Validation of EKF surface assimilation scheme

Report from stay at OMSZ Budapest
16/01/2017–17/02/2017

Viktor Tarjáni

Slovak Hydrometeorological Institute, Bratislava

Supervisors: Máté Mile and Helga Tóth

Hungarian Meteorological Service, Budapest

August 31, 2017

1 Introduction

During the previous stay several options have been suggested related to implementation of extended Kalman filter (EKF) surface analysis for AROME-HU at OMSZ. [1, p. 13-15] It was decided that priority should be given to 1-column validation as preliminary results of full-domain runs were not satisfactory and it is believed that properly implemented EKF could bring a superior result over the OI_MAIN approach which is now also tested at OMSZ.

In full-domain experiments CANARI (OI) analysis was used to provide gridded observations of air temperature and relative humidity at 2m height above a surface (T2M and HU2M). This is reasonable for operational surface analysis in which horizontal grid is given by driving atmospheric model and observations for each grid point have to be interpolated from relatively sparse observation network using CANARI.

For EKF validation purpose however such setup is not very suitable because CANARI gridded observations represents another source of error introduced into surface analysis. This is in fact CANARI analysis error for screen-level variables but from the EKF viewpoint it represents a contribution to observation error described by covariance matrix $\mathbf{R}$. Unfortunately, it is not easy to make a qualified guess about the errors introduced into surface analysis by using the gridded observations. They may consist not only from the random component but also from the systematic component (bias) which then produces also biased soil analysis. Whether it is over or under-estimated random errors magnitude or biased observations both will lead to degradation of performance or wrongly behaving EKF.

This was avoided by using screen-level observations taken at the identical location as model location instead of CANARI gridded observations. Here we used observation data taken at Debrecen-Kismacs meteorological site [2] during the year 2015. This site was selected because it provides a large set of observed variables. Besides screen-level measurements it collects also 10m measurements (used for offline SURFEX forcing) and soil temperature and wetness measurements at several depths (used for SURFEX initialization and in verification).

In present study we aim for thorough validation of current EKF implementation and its ability to compensate for deviations of model soil temperature and water content from their actual values. It is advantage if validation should be done under controlled conditions. This includes ability to distinguish different sources of error. For example using the station observation we are free of interpolation errors introduced by CANARI and observation error is then mainly due to vertical interpolation scheme which can be made under control by using 10m observations for forcing and testing for stable case when vertical interpolation scheme works best.

One can argue that using the exact observations in SURFEX initialization and forcing and also in EKF observation vector, leads to process and observation model with unrealistic low uncertainties, which is not the case in full-domain operative runs where such accurate data are seldom available for each grid-point. However with present setup we are able to easily introduce any error component (even biases) in controlled manner and study its impact on EKF behavior.

2 Description of the method

Temporal resolution of observation data taken at Debrecen-Kismacs was 10 minutes and they were available from 2015/01/01 00:00 to 2015/12/31 23:50, e.g. complete 2015 year was covered by measurements. Data were provided as single multi-column ASCII file each column representing specific observed variable with one row per record.

2.1 Forcing time series

Running the offline SURFEX requires providing forcing time series in NETCDF or ASCII format. We used ASCII format so that it could be easily manipulated or viewed with conventional tools.

Important change with respect to previous full-domain experiments was that forcing series were prepared directly from Debrecen-Kismacs observations and not from short-range AROME forecast (analysis) as before. Particularly tower measurements of 10m air temperature, 10m relative humidity and 10m wind and its direction were used. Additionally 10 minute cumulated precipitations, surface long-wave radiative flux, global short-wave radiation flux, diffuse short-wave radiation flux measurements were also used to prepare SURFEX forcing. The only missing important forcing quantity was surface pressure. Using observations from 10m height corresponds well with lowest model level of AROME-HU used to prepare forcing in full-grid SURFEX-EKF analysis.

Forcing time step was set equal to time resolution of observations (10 minutes) by default. This ensures most accurate forcing for SURFEX-EKF runs sine all available measurements are used in SURFEX forcing. SURFEX internal time step was set to 5 minutes (`XTSTEP_SURF = 300.0` in `&NAM_IO_OFFLINE`). Setting SURFEX time step same as forcing time step avoids interpolation from forcing points to SURFEX integration steps.

In full-grid setup such high frequency data are normally not available. Nevertheless it is possible to emulate lower temporal resolution by setting forcing time step integer multiple of observation time interval. For example to preserve original forcing temporal resolution 1 hour forcing time step could be used. Then only every 6-th record is used to prepare forcing series. One can also consider using moving average when downsampling observations to lower temporal frequency.

For 6 hour assimilation window with 1 hour forcing step forcing series contain 7 records (7 rows per each forcing file). First record corresponds to beginning of the assimilation window (typically 06:00), which equals to SURFEX initialization time (time of PREP file) and last record corresponds to end of the assimilation window (typically 12:00 of the same day).

2.2 Advantage of using observations in EKF analysis

By using the 2m observations T2M and HU2M taken at the exactly same horizontal location as location used in SURFEX-EKF runs, brings an advantage of minimizing the part of observation error coming from the horizontal shift between observation and model location. Error due to horizontal offset is part of representativeness error, which itself is part of observation error. In full-domain setup error due to horizontal offsets (grid-observation offsets) is usually important contribution to observation errors although it may be not so easy to quantify its magnitude. It is closely related to error of CANARI analysis of screen level variables.

The another important contribution to observation error should be the error of observation model $\mathcal{H}$. In this implementation of EKF this operator is part of the SURFEX model which calculates vertical profile of temperature, humidity and wind in surface boundary layer (allows to obtain 2m temperature and humidity) from soil surface and subsurface variables. Usually two different SBL schemes are used as observation operator in current implementation of EKF:

1. Interpolation scheme of Geleyn (diagnostic)
2. Canopy SBL scheme of Masson (prognostic)

Observation model error in case of diagnostic operator can come from two sources (see figure 1):

- forcing level error (error in upper BC)
- interpolation scheme error (error of shape of interpolating curve)

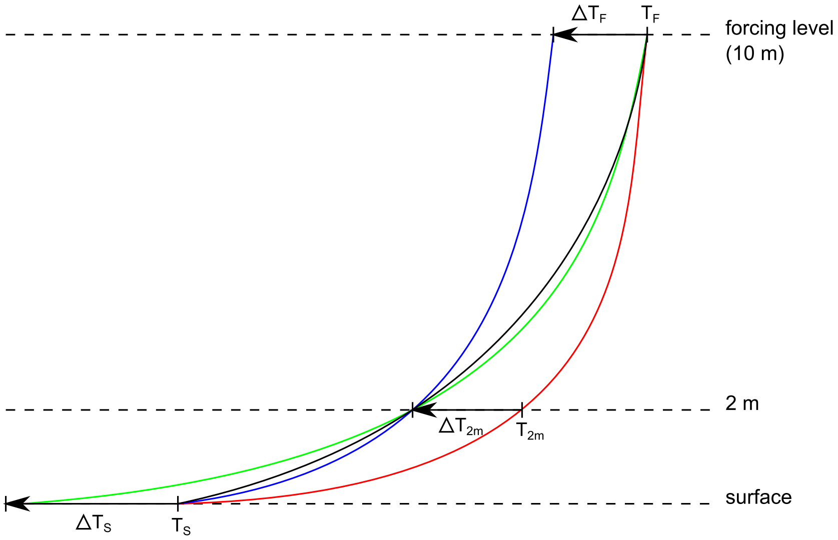

Figure 1: Illustration of possible sources of deviations of screen-level variables (here 2m air temperature) calculated by diagnostic interpolation routine. Red curve represents correct vertical profile of air temperature. Remaining three curves illustrate wrong vertical profiles with same error at screen level, but each due to different source. Blue curve: deviation due to error at forcing level, green: deviation due to error at surface, black: deviation due to error in shape.

Note that since observation error is defined as $\mathbf{y} - \mathcal{H}(\mathbf{x}_t)$ where $\mathbf{x}_t$ is true state of soil model, therefore error of vertical profile due to error in soil model state $\mathbf{x}_b - \mathbf{x}_t$ (error in bottom BC of vertical profile) is not part of observation operator error but it is rather background error. In this sense black and blue curves in figure 1 illustrates obs. operator error but green curve represent rather background error as seen in observation space, i.e. soil state error mapped to observation space (screen level variables).

Thus surface state error (background departure) represent a useful signal, because it is assumed that deviation of soil state in the SURFEX (ISBA) from actual soil state leads to deviations of 2m model variables from observed 2m values. Through Kalman gain matrix (weights) this deviation should be used to compensation of deviation of model soil variables.

In full-domain setup the offline SURFEX forcing is typically taken from lowest model level of driving atmospheric model analysis. Thus errors in SURFEX forcing variables could be identified with analysis error of driving model at lowest model level in this case.

In current 1-column experiments forcing errors are minimized by using temperature, humidity and wind observations from 10m height (approximately corresponds to height of lowest AROME model level above surface) and also by using measured radiative and precipitation fluxes.

Errors due to interpolation scheme can be reduced if assimilation is performed in those meteorological situations for which vertical interpolation scheme performs best. It is known that under stable, clear-sky conditions vertical interpolation scheme usually works optimally.

Using observations in SURFEX forcing allows also most accurate calculation of Jacobian matrices. Calculating the Jacobian matrices requires running offline SURFEX 5 times: reference + 4x perturbed, from beginning of the assimilation window until the end of the assimilation window. Normally assimilation window length is 6 hours, but for testing purposes it could be made also longer or shorter. In addition to already mentioned points high temporal resolution of forcing data allows accurate linearization of forward model or prognostic observation operator.

2.3 Physiography data (PGD)

Proper setup of physiography data is also very important. Since PGD parameters usually don't change with time (they are kind of climatic data) they can introduce considerable biases (systematic deviation from true state) in system if they are not properly set. This can lead to poor EKF behavior.

Soil parameters like root layer depth, deep layer depth are deduced from cover types present in simulated grid-box. Other important soil parameters like field capacity, wilting point, saturation are deduced from soil texture, e.g. clay and sand fractions at given location. Currently since no other data were available to us soil texture was retrieved from HWSD database and cover types were determined from ECOCLIMAP database. This lead to using only the Cover 172 consisting from mixture of C3 and C4 crops. This can be questionable because from photography of the Debrecen-Kismacs site it seemed that the predominant cover is grass.

2.4 Initialization of SURFEX-EKF

In current EKF implementation only NATURE tiles are taken into account. Principal numerical scheme for NATURE tiles in SURFEX is ISBA. Starting integration of ISBA requires specification of initial values of ISBA prognostic variables in addition to PGD data specification. Most importantly soil initial temperature and soil water profile must be given. Since we have available measurements of soil temperature and soil water content in different depths we used them to initialize ISBA scheme. Other possibility is initialization from guess values.

For 1-column SURFEX configuration we find it advantageous to initialize ISBA temperature and water profiles from prescribed uniform values set directly in `&NAM_PREP_ISBA` namelist in OPTIONS.nam file. By this we get rid of necessity to prepare additional input files, one for each ISBA variable, and possible horizontal interpolation.

2.4.1 Initialization of ISBA soil temperature

Independently whether 2 or 3 layer force restore variant is used, only two prognostic temperatures are considered by ISBA: superficial temperature TG1 which represents temperature of thick surface layer and TG2 which represents average of TG1 temperature over previous 24 hours. TG1 was initialized with observation from 10 cm depth sensor and TG2 was initialized with TG1 average during previous 24 hours.

Care must be taken when initializing soil temperature using the uniform prescribed values, set with `XTG_SURF`, `XTG_ROOT` and `XTG_DEEP` in `&NAM_PREP_ISBA` namelist, because SURFEX automatically adds correction to each prescribed soil temperature depending on grid box elevation. Final temperature written in `PREP.txt` thus differs from value prescribed in `OPTIONS.nam` and is calculated as:

$$\mathrm{TG}_{\mathrm{prep}} = \mathrm{TG}_{\mathrm{nml}} + k * \mathrm{ZS} \tag{1}$$

where $\mathrm{TG}_{\mathrm{nml}}$ is soil temperature prescribed in namelist, $\mathrm{TG}_{\mathrm{prep}}$ is corrected soil temperature written to PREP.txt, ZS is elevation of the grid-box in meters above mean sea level (it was set with `XUNIF_ZS` in `&NAM_ZS` namelist) and k is vertical gradient of temperature. SURFEX uses value $k = -0.0065\,\mathrm{K/m}$. Elevation of the grid-box was set with `XUNIF_ZS` in `&NAM_ZS` namelist.

We prefer to explicitly set ISBA initial temperature profile to specific values, however we were not able to find option in SURFEX to turn-off a temperature correction. We thus used simple trick to overcome this: Temperatures written to `&NAM_PREP_ISBA` namelist were calculated as:

$$\mathrm{TG}_{\mathrm{nml}} = \mathrm{TG}_{\mathrm{prep}} - k * \mathrm{ZS} \tag{2}$$

where $\mathrm{TG}_{\mathrm{prep}}$ is intended initial soil temperature, i.e. we subtract correction from intended initial temperature so that later when SURFEX automatically adds same correction, it is effectively eliminated and intended (uncorrected) temperature is written to output PREP.txt. Subtraction is done inside the control python script `run_ekf.py` which is also responsible to substitute calculated values to OPTIONS.nam file.

2.4.2 Initialization of ISBA soil water content profile

For soil moisture transport (including water transport) there are 2 variants of ISBA force-restore scheme in SURFEX: 2-layer (2-L) and 3-layer (3-L). Vertical coordinate $z[m]$ increases downward and soil surface is put at $z = 0$ in both variants. Layer configuration can be described with following set of equations:

$$\begin{aligned} z_1 &= \Delta_1 \\ z_2 &= \Delta_2 \\ z_3 &= \Delta_2 + \Delta_3 \end{aligned} \tag{3}$$

For 2-layer ISBA only first two equations make sense. Superficial layer extends from surface $z = 0$ to depth z_1 and has thickness Δ_1 ($\approx 1\,\mathrm{cm}$). Bulk soil layer extends from surface $z = 0$ to depth z_2 and has thickness Δ_2 ($\approx 2\,\mathrm{m}$), which means that bulk layer overlaps whole superficial layer. For 3-layer ISBA meaning of variables is little changed. Superficial layer is left as it is in 2-L version with same meaning for z_1 and Δ_1. But bulk layer is separated to two layers: root layer and deep layer. Root layer extends from soil surface $z = 0$ to depth z_2 ($< z_3$) having thickness Δ_2 ($\approx 1.5\,\mathrm{m}$) and it overlaps the superficial layer. Deep soil layer extends from base of root-zone layer z_2 to depth z_3 and has thickness Δ_3 ($\approx 0.5\,\mathrm{m}$). It doesn't overlap any other layer.

For ISBA 2-L prognostic variables WG1, WG2 represent average VWC in superficial and bulk soil layer respectively. For ISBA 3-L prognostic variables WG1, WG2, WG3 represent average VWC in superficial, root and deep layer respectively.

To initialize soil volumetric water content (VWC) profile we used possibility to prescribe uniform value for each of three main soil layer directly in namelist (`XHUG_SURF`, `XHUG_ROOT`, `XHUG_DEEP` in `&NAM_PREP_ISBA`) similarly as for soil temperature profile.

Unlike the prescription of temperature profile, no altitude correction is applied when initializing soil VWC profile during PREP. However prescribed values in `&NAM_PREP_ISBA` are assumed to be given as soil wetness index (SWI) rather then VWC[1].

Conversion from SWI to VWC is done in SURFEX file `prep_hor_isba_field.F90` file by following two formulas:

$$W_g = W_{\mathrm{wilt}} + \mathrm{SWI}\left(W_{\mathrm{fc}} - W_{\mathrm{wilt}}\right) \tag{4a}$$

$$W_g = \max\left(\min\left(W_g, W_{\mathrm{sat}}\right), W_{g,\mathrm{min}}\right) \tag{4b}$$

where W_g, W_{wilt}, W_{fc} and W_{sat} are respectively VWC, wilting point, field capacity and saturation (porosity), all given in m^3/m^3 and generally they are all functions of depth and thus change from layer to layer (but not for force-restore scheme).

Values W_{wilt}, W_{fc} and W_{sat} (so called hydrolimits) required by conversion are calculated using the empirical pedotransfer functions given the fractions of clay and sand in the soil (more details are in section 2.4.4).

As observations which we used to initialize soil water profile represent VWC in m^3/m^3 rather then SWI assumed in OPTIONS.nam, therefore simplest solution was to manually let calculate first hydrolimits and then in python control script calculate SWI from observed VWC using the equation (4a), and substitute calculated SWI values into `&NAM_PREP_ISBA` namelist in OPTIONS.nam file prior to execution of PREP program for reference offline run.

2.4.3 Initialization of soil water content error variances

While initialization of the soil water content profile is done inside the main SURFEX code during the execution of PREP program, initialization of background, model and observation error variances of soil water content for control layers is done inside the EKF code in file `varassim.f90` during the execution of VARASSIM program.

Observation error variance of soil water content (or better SWI) is only relevant if assimilating also satellite observations of SWI, therefore only background and model error variances are discussed here.

Initialization of background error variances σ_b^2 is done as a part of initialization of background error covariance matrix $\mathbf{B}$ inside the `LPRT` code block. Originally background error variances

---

[1] VWC values are in units m^3/m^3 are used by soil scheme but are also written to PREP.txt file

were calculated as:

$$\sigma^2_{b[WG2]} = \left[\sigma'_{b[WG2]}(W_{fc} - W_{wilt}) + W_{wilt}\right]^2 \tag{5a}$$

$$\sigma^2_{b[WG1]} = \left[\sigma'_{b[WG1]}(W_{fc} - W_{wilt}) + W_{wilt}\right]^2 \tag{5b}$$

where primed sigma's on the right are standard deviations set through namelist (`XSIGMA_M(i)` in `&NAM_VAR`) and they are assumed to be given as soil wetness index (SWI), while unprimed sigma's on the left are actual values used during the assimilation given as volumetric water content in $[m^3/m^3]$. W_{fc} and W_{wilt} are respectively soil field capacity and wilting point both given as VWC in $[m^3/m^3]$. More details on their calculation are given in section 2.4.4.

Initialization of model error variances σ^2_q is done as a part of $\mathbf{Q}$ matrix initialization which is executed in `LANA` block but only if `LBFIXED=FALSE`. Originally they were calculated as:

$$\sigma^2_{q[WG2]} = q_s q_s \left[\sigma'_{q[WG2]}(W_{fc} - W_{wilt}) + W_{wilt}\right]^2 \tag{6a}$$

$$\sigma^2_{q[WG1]} = q_s q_s \left[\sigma'_{q[WG1]}(W_{fc} - W_{wilt}) + W_{wilt}\right]^2 \tag{6b}$$

where q_s is empirical scaling factor set in the namelist which represents ratio between model and background error standard deviations.

In (5) and (6) wilting point W_{wilt} is added as a part of conversion from SWI to VWC. However, we have found this unreasonable and no such addition was found in the original J. F. Mahfouf code. We have thus modified those formulas by removing addition of W_{wilt}:

$$\sigma^2_{b[WG2]} = \left[\sigma'_{b[WG2]}(W_{fc} - W_{wilt})\right]^2 \tag{7a}$$

$$\sigma^2_{b[WG1]} = \left[\sigma'_{b[WG1]}(W_{fc} - W_{wilt})\right]^2 \tag{7b}$$

and

$$\sigma^2_{q[WG2]} = q_s q_s \left[\sigma'_{q[WG2]}(W_{fc} - W_{wilt})\right]^2 \tag{8a}$$

$$\sigma^2_{q[WG1]} = q_s q_s \left[\sigma'_{q[WG1]}(W_{fc} - W_{wilt})\right]^2 \tag{8b}$$

2.4.4 Calculation of soil hydrolimits

Soil hydrolimits (soil hydraulic parameters), namely wilting point W_{wilt}, field capacity W_{fc} and saturation W_{sat} are needed in two places:

1. During the PREP to convert initial SWI of each soil layer to volumetric water content in the soil layer $[m^3/m^3]$, see section 2.4.2 and equations (4a), (4b)
2. During the VARASSIM to convert background and model error variances from SWI (set in the namelist) to volumetric water content (used in numerical scheme), see section 2.4.3 and equations (7), (8)

As hydrolimits are usually not directly available (not measured), they are calculated from the values of clay and sand fractions[2] in given soil column using the empirical pedotransfer function (PTF for short). As they depend only on soil type they are assumed to be constant in time, but they can change with depth (each soil layer can be assigned different hydrolimits).

---

[2] These fractions can be manually set in OPTIONS.nam or more usually they are read from input HWSD files.

In EKF code (`varassim.F90`) identical value of W_{wilt} and W_{fc} are assigned to each ISBA soil layer using the fixed pedotransfer function (CH78). In SURFEX different PTF can be chosen through namelist option `CPEDO_FUNCTION` in OPTIONS.nam, except if force-restore soil scheme (2-L or 3-L) is used. In that case SURFEX always resets back to Clapp and Hornberger 1978 (CH78) PTF independently on value of `CPEDO_FUNCTION` in `&NAM_ISBA` namelist. Namelist value is taken into account only if diffusion soil scheme is used, but there is then possibility of inconsistency when for example soil water profile is initialized using the CO84 PTF (set in SURFEX namelist) but error variances in EKF are initialized using the CH78 PTF (hard-coded in `varassim.F90`).

For consistency it is important to ensure that VWC profile as also EKF variances are converted from SWI to VWC using the same values of hydrolimits in each model layer which imply using the same PTF. We have thus modified EKF source code (`varassim.F90`) as follow:

- Following lines were added:

  ```
  !! Read type of ISBA soil scheme (2-L, 3-L, DIF)
  CALL READ_SURF(YPROGRAM,'ISBA',YISBA,IRESP)

  !! Select pedo-transfer function
  IF (IVERSION>=7) THEN
    CALL READ_SURF(YPROGRAM,'PEDOTF',YPEDOTF,IRESP)
  ELSE
    YPEDOTF = 'CH78'
  ENDIF
  ! Only Clapp and Hornberger 1978 with Force-Restore scheme
  IF(YISBA/='DIF')THEN
    YPEDOTF='CH78'
  ENDIF
  ```

- Following lines were replaced

  ```
  DO I=1,NSIZE_NATURE
    COFSWI(I)=0.001*(89.0467*((100.*ZCLAY(I))**0.3496)-37.1342*((100.*ZCLAY(
        I))**0.5))
    WWILT(I)=0.001*37.1342*((100.*ZCLAY(I))**0.5)
    SMSAT(I)=0.001*(-1.08*100*ZSAND(I)+494.305)
  ENDDO
  ```

  with

  ```
  WWILT(:) = WWILT_FUNC(ZCLAY(:),ZSAND(:),YPEDOTF)
  WFC(:) = WFC_FUNC(ZCLAY(:),ZSAND(:),YPEDOTF)
  WSAT(:) = WSAT_FUNC(ZCLAY(:),ZSAND(:),YPEDOTF)
  COFSWI  = WFC - WWILT
  ```

This ensures that soil hydro-limits are calculated with exactly same equations in SURFEX as in EKF.

3 Test run

For test run clear-sky no precipitation period was searched for in observation data. One such started on July 5, 2015 and ended on July 9, 2015. Highest measured wind speed during the period was 9.0 m/s.

SURFEX initial time was set to July 5, 2015 at 06:00 UTC. For initial run 2-layer ISBA was used without CANOPY model. TG1 and WG1 were initialized with observations from 10 cm depth while TG2 temperature was initialized to average of TG1 during previous 24 hours and WG2 was initialized with average volumetric water content in bulk layer calculated from observations.

4 Summary

For the purpose of 1 column validation several python scripts were written. One for preparation of forcing time series from original observation file. Another one for EKF assimilation execution control, i.e. control script `run_ekf.py`. Modification of original EKF code `varasim.f90` was also necessary. Following changes were done:

- Reading of CANARI 2m gridded observations was replaced by reading of 2m station observations
- Addition of W_{wilt} in calculation of background and model error covariance matrices was removed
- Calculation of soil hydrolimits in EKF code was made consistent with SURFEX code

Validation of the EKF surface assimilation will continue on. More results and detailed discussion will be given in the next report.

5 Acknowledgements

I would like to acknowledge following people for their help and assistance related with my Budapest stay: *Mile Máté, Tóth Helga, Szintai Balázs, David Lancz, Kullmann László, Ágh Attila and Roman Zehnal.* Particularly I would like to thank Mr Nagy Zoltán who readily prepared for us a set of observation data from Debrecen-Kismacs and Debrecen-Látókép meteorological sites.

References

[1] V. Tarjáni. *Surface assimilation using extended Kalman filter, Report from stay at OMSZ Budapest 05–16/10/2015, 02–13/11/2015.* 2016. URL: http://www.rclace.eu/File/Data_Assimilation/reports/vt_report_budapest.pdf.

[2] Z. Nagy et al. "Debrecen-Kismacs és Debrecen-Látókép mérőállomás talajnedvesség adatsorainak elemzése". 42. Meteorológiai Tudományos Napok. 2016. URL: http://www.met.hu/doc/rendezvenyek/metnapok-2016/12_NagyZoltan.pdf.