normalized 95% confidence intervals of $CV_g(K_s)$ have been estimated through the non-parametric bootstrap method for a total of 79 combinations area - sample number. Finally, the widths of the normalized 95% confidence intervals have been plotted against area for different sample numbers. The resulting curves suggest how the confidence in the estimation of $CV_g(K_s)$ appears invariant when an equal number of measurements is considered on plots of increasing dimensions, at least in the range of both number of measurements and plot area investigated. Furthermore, on a plot of specific dimensions, a better estimate of $CV_g(K_s)$ can be achieved by increasing the number of measurements in the area. For example, for a plot of 345 m^2 sampled at 5 locations the width of the confidence interval is about 2.5 times the average $CV_g(K_s)$, while 16 measurements on the same plot determine an interval width almost halved (about 1.3 times the average $CV_g(K_s)$), i.e. a reliability associated with the $CV_g(K_s)$ estimation about doubled. On the other hand, when the same number of samples is taken in bigger areas, the width of the confidence interval remains invariant.

Figure 6 is very useful to derive a minimum number of samples that have to be taken in an area of specific dimensions as a tradeoff between accuracy (or uncertainty) and time required for the measurements.

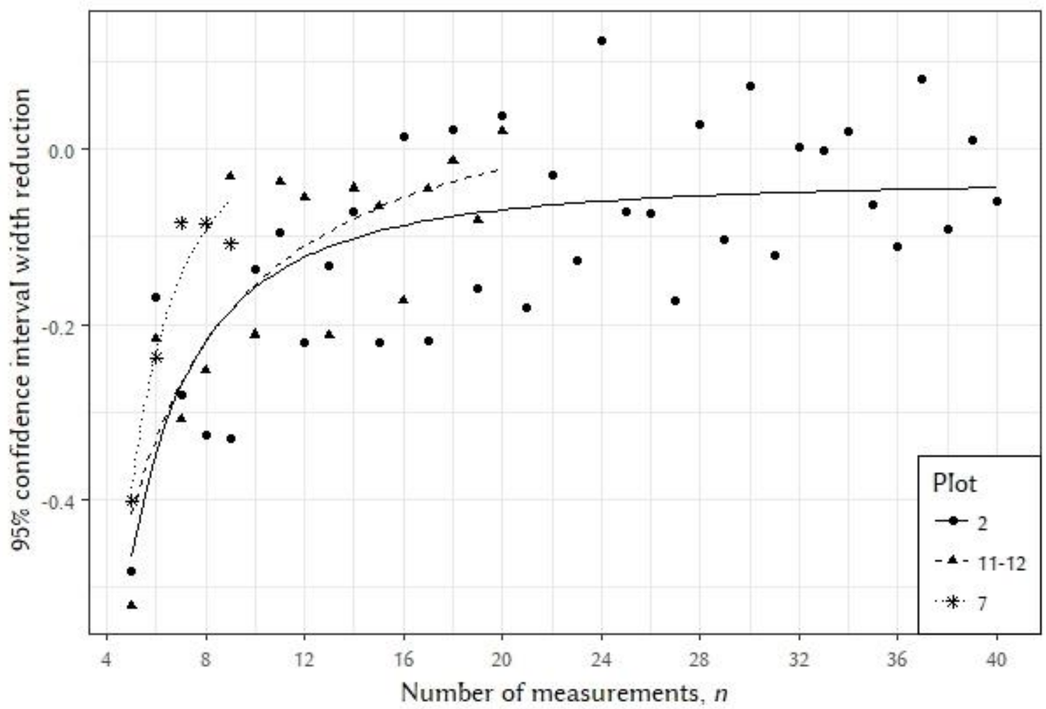

Figure 5: Reduction of the width of the non-normalized 95% confidence interval of the K_s geometric coefficient of variation, $CV_g(K_s)$, obtained using one extra measurement in the bootstrap method. For a specific number of measurements, n, the corresponding point represents the difference between the width associated with n measurements and with n-1 measurements. K_s is the saturated hydraulic conductivity.

4. Conclusions

From a dataset of 131 measurements of saturated hydraulic conductivity K_s earlier performed by double-ring infiltrometers in a small Austrian catchment over 12 plots belonging to both grassland areas and areas devoted to different agricultural practices (Picciafuoco et al., 2019), 69 observations from three grassy plots have been taken out. A statistical characterization of K_s at the plot scale has been made for studying problems linked with the assessment of its coefficient of variation. An analysis of uncertainty based on the non-parametric bootstrap method has been performed for estimating the 95% confidence interval around the coefficient of variation of K_s for different numbers of observations and different plot areas. The coefficient of variation $CV_g(K_s)$ is calculated considering the geometric mean $<K_s>$. The width of the normalized confidence interval obtained with a specific number of measurements is almost invariant with plot size while decreases with increasing the number of measurements in a specific area. These outcomes define a methodology useful for determining the minimum number of measurements to be performed in an area of specific dimensions to have a fixed level of uncertainty in $CV_g(K_s)$ estimate.
This study does not involve a specific analysis of spatial correlation of K_s measurements. If present the spatial correlation would have the effect of reducing the width of the confidence interval for a specific number of measurements in comparison with uncorrelated data. The same confidence in the estimates of $CV_g(K_s)$ would be therefore achieved with a smaller number of observations. In any case, the assumption of independent data allows to transfer to other geographical areas the results regardless of the plot correlation structure, even though they may represent an upper limit of the required sample density in the presence of correlation.

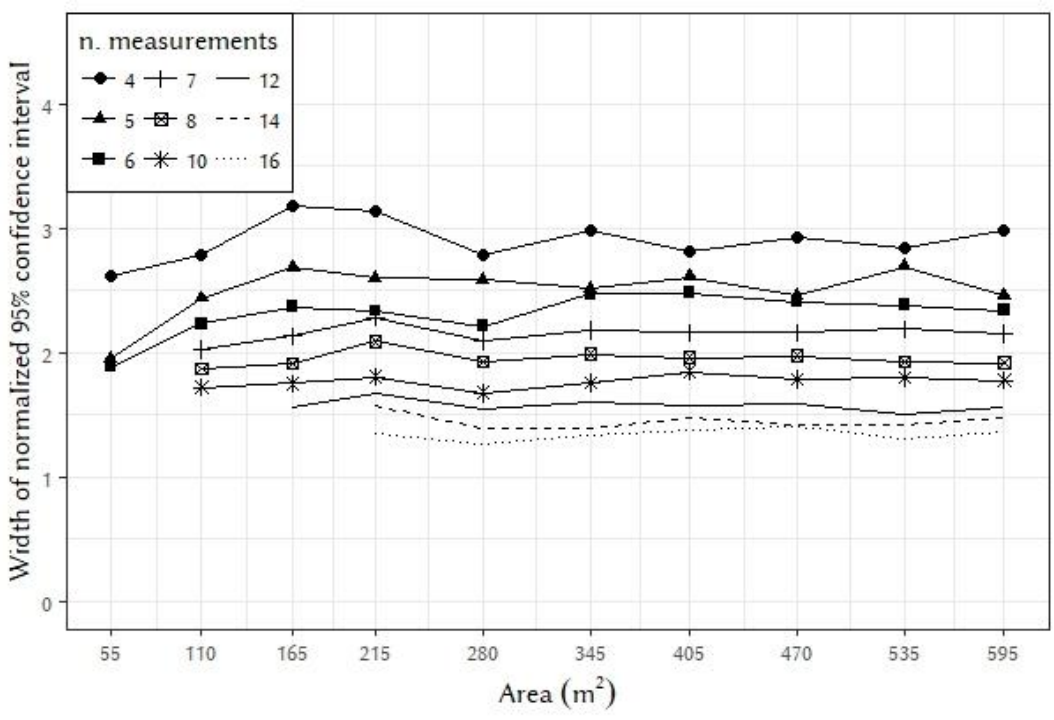

Figure 6: Width of the normalized 95% confidence intervals of the K_s geometric coefficient of variation, $CV_g(K_s)$, as a function of plot size. Lines represent the number of measurements used

for the evaluation of the confidence interval for different values of area. K_s is the saturated hydraulic conductivity.

Acknowledgements

The authors would like to acknowledge the Italian Ministry of Education, University and Research (PRIN 2015).

References

Ahmed, F., Gulliver, J., Nieber, J. (2015). Field infiltration measurements in grassed roadside drainage ditches: Spatial and temporal variability. Journal of Hydrology. 530, 604-611, https://doi.org/10.1016/j.jhydrol.2015.10.012.

Armstrong, R., Slade, S., Eperjesi, F. (2000). An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthalmic and Physiological Optics*, 20(3), 235-241, https://doi.org/10.1046/j.1475-1313.2000.00502.x.

Baiamonte, G., Bagarello, V., D'Asaro, F., Palmeri, V. (2017). Factors influencing point measurement of near-surface saturated soil hydraulic conductivity in a small sicilian basin. *Land Degradation & Development*, 28 (3), 970–982, https://doi.org/10.1002/ldr.2674.

Blöschl, G., Blaschke, A., Broer, M., Bucher, C., Carr, G., Chen, X., Eder, A., Exner-Kittridge, M., Farnleitner, A., Flores-Orozco, A., Haas, P., Hogan, P., Kazemi Amiri, A., Oismüller, M., Parajka, J., Silasari, R., Stadler, P., Strauss, P., Vreugdenhil, M., Wagrner, W., Zessner, M. (2016). The Hydrological Open Air Laboratory (HOAL) in Petzenkirchen: A hypothesis-driven observatory. *Hydrology and Earth System Sciences*, 20(1), 227-255, https://doi.org/10.5194/hess-20-227-2016.

Carpenter, J., Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19(9), 1141-1164.

Corradini, C., Flammini, A., Morbidelli, R., Govindaraju, R. (2011). A conceptual model for infiltration in two-layered soils with a more permeable upper layer: From local to field scale. *Journal of Hydrology*, 410(1-2), 62-72, https://doi.org/10.1016/j.jhydrol.2011.09.005.

Flammini, A., Morbidelli, R., Saltalippi, C., Picciafuofo, T., Corradini, C., Govindaraju, R.S. (2018). Reassessment of a semi-analytical field-scale infiltration model through experiments under natural rainfall events. *Journal of Hydrology*, 565, 835-845.

Loague, K., Gander, G.A. (1990). R-5 revisited, 1, Spatial variability of infiltration on a small rangeland catchment. *Water Resources Research*, 26(5), 957-971.

Morbidelli, R., Saltalippi, C., Flammini, A., Cifrodelli, M., Picciafuoco, T., Corradini, C., Govindaraju, R. (2017). In-Situ Measurements of Soil Saturated Hydraulic Conductivity: Assessment of Reliability through Rainfall-Runoff Experiments. *Hydrological Processes*, 31(17), 3084-3094, https://doi.org/10.1002/hyp.11247.

Papanicolaou, A., Elhakeem, M., Wilson, C.G., Lee Burras, C., West, L.T., Lin, H., Clark, B., Oneal, B.E. (2015). Spatial variability of saturated hydraulic conductivity at the hillslope scale: understanding the role of land management and erosional effect. *Geoderma*, 243–244, 58–68. ttps://doi.org/10.1016/j.geoderma.2014.12.010.

Picciafuoco, P., Morbidelli, R., Flammini, A., Saltalippi, C., Corradini, C., Strauss, P., Blöschl, G. (2019). On the estimation of spatially representative plot scale saturated hydraulic conductivity in an agricultural setting. *Journal of Hydrology*, 570, 106-117.

Sharma, M., Barron, R., Fernie, M. (1987). Areal distribution of infiltration parameters and some soil physical properties in lateritic catchments. *Journal of Hydrology*, 94(1-2), 109-127, https://doi.org/10.1016/0022-1694(87)90035-7.

Snedecor, G., Cochran, W. (1980). *Statistical Methods* (7th ed.). Iowa State University Press.