Digitizing archive atmospheric electric potential gradient data for scientific research

Levente Magos[1,2], Tamás Bozóki[1,3], István Bozsó[1,4], József Bór[1], András Horváth[1,5], Lukács Kuslits[1], Máté Timkó[1,4] and ***Attila Buzás[1,4]****

(1) Institute of Earth Physics and Space Science, Sopron, Hungary
(2) John von Neumann Faculty of Informatics, Óbuda University, Budapest, Hungary
(3) Doctoral School of Environmental Sciences, University of Szeged, Szeged, Hungary
(4) Doctoral School of Earth Sciences, Faculty of Science, Eötvös Loránd University, Budapest, Hungary
(5) Faculty of Information, Technology and Bionics, Péter Pázmány Catholic University, Budapest, Hungary

**** Contact author:***
buzas.attila@epss.hu

Introduction

- The **atmospheric electric potential gradient (PG)** has been measured in the **Széchenyi István Geophysical Observatory** in Nagycenk (NCK; near Sopron, Hungary) for more than six decades, since 1961.
- **Problem**: measurements were **recorded on photographic paper** because digital data acquisition was introduced late (in 2011). Digital PG data are **only available in the format of hourly averages**.

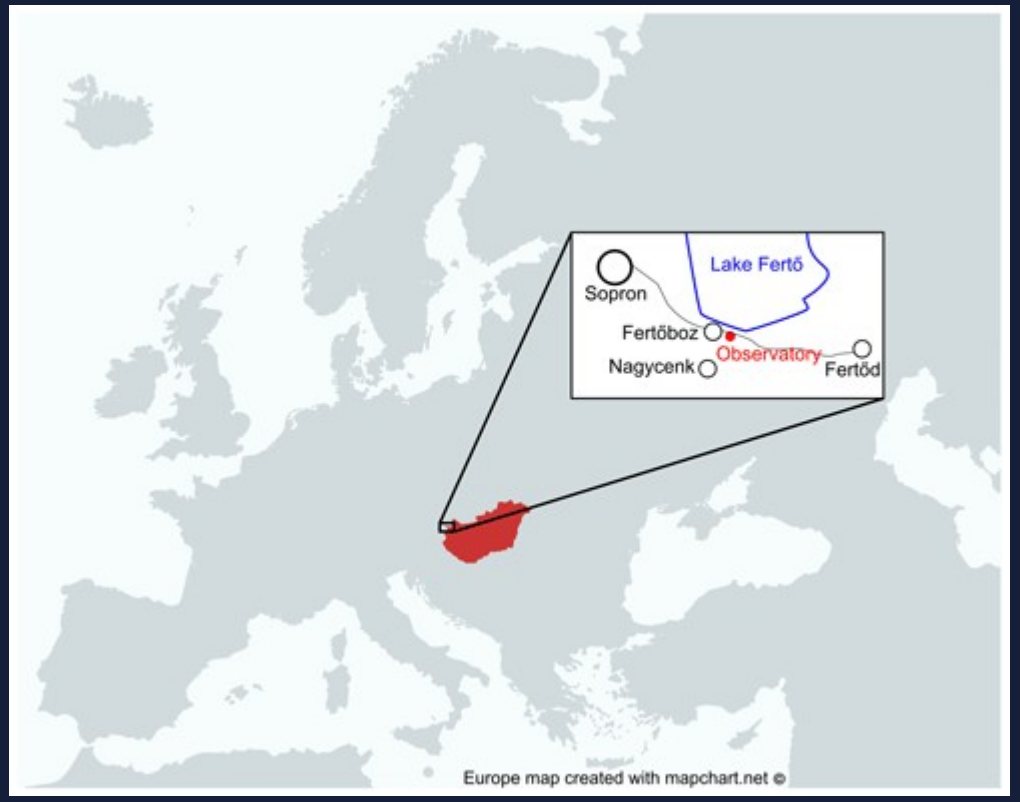


Figure source: Bór et al., Measurements of atmospheric electricity in the Széchenyi István Geophysical Observatory, Hungary. Hist. Geo Space. Sci., 11, 53–70, 2020.

https://doi.org/10.5194/hgss-11-53-2020

Objective of the project

- To **obtain a high time-resolution PG dataset** from the analogue photographical records.

Implementation:

- Creation of an **image processing algorithm**.
- **Scanning** the photo paper rolls of the PG measurement (Figure 1.)
- **Determination of numerical values** (Figure 2.)
  - The value of PG at a given time [V/m].
- **Digitization of a longer period (1999-2009).**

Figure 1.

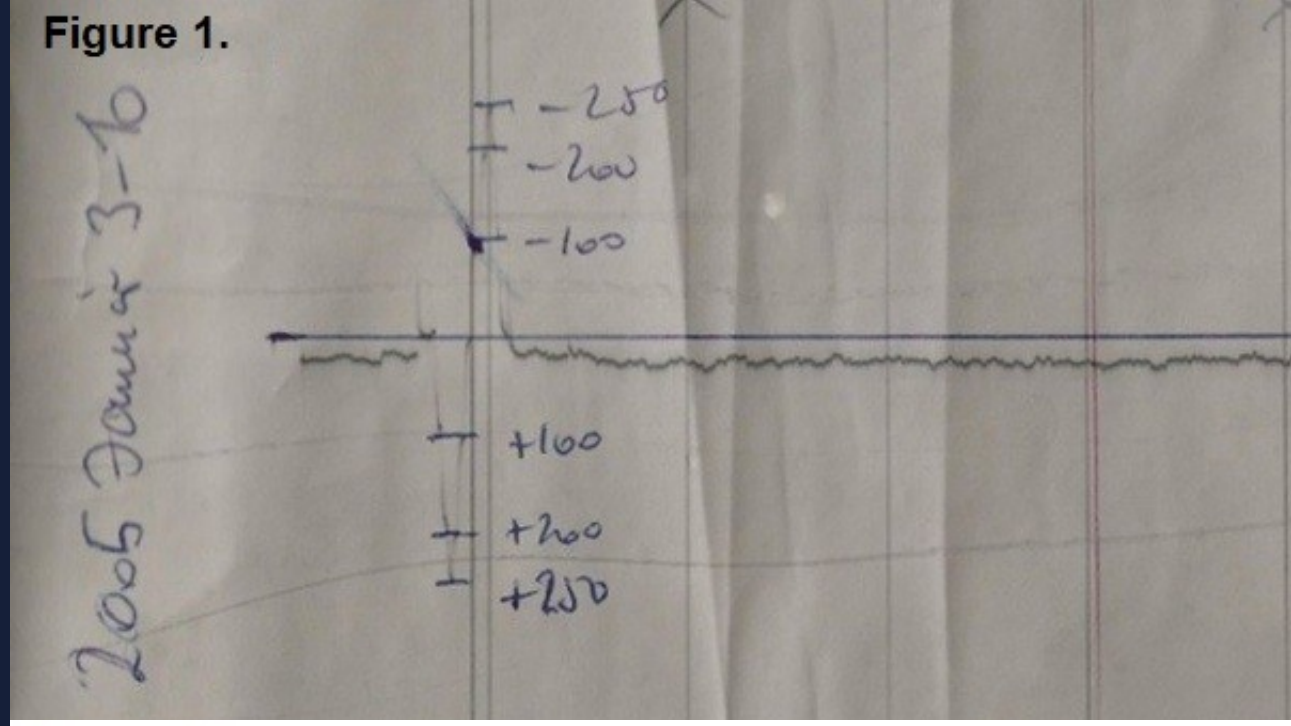


Figure 2.

Date [UTC]	PG [V/m]
2005-01-06 01:59:32;	40.37
2005-01-06 02:00:15;	38.12
2005-01-06 02:00:57;	35.88
2005-01-06 02:01:40;	33.64
2005-01-06 02:02:23;	31.4
2005-01-06 02:03:05;	29.15
2005-01-06 02:03:48;	24.67
2005-01-06 02:04:31;	22.43
2005-01-06 02:05:13;	20.18
2005-01-06 02:05:56;	17.94
2005-01-06 02:06:39;	20.18
2005-01-06 02:07:21;	17.94
2005-01-06 02:08:04;	20.18
2005-01-06 02:08:47;	22.43

Digitization of PG measurements

- **Semi-automatic** approach: automatic digitization but the validation of the result and finding the optimal parameters for the digitization requires supervision from a user.
- **Edge detection** (Figure 3.)
  - Edge: sudden intensity change -> interface between object and background,
  - Detect reference axes and the graph of PG measurement,
  - Canny edge detection algorithm.
- Software implementation
  - Python programming language,
  - OpenCV, NumPy, SciPy, imageio libraries.

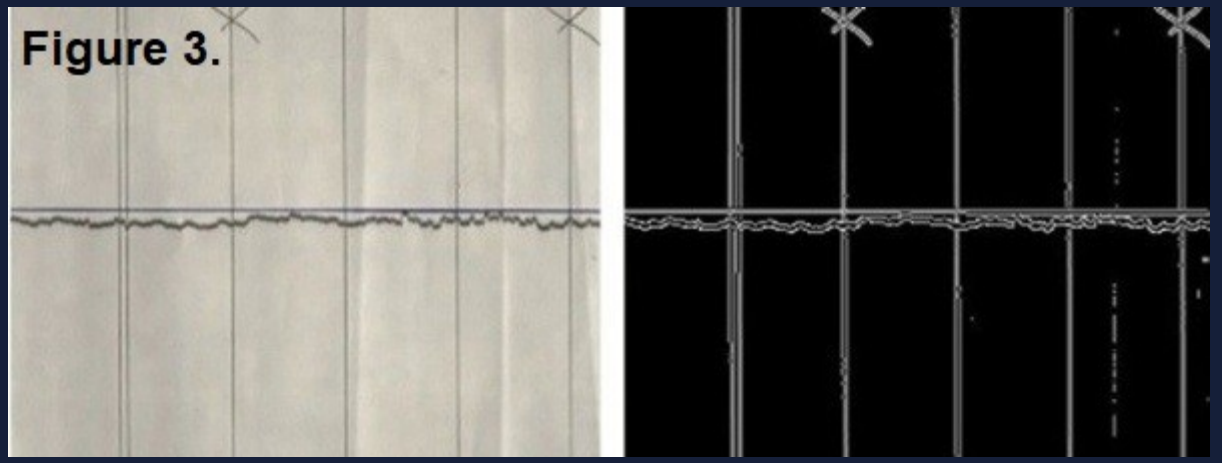
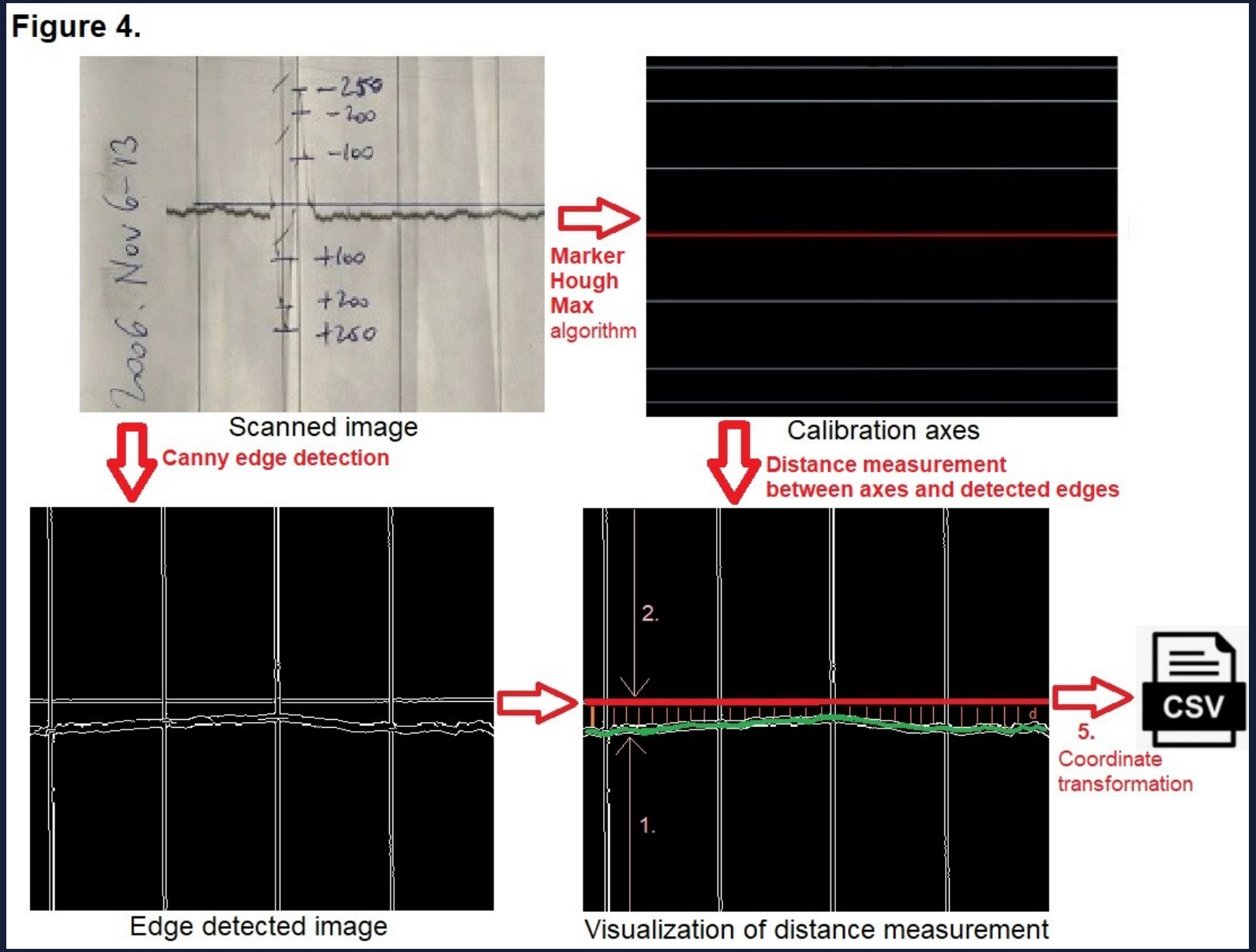
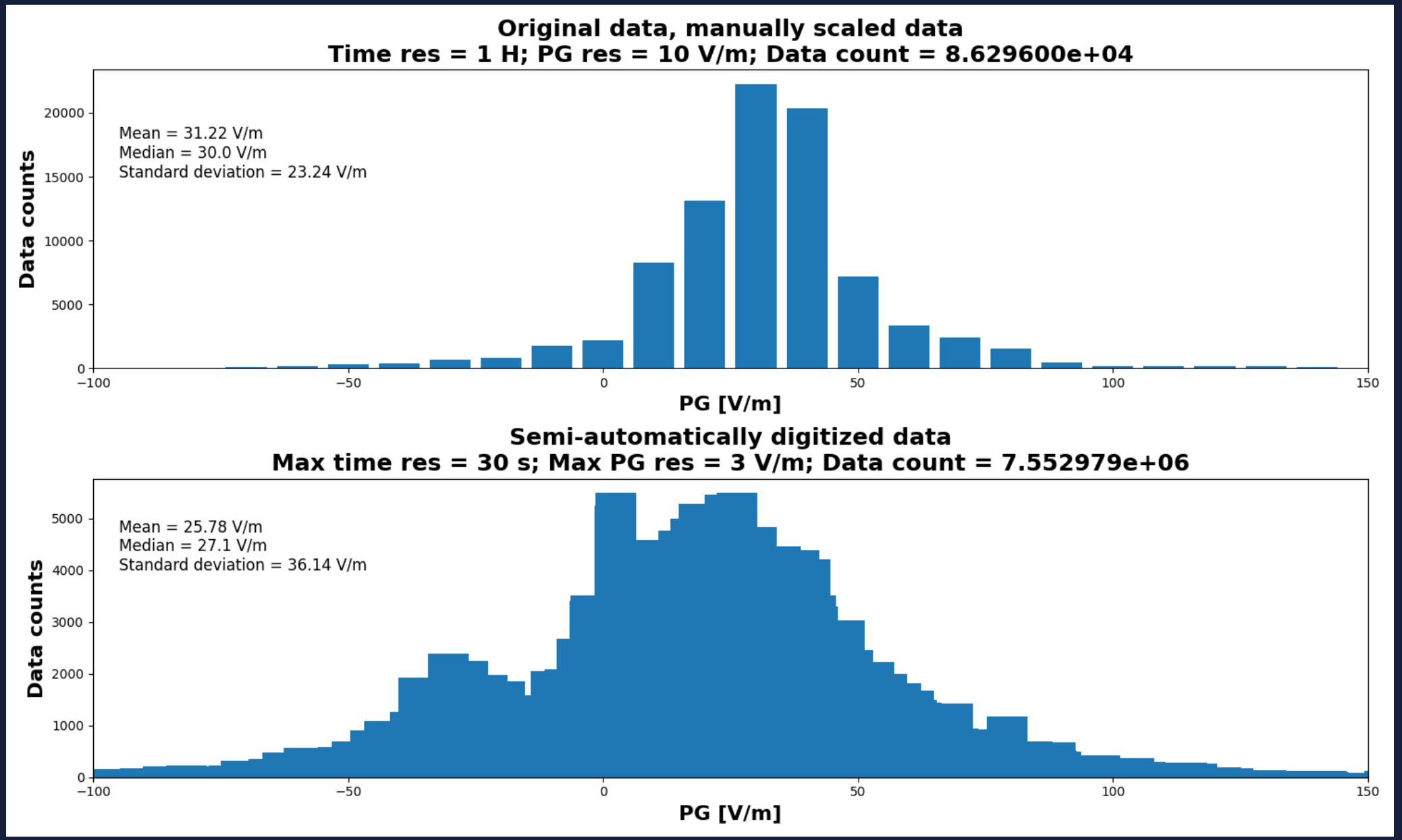


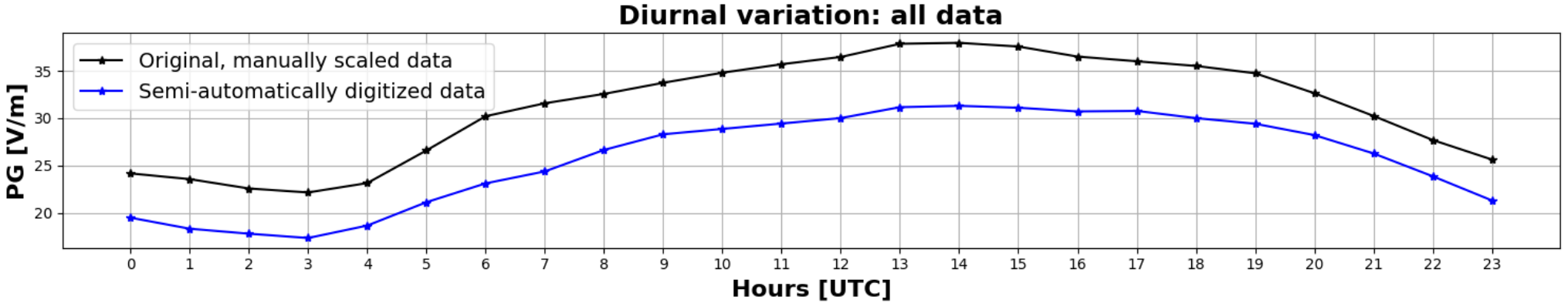
Image processing algorithm

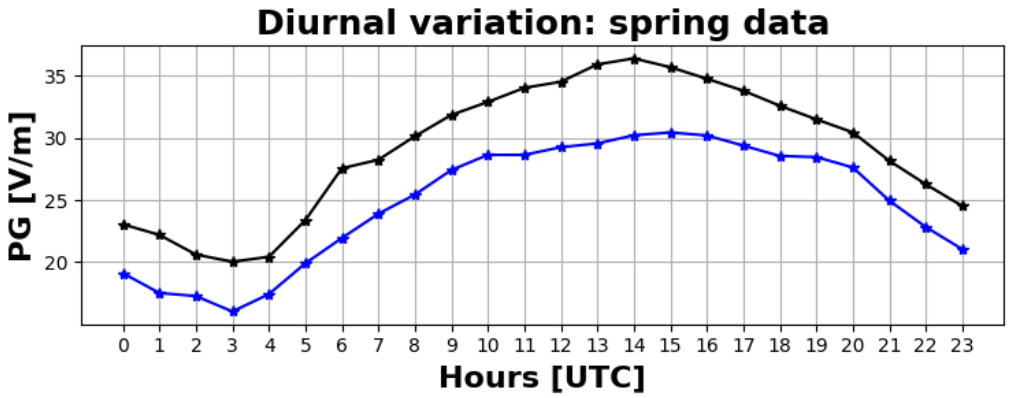


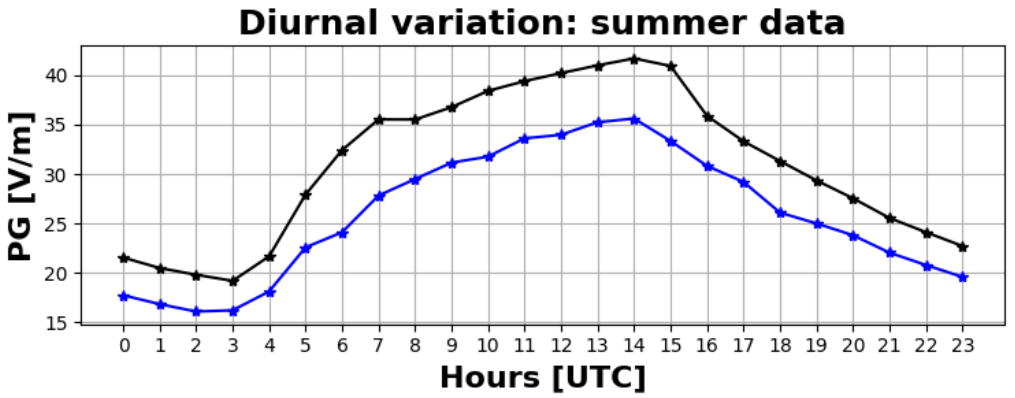
Validation

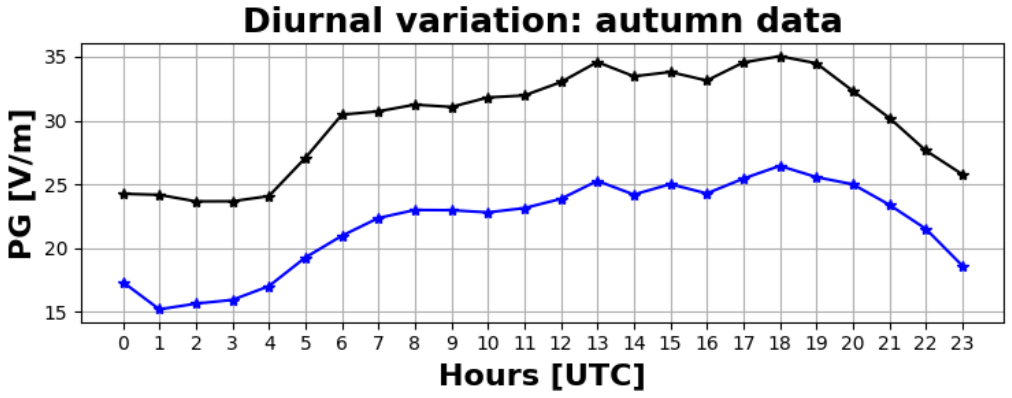

Original data, manually scaled data
Time res = 1 H; PG res = 10 V/m; Data count = 8.629600e+04
Mean = 31.22 V/m
Median = 30.0 V/m
Standard deviation = 23.24 V/m
Data counts
PG [V/m]
Semi-automatically digitized data
Max time res = 30 s; Max PG res = 3 V/m; Data count = 7.552979e+06
Mean = 25.78 V/m
Median = 27.1 V/m
Standard deviation = 36.14 V/m
Data counts
PG [V/m]

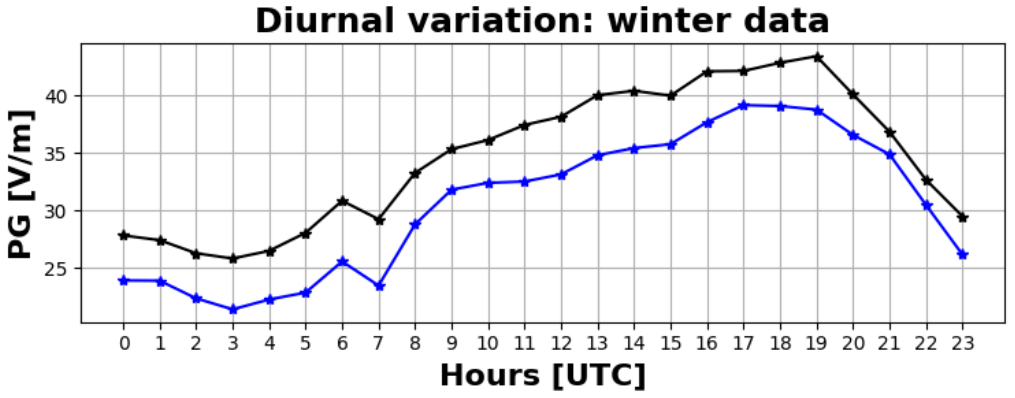
Validation


Diurnal variation: all data
Original, manually scaled data
Semi-automatically digitized data
PG [V/m]
Hours [UTC]


Diurnal variation: spring data
PG [V/m]
Hours [UTC]


Diurnal variation: summer data
PG [V/m]
Hours [UTC]


Diurnal variation: autumn data
PG [V/m]
Hours [UTC]


Diurnal variation: winter data
PG [V/m]
Hours [UTC]

Conclusions

- **The algorithm has proved to perform well on fair-weather PG data and produces high time resolution (~30 s) data** from the original, photographic records.
- **Resolution of the digitized data**: ca. 30 seconds, ca. 3 V/m.
- The **number of data points** increased by two orders of magnitude.
- **Runtime of the algorithm**: 30 seconds (digitization of a scanned image which contains weekly data set).
- **Compression rate**: 25x (10 MB picture -> 0,4 MB numerical value).
- **Plans for future**:
  - Code optimization,
  - digitizing images based on AI/deep learning solutions.

Thank you very much for your attention!

Attila Buzás
buzas.attila@epss.hu