Archimer
https://archimer.ifremer.fr

An Advanced Algorithm to Retrieve Total Atmospheric Water Vapor Content From the Advanced Microwave Scanning Radiometer Data Over Sea Ice and Sea Water Surfaces in the Arctic

Zabolotskikh Elizaveta, V [1], Khvorostovsky Kirill S. [1], Chapron Bertrand [2]

[1] Russian State Hydrometeorol Univ RSHU, Satellite Oceanog Lab SOLab, St Petersburg 195196, Russia.
[2] Inst Francais Rech Exploitat MER IFREMER, Ctr Brest, Plouzane 29280, France.

Corresponding authors : email addresses : liza@rshu.ru ; ; kirill@rshu.ru ; bertrand.chapron@ifremer.fr

Abstract :

An advanced algorithm for atmospheric water vapor column (WVC) retrieval from the Advanced Microwave Scanning Radiometer (AMSR) measurements over the Arctic sea ice (SI) and open ocean waters is presented. The algorithm is built on the physical modeling of the brightness temperature (BT) of the microwave radiation of the SI-open ocean-atmosphere system at the AMSR frequencies and polarizations. The BTs are calculated using a data set of the SI, atmospheric, and oceanic parameters changing in the range of their natural variability in the Arctic, and using the SI microwave emission coefficients varied according to the published experimental data. The inverse operator explores neural networks (NNs), trained on an ensemble of modeled BTs. The algorithm is applied both to the AMSR-E and to the AMSR2 measurement data. Validation of the algorithm is performed with radiosonde (r/s) WVC measurements from the four Arctic coastal stations at different SI conditions during 2014-2017. The results of the application of the new algorithm to satellite radiometer measurements are also compared with the Era-Interim reanalysis WVC, as well as with other satellite WVC products, based on the data of the Moderate Resolution Imaging Spectrometer (MODIS) and on the data of the Advanced Microwave Sounding Unit-B (AMSU-B) for 2008 and 2015. To justify the usage of the Era-Interim WVC as a reference data set for the algorithm accuracy estimation in the Arctic area, Era-Interim WVC is also compared with the r/s WVC measurements.

Keywords : Advanced Microwave Scanning Radiometer (AMSR), Arctic, atmospheric water vapor column (WVC), brightness temperature (BT), neural networks (NNs), retrieval algorithm, sea ice (SI)

I. INTRODUCTION

Water vapor is one of the major natural greenhouse gases. It is a key parameter in the global hydrological cycle and a component of climate change and ocean - atmosphere energy exchange studies [1]. Study of the atmospheric water vapor distribution and variability can supposedly help to understand the climate response to increasing greenhouse gas concentrations [2]. A number of recent studies investigated the water vapor column (WVC) climatic trends and their accuracy exploring different data sets [3]–[12]. These studies explore different WVC data sets and time periods. The accuracy of WVC data is discussed in most of them. Large deviations between WVC data sets over some areas, including Arctic region, were found from the analysis of the 22 WVC satellite-based and re-analysis data sets.

The WVC changes over the Arctic area are poorly studied not least due to the scarcity of direct observations [13]. Radiosonde (r/s) WVC measurement data have long been considered as a source of the most accurate WVC providing precise moisture data at different pressure levels with high vertical resolution [14]. However, being available only from the sparse r/s stations, not all r/s humidity observation can be qualified as complete to calculate WVC [15]. Moreover, the raw radiosonde humidity records contain large errors and biases which need to be homogenized before their usage for the studies of long-term changes [16]. Some satellite retrievals of moisture related quantities are even used as reference data to estimate radiosonde data accuracy [17]. Nevertheless, the r/s WVC measurements are still most widely used both for validation purposes [18]–[21] and in climate studies of the WVC trends [9], [12], [22].

A highly accurate WVC data may be obtained from the observations at ground-based Global Positioning System (GPS) networks [23]. One of the most important advantages of the GPS WVC data is their high temporal resolution and providing homogenized data, which may be used both for the correction of r/s WVC measurements and for validation studies [24]. Unfortunately, low elevation angles, poor visibility of GEO satellites, sparse monitoring

infrastructure and ionospheric effects limit usability of the Global Navigation Satellite Systems (GNSS) in the Arctic [25].

Some other ground-based instruments, such as microwave radiometers, Raman lidars, multifilter rotating shadow band radiometers and sun photometers are also capable to measure WVC. But only satellite remote sensing measurements provide regular basin-wide WVC estimates in the Arctic. A detailed review of various satellite-based techniques for WVC retrievals as well as the description of different WVC datasets, their advantages and limitations is given in [2]. Here we will outline the general features of water vapor sensors in terms of the Arctic cloud cover that exceeds 65% as measured by the different sensors [26].

Visible and near IR methods are based on attenuation measurements of solar radiation over cloudless areas. The WVC retrievals from these sensors are limited to the daylight portion of the swath [19]. Another group of methods explores thermal IR radiation measurements. These methods are independent of the day light but under cloudy conditions they allow measuring WVC only over clouds [27]. The observed WVC underestimation in many climatological studies is attributed to exploiting only the clear sky scenes [6], [28], but the amplification of this bias in the Arctic have not yet been investigated. In particular, the WVC distribution, trends, and anomalies over the Arctic for the period 2001–2015 obtained in [3] using the WVC monthly product of moderate resolution imaging spectrometer (MODIS) data incorporate inevitable errors since this product is generated only from cloudless data.

Recent progress in satellite atmospheric parameter monitoring, independent of clouds and sun light, is associated with the deployment of a large networks of limb-sounding satellites, using GPS radiooccultation (RO) [29]. GPS RO ability to penetrate through clouds and insensitivity to surface emissivity ensures avoiding biases typical for WVC measurements from other satellite missions [2]. The GPS RO measurements have a potential to fill the gap in observing the WVC over the Arctic. Unfortunately, the RO sounding often does not reach the surface in the Arctic that limits the RO capabilities to measure WVC in this region. [30].

Regular long-term observations of water vapor are provided by satellite passive microwave instruments. Both satellite microwave imagers and sounders allow retrieving WVC for clear and cloudy sky conditions, though not in the presence of strong scattering by cloud/precipitation hydrometeors [31]. Both WVC and total cloud liquid water have been successfully retrieved from such satellite passive microwave instruments (PM), as the Scanning Multichannel Microwave Radiometer (SMMR) on board the Nimbus 7, the Special Sensor Microwave/Imager (SSM/I) and the Special Sensor Microwave Imager and Sounder (SSMIS) on board the DMSP series satellites, the WindSat radiometer, the Microwave Instrument onboard the Tropical Rain Measurement Mission (TRMM) satellite (TMI) and the series of the Advanced Microwave Scanning Radiometers (AMSR, AMSR-E and AMSR2) [32]. Many algorithms for WVC retrieval from PM imager data have been developed for non-precipitating atmosphere over ocean surface during the last decades [33]–[36]. WVC retrieval over other surface types is difficult due to their complex and variable emissivity.

Recent studies demonstrated possibilities of WVC retrieval from PM imager measurements over the surfaces, different from open oceans. These studies are based on PM measurements at the frequencies close to the frequency of water vapor resonant absorption at 22.235 GHz [37]. These methods assume that the surface electromagnetic properties are the same at close frequencies [38], [39]. However satellite WVC products, based on PM imager data, are still available only over open oceans [34], [40] due to high retrieval errors over the surfaces with large variability of microwave emissivity. The mean retrieval error of ~ 4.7 kg/m^2 reported in [39] is close to monthly regional natural WVC variability, thus the WVC estimates do not provide valued information.

The major progress in WVC retrieval from satellite measurements over the Arctic sea ice is associated with the usage of satellite passive microwave sounder data, and has been reported by the scientists from the University of Bremen [41]. The key concept of their method was suggested in [20], where the authors used several channels near the 183.31 GHz water absorption

line, for which microwave radiation is featured by similar surface emissivity but different water vapor absorption behavior. The usage of the 183.31 GHz channels along with the channel at the 150 GHz window frequency allowed successful retrieval of WVC lower than ~7 kg/m^2 over the sea ice from such satellite instruments as the Advanced Microwave Sounding Unit-B AMSU-B (AMSU-B) or the Microwave Humidity Sounder (MHS). In the retrieval process the authors of [41] initially exploited the 89 GHz channel measurements along with several assumptions on the sea ice emissivity. However, the WVC values were estimated with large errors due to high variability and lack of knowledge about sea ice emissivity at 89 GHz. The following version of the algorithm was supposed to be valid for WVC < 15 kg/m^2 over the sea ice and for WVC < 7 kg/m^2 over the sea water with an error of ~ 3 kg/m^{2}over the sea ice surfaces [42]. Then the authors extended the method to be applicable for a lager range of WVC over the sea water surface by the development of the sub-algorithm specifically for ice-free and partially ice-covered areas [41]. The new algorithm was tested against the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim [43] reanalysis data over the whole Arctic and against the other two WVC PM satellite products over open ocean areas. The reported error for summer months of ~ 5.7 kg/m^2 seem to be too high to to exploit this product.

Reviewing published research studies, we can conclude that despite the highly essential role of water vapor in the Earth's climate, there is still lack in reliable WVC data. The results of climatic WVC studies are sometimes inconsistent. E.g. in [11] the positive WVC trends are found from GPS, radiosonde, and PM data both over land and over oceans, with larger change rates over oceans. In contrast the WVC trends obtained in [12] from the global COSMIC RO and radiosonde observations are positive only over land, while over ocean they are negative. Thus, the reliability of existing WVC data for the Arctic region needs to be further carefully investigated due to various inherent limitations.

Here we present a new method for WVC retrieval from the measurements of the AMSR-E and AMSR2 instruments with the similar sets of channels and characteristics. The methodology

used to estimate WVC is the same as in the algorithms, developed in [18], [32], [44] for the AMSR-E, SSM/I and AMSR2 data, and applied for the measurements over open ocean The methodology includes: 1) numerical simulation of the AMSR2 measured brightness temperatures (T_Bs) for a large number of surface and atmospheric parameters varying within the range of their natural variability, using physical T_B model, and 2) inversion of T_Bs into WVC by applying Neural Networks (NNs), trained with an ensemble of modeled T_Bs. The new method is based on the numerical simulation of the AMSR measured T_Bs both over sea ice (SI) and open water (OW), thus providing WVC estimates for the most part of the Arctic region. Validation of the algorithm is done with the ERA-Interim reanalysis and radiosonde data for the whole range of the WVC variability in the Arctic region. We also compared the results of the algorithm application to the AMSR-E data with the two publically available satellite-based data sets of the WVC in the Arctic. The first one is based on the application of the advanced algorithm to the AMSU-B data [41]. The second presents a MODIS WVC product, used in [3] to derive the water vapor trends in the Arctic atmosphere. We validated these products with the Era-Interim WVC data to compare the accuracies of WVC from these datasets with the accuracy of our AMSR-based retrievals. In addition verification of the Era-Interim WVC in the Arctic was done with the r/s ground-truth data to justify our conclusions.

II. INSTUMENT AND DATA DESCRIPTION

The AMSR – Earth Observing System sensor (AMSR-E) – is the first of well calibrated Japan AMSR instruments launched on 4 May 2002 onboard the Aqua satellite and operated until 4 October 2011. It was succeeded by the AMSR2 launched on 18 May 2012 onboard the GCOM-W1 satellite. Both instruments provide global measurements of the brightness temperatures (T_Bs) of vertically (V) and horizontally (H) polarized microwave radiation at the frequencies 6.9, 10.7, 18.7, 23.8, 36.5, 89.0 GHz with the equator crossing of descending and ascending orbits at 01:30

and 13:30 local time [45]. The AMSR2 basic characteristics are the same as those of its predecessor but with several enhancements, including additional T_B channels at 7.3 GHz (V and H). The main characteristics of the AMSR-E and AMSR2 instruments are presented in Table I.

The AMSR-E and AMSR2 data of different processing levels are freely distributed by the GCOM-W1 Data Providing service of the Japan Aerospace Exploration Agency (JAXA) through the G-Portal (https://gportal.jaxa.jp). In our study we use T_B data from the Level 1R (L1R) and Level 3 (L3) products. In the L1R product, that was made available comparatively recently, the calibrated T_B values are resampled to a common resolution that corresponds to the footprint center and scan number of the measurements at 89.0 GHz to account for different channel resolution. The L3 product provides the daily and monthly averaged geophysical parameters and T_Bs, calculated from the L1R product for the two grid resolutions of 0.1° and 0.25° separately for ascending and descending orbit paths.

The L1R product is used for matching with the r/s data to enable precise collocation in time, while the L3 product is appropriate for comparison with the daily averaged WVC products.

The following data sets are used additionally in this study:

- The Era-Interim reanalysis data on air temperature, humidity, pressure, liquid water content profiles, sea water temperature, sea ice temperature and sea ice concentration for 2°×2° grid are used for the AMSR T_B model calculations.

- The r/s WVC measurements are used for validation of the WVC retrievals as well as for evaluation of the Era-Interim reanalysis data. We use the r/s profiles for three-years period from the four Russian Arctic coastal stations: 20744 Malye Karmakuly (72.36N, 52.7E), 20674 Ostrov Dikson (73.5N, 80.40E), 21432 Ostrov Kotelnyj (76.0N, 137.86E) and 20292 Gmo Im.E.K. Fedorova (77.71N, 104.3E). The r/s data, obtained by Russian RF-95 sondes, were downloaded from the University of Wyoming archive (http://weather.uwyo.edu/ upperair/sounding.html). These radiosondes have an accuracy of 7 % for relative humidity measurements. The data were

converted to absolute humidity with the new approximation formula for vapor pressure proposed in [46].

- The WVC Era-Interim reanalysis surface level daily data were downloaded from the ECMWF (https://apps.ecmwf.int/datasets/data/interim-full-daily) for the Arctic region at the grid of 0.75° × 0.75°. After evaluation they were used for the algorithm validation and as a benchmark data set to estimate the accuracy of other WVC satellite products.

- In the validation process we also exploit the Level 3 sea ice concentration (SIC) daily product of University of Bremen (https://seaice.uni-bremen.de) to determine the surface type (sea ice or open water). The SIC in this product is retrieved using the ARTIST Sea Ice algorithm (ASI) from the AMSR2 measurements at 89 GHz [47].

- The algorithm performance is also analyzed against the two satellite products:

- Daily WVC product, derived from the AMSU-B measured radiances [41]. The dataset of geo-referenced WVC data for the 4 months of 2008 – March, June, September and December – is available at https://doi.org/10.1594/PANGAEA.888100.

- Daily WVC product, derived from the Terra MODIS Near IR and IR channel measurements (https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/61/ MOD08_D3/2008).

III. ALGORITHM DEVELOPMENT

In general the algorithm development repeats the scheme, described in details in [18]. Numerical simulation of the brightness temperatures (T_Bs) at the AMSR frequencies, polarizations and incidence angle for the full range of atmospheric, oceanic and sea ice parameters allows obtaining a dataset of T_Bs and corresponding WVC values. The radiation transfer model is restricted by non-precipitating conditions. This restriction implies a simplified form of the Radiation Transfer Equation (RTE) in absorption-emission approximation and

neglecting microwave scattering on large cloud/rain drops and ice particles with a size $r > \lambda/2\pi$. For all the AMSR frequencies except 89 GHz such approximation is valid for clear and cloudy atmosphere [48]. At 89 GHz channels (wave length $\lambda \sim 3mm$) cloud liquid and ice water particles may produce non-negligible scattering effect. The problem in T_B modeling under any conditions, including radiation scattering, lies in the large influence and variability of water particle parameters (e.g. distribution of sizes and forms) on RTE and the absence of data to account for this influence.

In the absence of scattering the brightness temperature $T_B^{H,V}(\nu, \theta)$ of the sea ice – ocean – atmosphere system at frequency ν, incidence angle θ and polarization P (vertical-V or horizontal-H) can be written as:

$$T_B^{H,V} = T_a^{\uparrow} + [r^{H,V}_{eff} \cdot T_a^{\downarrow} + T_{s\_eff}^{H,V} + 2.7 \cdot r^{H,V}_{eff} \cdot e^{-\tau}] \cdot e^{-\tau}, \tag{1}$$

$$T_a^{\uparrow} = \frac{1}{\cos\theta} \int\limits_0^{\infty} T(h)\alpha(h) \cdot \exp(-\frac{1}{\cos\theta} \int\limits_h^{\infty} \alpha(h')dh')dh, \tag{2}$$

$$T_a^{\downarrow} = \frac{1}{\cos\theta} \int\limits_0^{\infty} T(h)\alpha(h) \cdot \exp(-\frac{1}{\cos\theta} \int\limits_0^{h} \alpha(h')dh')dh, \tag{3}$$

$$\tau = \frac{1}{\cos\theta} \int\limits_0^{\infty} \alpha(h')dh', \tag{4}$$

where τ is an atmospheric optical thickness, h – a vertical coordinate, $T(h)$ – atmospheric temperature, $T_a^{\uparrow}, T_a^{\downarrow}$ are atmospheric upwelling and downwelling radiation at frequency ν correspondingly. The atmospheric constituents of the radiation are the functions of molecular oxygen (α_o), and water vapor (α_{wv}) and liquid water (α_{lw}) absorption coefficients : $\alpha = \alpha_o + \alpha_{wv} + \alpha_{lw}$. Their models, as in [18], [32], [44], [49], are taken from [50] for α_o and α_{lw}, and from [51] for α_{wv}. Effective surface radiation $T_{s\_eff}^{H,V}$ is defined by the parameters of the underlying surface. Since we consider the radiation of the sea ice – open ocean water – atmosphere system, it can be written as:

$$T_{s\_eff}^{H,V} = \sum_{n} C_n \cdot \varepsilon_{icen}^{H,V} \cdot T_{icen} + (1 - \sum_{n} C_n) \cdot \varepsilon_w^{H,V} \cdot T_w \qquad (5)$$

where C_n is a partial concentration of the sea ice of type n, $\varepsilon_{icen}^{H,V}$ - its emission coefficient

on H and V polarization, T_{icen} - its surface temperature, $(1 - \sum_{n} C_n)$ - open water concentration,

$\varepsilon_w^{H,V}$ - open sea water emission coefficient, T_w - sea surface temperature. Effective reflection

coefficient $r_{eff}^{H,V}$ in (1) can be written as:

$$r_{eff}^{H,V} = 1 - (\sum_{n} C_n \cdot \varepsilon_{icen}^{H,V} + (1 - \sum_{n} C_n) \cdot \varepsilon_w^{H,V}) \qquad (6)$$

The open water microwave emission $\varepsilon_w^{H,V}$ consists of two parts – calm salt water radiation

and wind-induced radiation [52]. The calm sea emission model is taken from [53] whereas the

wind induced signal is modeled differently for different channels. For the lower frequency

channels (in K and Ka bands) we follow [52] model, for the 89 GHz channels we use our recent

findings [49].

The sea ice emission coefficient is a function of many sea ice parameters. The most

significant factor governing sea ice microwave emission is the sea ice type. But such parameters

as snow presence and its properties, ice formation history and its current state also influence

$\varepsilon_{icen}^{H,V}$. In this study we use two sea ice types with distinct properties – first year ice (FY) and

multi-year ice (MY). The corresponding emission coefficients $\varepsilon_{FY}^{H,V}$ and $\varepsilon_{MY}^{H,V}$ are set

randomly within the ranges of variability taken from experimental and modeling studies [54],

[55]: $\varepsilon_{FY}^{V} \sim 0.93 - 0.91$; $\varepsilon_{FY}^{H} \sim 0.85 - 0.9$; $\varepsilon_{MY}^{V} \sim 0.93 - 0.55$; $\varepsilon_{MY}^{H} \sim 0.83 - 0.51$.

Calculations of $T_B^{H,V}$ are performed for all the AMSR frequencies at which microwave

radiation is sensitive to WVC, i.e. 18.7, 23.8, 36.5 and 89 GHz. Air temperature, humidity,

pressure and liquid water content profile data as well as T_w and T_{ice} data from the Era-Interim

reanalysis are used for calculations. The full range of parameter variability and

representativeness of the final match-up dataset of $T_B^{H,V}$ and WVC are ensured by the usage of

data for the whole year 2015 for latitudes higher than 65 degree with a grid of 2°×2°. Era-Interim data on sea ice concentration C are used to calculate surface emission. Partial concentrations of MY or FY sea ice type, C_{MY} and C_{FY}, are set up randomly for $C > 0$ so that $C_{MY} + C_{FY} = C$. Since the reanalyses tend to underestimate high values of sea surface wind speed [56], it is set randomly with uniform distribution within the range of 0 - 35 m/s for $C < 1$.

Then we added a normally distributed radiometric noise with 0.5 K equivalent temperature to the simulated T_B values. The WVC values are derived from the Era-Interim data by the integration of humidity profiles. The resulting matchup dataset consisted of 375 231 sets of eight T_Bs at 18.7, 23.8, 36.5 and 89.0 GHz on horizontal (H) and vertical (V) polarizations and the corresponding WVC value.

Then we applied an NN technique to establish the relationship between the AMSR T_Bs simulated measurements and the WVC values. An NN is able to approximate any continuous or nearly continuous dependence given a representative data set for NN training [57]. NN functions of Multilayer Perceptron (MLP) type have demonstrated their efficiency to solve remote sensing inverse problems [58] including WVC retrieval from satellite PM data [38], [59]. Performance of an NN-based algorithm depends primarily on the choice of a representative data set for NN training and an optimal configuration of the NN [60].

To invert simulated T_B values to WVC a standardNN of MLP type with feedforward backpropagation of errors is used. This task belongs to the simplest class of NN approximation. In [61] it is shown that a single layer feedforward NN with a sigmoid transfer function can perform any mapping given sufficient number of hidden neurons.

Since the ensemble of T_Bs and WVC is generated from the whole year of the Era-Interim data to account for seasonal variations we suppose that it represents the whole range of the parameter variability in the Arctic. Although contribution of the inter-annual parameter variability may increase this range, we believe that it is not reasonable to further enlarge the dataset.

The whole matchup dataset was randomly divided into two subsets consisting of 187287 and 187944 sets of the T_Bs and WVC values to use them for NN training and testing respectively. Utilizing this comparatively large amount of data ensures representing the majority of possible combination of atmospheric and oceanic parameters required for NN training and for reliable estimation of NN performance from its testing [62].

A neural network consists of a number of interconnected nodes, and an MLP generally consists of the three types of layers. The first layer of input nodes and the third layer of output nodes are represented in our case by the T_Bs and WVC respectively. The second layer is a hidden layer that transfers the signals modified with a non-linear sigmoid function from the input layer to the output layer. An increase in the number of hidden nodes enables an NN to learn more complex problems, but the capacity for generalization is reduced [57]. There are still no rules for the choice of NN complexity for various remote sensing problems. Practically an optimal NN configuration is searched from the numerous experiments of training and testing different NN models.

Since we consider a complicated underlying surface – atmosphere system, including sea ice and sea water surfaces, we tried various sets of NNs inputs. Involving the measurements at lower frequency channels (6.9 and 10.65 GHz) resulted in increase of the least root mean square difference σ, although they were supposed to be helpful for discrimination of sea ice and ocean surfaces due to their insensitivity to water vapor. The best theoretical results in terms of σ, calculated for the testing data set, are achieved when we use as input data the T_Bs measured at 18.7, 23.8, 36.5 and 89 GHz channels – altogether 8 inputs.

A hyperbolic tangent is used as a sigmoid function applied to the T_Bs, and the formula to calculate WVC with NN algorithm is:

$$WVC^{NN} = b_0 + w_0 \cdot \tanh(\sum_{n=1}^{N_{optimal}} (b_1 + w_{1n} \cdot \tanh(\sum_{i=1}^{8} (b_{2n} + w_{2in} \cdot T_{B_i})))) , \qquad (7)$$

The optimal number of nodes (neurons) in the hidden layer $N_{optimal}$ is determined from a series of experiments of NN training and testing. The process of training in the feedforward backpropagation NN consists of initialization of node weights, calculation of the output and comparison of the calculated output with the corresponding match-up parameter. The total root mean square error is then back-propagated through the network and the weights are altered so that to ensure the error decreases during the following processing of inputs. In our study we used a modified Levenberg-Marquardt method to solve a nonlinear least squares problem:

$$\sum_{k=1}^{M} (WVC^{NN}{}_k - WVC_k)^2 \rightarrow \min \ , \tag{8}$$

where M is the number of training data.

The process of an optimal NN development is shown in Fig. 1. At each step L the training is fulfilled a dozen times with different initial weights to avoid local minimum [57]. While the training error σ_{train} is gradually reduced with an increase of the hidden node number N, a decrease of the testing error σ_{test} slows down followed by its rise for N > 7. Thus the least σ_{test} of 1.2 kg/m^2 for WVC retrieval over the Arctic SI and OW surfaces corresponds to the $N_{optimal}= 7$ with fully connected input and hidden nodes. (Fig. 2). The NN weights are given in Table II.

We analyzed the algorithm sensitivity to the eight input $T_B s$ to check its robustness both to systematical biases and to random noise in channel measurements. As expected, the algorithm proved to be most sensitive to the vertically polarized measurements at 23.8 GHz (TB23V) near the water vapor resonant absorption line center at 22.234 GHz. Adding the systematical bias of 2 K [45] led to σ_{test} increase of ~ 50%. The influence of the random noise with the noise equivalent temperature (NET) of 2K at this channel is twice weaker: σ_{test} increased by 25%. Therefore errors in TB23V channel measurements lead to 1 kg/m^2 on each 1 K of the systematical bias and +0.5 kg/m^2 on each 1 K of NET. The other seven channels had much lower weighting factors in the algorithm with the errors from bias and random noise not exceeding 15%. Surprisingly the variations of the $T_B s$ at 89 GHz, sensitive to WVC, have the least effect on

the performance of the algorithm. This result may be considered as positive since the higher frequency channels have the largest NET. Thus adding the systematical biases and the random noise to the algorithm inputs results in remarkable bias in WVC retrievals only for the TB23V channel.

Though the error (σ_{test}) of 1.2 kg/m^2 seems to be low as compared with published retrieval errors for alternative methods and alternative instruments [19], [41], it may be far from actual one. The reasons are related to possible model inconsistencies and calibration issues as described in [63]. This error characterizes only the NN capability to reproduce well the dependency of WVC on $T_B s$, represented by model data. Therefore we analyzed the algorithm performance after its application to satellite measurements by comparison of WVC estimates with ground-truth data. This analysis is described in the following sections IV and V.

IV. ALGORITHM VALIDATION

A. *Validation with radiosonde data*

The NNs algorithm, described in Section III, is applied to the AMSR2 Level 1R T_B measurements at 18.7, 23.8, 36.5 and 89 GHz on vertical and horizontal polarization after accounting for model/calibration inconsistencies by adding a constant bias correction to T_B values, substantiated in [63]. Though these biases had been found for the AMSR-E instrument, they can be applied to the AMSR2 data as well because JAXA provides consistent AMSR data record based on cross calibration analysis that accounts for the differences in sensor properties.

The r/s WVC measurements from the Russian Arctic coastal radiosonde stations over the period of 2014 – 2017 are used to construct a database of matched up AMSR2 and r/s measurements. Although the r/s stations measure WVC twice a day we used only the measurements at 0:00 UTC for temporal collocation within 1 hour. To avoid land contamination

on the results of the algorithm application we took the AMSR2 measurements over the sea points in the vicinity of the stations. The distance between the station and selected point does not exceed 60 km, and is not less than 48 km (station 20674 on Fig. 3). The selected stations meet the following criteria to be chosen as a source of WVC ground-truth data: their elevation is not too high for inconsistencies; they provide measurements in the whole range of the Arctic WVC variability; the seas near the stations are covered by the sea ice in winter and spring seasons, thus providing WVC data both over the open water and over the sea ice We selected the r/s profiles with humidity measurements available at least up to 10 mB.

After initial quality control (screening physically not valid profiles of humidity) every WVC value derived from the r/s profile is matched-up with the AMSR2 measurements over the selected points near the station. Locations of these points and stations are presented in Fig. 3 along with an example of WVC field on March 11, 2014, retrieved with the developed algorithm. The selected AMSR2 measurement points and the r/s station positions are marked with green and red circles 'respectively. The WVC field, presented in the figure, are retrieved from the AMSR2 descending path 183D that covers all four stations few minutes after midnight, and is used for comparison with the r/s measurements taken at 0:00 UTC.

The WVC distribution derived from the 1852 selected r/s profiles covers the full range of the WVC variability (Fig. 4a). The largest WVC values are observed in summer months over OW surfaces, defined as the areas with SIC < 10%. Most of the r/s measurements over SI (SIC > 90%) do not exceed 10 kg/m^2 with a maximum number of samples at 3-4 kg/m^2.

The scatter plot of WVC, retrieved from the AMSR2 data with the new algorithm (WVC$_{AMSR2}$), versus WVC, measured by radiosondes (WVC$_{r/s}$), is presented in Fig. 5. The correlation coefficient R = 0.96, the slope is 0.9, the intercept is 0.6 kg/m^2.The total root mean square error of 1.5 kg/m^2 is calculated as:

$$\sigma_{AMSR2,r/s} = \sqrt{\frac{\sum_i (WVC^i_{AMSR2} - WVC^i_{r/s})^2}{N-1}}, \tag{9}$$

where i is the number of a match-up in the dataset. The highest difference between WVC$_{AMSR2}$ and WVC$_{r/s}$ (up to 5 kg/m^2) is observed over open water providing the main contribution to the error. Considering the WVC only over SI reduces the $\sigma_{AMSR2,r/s}$ to 1 kg/m^2. The relative error σ_{rel}, defined as $\sigma_{AMSR2,r/s}$ divided by the mean WVC, is 20% over OW and 25% over SI. Overall, the algorithm provides reasonable estimates both over SI and OW surfaces, although slight underestimation of the WVC values higher than 25 kg/m^2 can be noted from Fig. 5. The largest deviations of the WVC$_{AMSR2}$ from the WVC$_{r/s}$ occur for the cases with time difference exceeding 40 minutes that, obviously, can be explained by the fast changes in the atmosphere during this time.

The amount of r/s data we use for validation is not sufficient to draw solid conclusion about the algorithm accuracy under various Arctic conditions. Reanalysis data present a consistent and numerous dataset that may serve as a source of ground-truth data. Therefore we have validated the algorithm also with the European Centre for Medium - Range Weather Forecasts (ECMWF) Interim Reanalysis (Era-Interim) data on WVC. To ensure that the WVC from the Era-Interim Reanalysis can be used as ground-truth data we first investigate its accuracy in the Arctic region.

B. Validation of Era-Interim reanalysis WVC with radiosonde data

Reanalysis data present consistent global data sets with hundreds of available parameters. Reanalysis process tremendous number of observation data, assimilating them into a consistent database that allows studying various processes. Though reanalysis data sets are based both on in-situ data and models, they are often considered as observational data sets. In contrast, some studies discuss the caution needed when replacing observation data by reanalysis data [64]. It is generally recognized that the accuracy of reanalysis data on geophysical parameters needs to be estimated before exploiting them for scientific studies [65]. Accuracy of data is typically

estimated with some ground-truth data [66]. To some extent the reanalysis data are ground-truth themselves, but obviously this is not the case for remote Arctic areas.

The Era-Interim reanalysis is produced at a horizontal spatial resolution of about 79 km. It incorporates an improved representation of the hydrological cycle and 4D-Var assimilation scheme of WVC satellite retrievals. The largest amount of data comes from polar-orbiting and geostationary satellite observations, which are mostly assimilated as brightness temperatures. A part of assimilated data are the ground-truth WVC measured from radiosondes, GPS, ships, weather stations etc. The accuracy of the WVC Era-Interim reanalysis was regionally estimated with ground-based measurements in many studies. This data set is confirmed to be one of the most homogeneous and consistent among the modern reanalysis. In [67] the authors compared the Era-Interim WVC with WVC from the r/s, radio occultation and GPS measurements and found high consistency and no significant biases on a global scale. Despite some shortcomings in the near - surface variables, this reanalysis is the best of available for the Arctic region [68]

To evaluate an accuracy of the Era-Interim data on WVC in the Arctic region we use the same r/s stations for the period of 2014 – 2017 and both 'times of observation: 0:00 UTC and 12:00 UTC.

We found almost no systematical bias between WVC_{Era-I} and $WVC_{r/s}$ using 8339 collocated estimates. The scatter plot of the WVC_{Era-I} versus $WVC_{r/}$ and their monthly means along with the root mean square error $\sigma_{ERA-I,r/s}$ are presented in Fig.6. The $\sigma_{ERA-I,r/s}$ is 1.6 kg/m^2 for the whole data set, that corresponds to $\sigma_{rel} \sim 20\%$, and slightly lower in winter ($\sim 0.7 - 1$ kg/m^2) than in summer ($2 - 2.5$ kg/m^2). 97% of WVC_{Era-I} data do not scatter more than $\pm 2\sigma_{Era-I,r/s}$, and 85% of data fall within $\pm\sigma_{Era-I,r/s}$. These characteristics confirm that accuracy of the WVC Era-Interim reanalysis in the Arctic region is sufficient for using this data set as a source of vast benchmark validation WVC data. Apparently the reanalysis WVC data are not independent from the r/s measurements, assimilated during reanalysis process. However the atmospheric models with their biases, the errors, induced by the assimilation schemes and the number of

observations varying in space and time affect the reanalysis quality [2] thereby justifying estimation of their accuracies.

C. Validation with Era-Interim data

Validation of the algorithm with the Era-Interim WVC data is performed for the period of one year. We applied the algorithm both to the AMSR2 and to the AMSR-E data. Comparison of the AMSR-E (WVC$_{AMSR-E}$) and AMSR2 (WVC$_{AMSR2}$) retrieved WVC with WVC$_{Era-I}$ was done for daily averaged data for the period of 2008 and 2015 respectively. The year of 2008 was selected because the alternative satellite product, based on the AMSU-B satellite data [41], is available for that period allowing comparing the accuracy of our WVC estimates with that products, presented in the next section.

The algorithm is applied to the AMSR-E and AMSR2 daily averaged Level 3 T_B data and the results are gridded on the Era-Interim grid of 0.75°×0.75°. The SIC data of the University of Bremen, also gridded on the Era-Interim grid, are used to identify the surface type.

The Era-Interim reanalysis provided much larger amount of WVC validation data than r/s measurements. For example for the period of the year 2015 we have 2 380 351 valid WVC values over the whole Arctic defined as the ocean area north of 65° (SI and OW surfaces) against 1852 r/s values over 4 stations for ~ 4 years. Fig. 4(b) shows WVC distribution over SI and OW surfaces according to the Era-Interim reanalysis data for the period of 2015. In particular, the large WVC values of more than 10 kg/m^2 over SI are better represented in the Era-Interim data comprising 17% of the dataset, while only 3% of such samples are available in the r/s data. We analyzed the algorithm performance for the total matched-up dataset as well as for the monthly means. To assess the significance of the WVC retrievals we compared the root mean square error $\sigma_{AMSR2,Era-I}$ estimated as in (9) with the WVC standard deviation $\sigma_{natural}$, that is the characteristic of the range of WVC natural variability, calculated as

$$\sigma_{natural} = \sqrt{\frac{\sum_{i=1}^{N}(WVC_i - <WVC>)^2}{N}} \, , \tag{10}$$

where N is the number of WVC values within the whole year and monthly datasets. Any algorithm provides valuable knowledge about a geophysical parameter only if the natural variability is larger than the retrieval error: in the equation $\sigma_{natural} = k \ \sigma_{AMSR2,Era-I}$, $k > 1$. For the AMSR-2 retrievals we found that for the whole Arctic the monthly k values range between 2.1 and 3.5, being the lowest in July and August when high WVC values are observed. For the whole match up data set the $\sigma_{AMSR2,Era-I}$ is 1.07 kg/m^2 corresponding to a relative error $\sigma_{rel} = \sigma_{AMSR2,Era-I}/<WVC_{Era-I}> \sim 15\%$. The correlation coefficient R is 0.98 with insignificant difference between the mean WVC values of -0.07 kg/m^2..

The scatter plots of the AMSR2 and AMSR-E retrieved WVC versus Era-Interim WVC as well as corresponding statistics separately for SI and OW surfaces are presented in Fig. 7 and 8, and in Table III. The results of the algorithm application to the AMSR2 and AMSR-E data are quite similar except slightly enhanced errors for the WVC$_{AMSR-E}$ as compared to WVC$_{AMSR2}$. This is likely caused by the differences in the atmospheric and oceanic conditions in 2008 and 2015.

Fig. 9 shows annual distribution of WVC and retrieval errors over SI and OW surfaces along with error bars illustrating the range of WVC variability. In terms of the absolute retrieval error there is no significant difference in the algorithm performance over SI and OW, except in summer months, when they are higher by $\sim 0.3 - 0.5$ kg/m^2 over SI. But the natural variability of WVC over the sea ice is lower, so that k =1.5 – 2.5 throughout a year, whereas over OW k = 4-5, ensuring more valued information about WVC The corresponding comparison for the AMSR-E WVC retrievals is presented in the next section along with the comparison with the other WVC products.

We also checked the consistency of the new algorithm with that developed previously for open water areas [32] (ZC$_{OW}$). One could expect that the ZC$_{OW}$ algorithm should have higher accuracy than the newly developed as it does not use measurements at the 89 GHz channel. Nevertheless, as compared with the Era-Interim WVC data, the ZC$_{OW}$ does not exhibit any noticeable advantages either in terms of the error or in terms of bias and correlation throughout the whole year (Table III). The only small advantage of ZC$_{OW}$ is related to lower number of the WVC estimates that differ from the ERA-Interim data by more than $\sigma_{AMSR2,Era-I}$ i.e. when $|\text{WVC}_{\text{AMSR2}} - \text{WVC}_{\text{Era-I}}| > \sigma_{AMSR2,Era-I}$. The percentage of such data for the new algorithm is 23%, whereas for ZC$_{OW}$ it is 21%. The difference of 2% we attribute to weather events, associated with high liquid or ice water content in the atmosphere, possible scattering of microwave radiation at 89 GHz, not accounted by physical modeling and thus deteriorating the new algorithm performance.

V. COMPARISON WITH OTHER SATELLITE WVC PRODUCTS IN THE ARCTIC

The main advantage of the newly developed algorithm is its ability to retrieve WVC not only over OW but also over SI. To evaluate the benefits of this algorithm we validated the MODIS-based and AMSU-B-based satellite products on WVC over the Arctic SI and OW surfaces against the data from the Era-Interim reanalysis in a similar way as for the AMSR retrievals in section IV. For analysis we use the period of the year 2008, for which all these datasets are available.

The daily WVC product derived from the AMSU-B measured radiances (WVC$_{\text{AMSU-B}}$) provides the WVC estimates valid over SI and OW within the range of 0 - 15 kg/m^2 [41]. We used the latest AMSU-B product generated with the new version of the algorithm that was found to be the most accurate. The WVC derived from the Terra MODIS measurements (WVC$_{\text{MODIS}}$) is a daily product though containing limited amount of data due to significant cloud coverage on

the MODIS scenes. Both the WVC products are gridded on the Era-Interim reanalysis grid for pixel to pixel comparison as well as the WVC retrievals from the AMSR-E data.

Also the same SIC product provided by the University of Bremen [47] is used to distinguish WVC data over SI and OW surfaces.

A. Comparison with MODIS WVC product

The results of the comparison of the MODIS WVC and the AMSR-E retrievals with the Era-Interim data over the Arctic region are presented in Figs. 10 and 11, and in Table IV. The MODIS WVC data are much more scattered than those retrieved from AMSR-E measurements both over SI and over OW surfaces as illustrated on the scatter plots in Figs. 8 and 10 and indicated from the comparison of the root mean square errors and correlation coefficients (Tables III and IV). In addition, the bias of -1.1 kg/m^2 as well as the scatter plots in Fig. 10 reveal overall underestimation of WVC in the MODIS product, that can be particularly attributed to excluding the scenes with cloud cover.

The annual behavior of the monthly mean WVC values (<WVC$_{MODIS}$> and <WVC$_{Era-I}$>) and the root mean square errors $\sigma_{MODIS,Era-I}$ and $\sigma_{AMSR-E,Era-I}$ are shown in Fig.11. The largest underestimation of the WVC$_{MODIS}$ as compared to the WVC$_{Era-I}$ is observed in summer, especially over SI surface where the bias exceeds 3 kg/m^2, and the errors are higher than the natural variability of WVC. In contrast, $\sigma_{AMSR-E,Era-I}$ is as low as ~ 1.2 – 1.3 kg/m^2 and no systematical bias is observed for WVC$_{AMSR-E}$ during the whole year. In fact WVC$_{MODIS}$ can be considered as a reliable WVC product only in winter, although slight underestimation over SI and overestimation over OW are still observed even during winter months.

B. Comparison with AMSU-B WVC product

The four-month AMSU-B product (March, June, September and October) available in the product is used for the comparison with the Era-Interim WVC data. The results, presented in Table IV and Figs. 12 and 13, show that despite a small overall bias, the large root mean square errors and low correlation are observed, especially, in summer months (June and September). For example, in September $\sigma_{AMSU-B,Era-I} = 4.3$ kg/m^2 and R = 0.15 .

The scatter plots in Fig 12 show that for the Era-Interim WVC higher than $\sim$ 7 kg/m^2 the AMSU-B estimates are significantly deviate from the 1:1 line: over OW the WVC is underestimated, while over SI it is overestimated for the WVC$_{Era-I}$ within the range of 7-12 kg/m^2 and underestimated for the WVC$_{Era-I}$ higher than 15 kg/m^2. Although excluding WVC$_{Era-I}$ exceeding 15 kg/m^2 results in notable improvement of the AMSU-B product accuracy over OW (by ~30%), the observed underestimation is still remarkable. Besides, in this case, a significant portion of the Arctic region remains uncovered by WVC retrievals. The largest discrepancies in the monthly averaged WVC values (see Fig. 13) are found for June and September: the WVC$_{AMSU-B}$ estimates are overestimated over SI and underestimated over OW in contrast with a good agreement between the AMSR-E retrievals and the Era-Interim data. The lower values of the $\sigma_{AMSR-E,Era-I}$, especially for June and September are also among the advantages of the new algorithm.

Figs. 14 and 15 illustrate spatial distribution of WVC from the ERA-Interim reanalysis and its difference from the satellite WVC products for March 13, 2008 and June 3, 2008 respectively. The fields of the WVC differences illustrate general features of the products typical for winter and summer. 1) WVC$_{MODIS}$ is mostly reliable in winter but large areas over OW are not covered by data due to cloud cover. In summer this product has large errors both over SI and over OW; 2) WVC$_{AMSU-B}$ in winter is accurate over SI, but significantly underestimates WVC$_{Era-I}$ over OW areas with comparatively high WVC. In summer this underestimation is more dramatic and can exceed for some areas the average value of $\sigma_{AMSU-B,Era-I}$ in several times. The overestimation over SI in summer is also evident almost for the whole Arctic area. 3) WVC$_{AMSR-E}$ is highly

accurate in winter both over SI and over OW. In summer deviations from WVC_{Era-I} increase but remain lower than the deviations of the alternative products.

VI. CONCLUSION

In attempts to fill the gap in the Arctic WVC data a new algorithm for WVC retrieval from the AMSR instruments (AMSR-E and AMSR2) is developed. The algorithm presents an advancement of the previous algorithms developed in [18], [32], applicable over open oceans only. The main advantage of the new algorithm consists in the expansion of the applicability area to sea ice surfaces, thus covering the most of the Arctic region. The algorithm is based on the results of the numerical simulation of the AMSR measured brightness temperatures over sea ice and open water areas. The inversion method is based on a Neural Networks approach. Extensive validation of the algorithm was done using as benchmark data sets the ERA-Interim reanalysis and radiosonde measurements for the whole range of WVC variability in the Arctic region. The validation of the algorithm applied to the AMSR2 data using the r/s measurements from the four Arctic coastal stations during 2014 – 2017 proved the high accuracy of the WVC retrievals both over SI and OW with the overall root mean square error of 1.5 kg/m^2. Before using the ERA-Interim reanalysis as a validation dataset, it was also compared with the same r/s data. This comparison revealed root mean square error of 1.6 kg/m^2 and no systematical bias. It made possible to validate the algorithm exploiting the Era-Interim WVC data over the entire Arctic region for a whole year both for the AMSR2 and for the AMSR-E retrievals. We found high correlation of the retrieved WVC (0.96 – 0.99) and the absence of WVC bias both over SI and over OW for both instruments. The root mean square errors for SI and OW surfaces are close and vary within the range of 0.9 – 1.2 kg/m^2, but the relative error is almost 3 times higher over SI. We also checked the performance of the new algorithm in terms of consistency with that developed previously and applicable only over OW. We did not find any substantial difference

between the results. So we may conclude that the new algorithm does not deteriorate WVC retrievals over OW but offers the possibility to estimate WVC over SI. As expected variability of WVC over SI is low as compared to that over OW. Therefore a significance of the AMSR derived retrievals over SI (assessed as a relation of the WVC natural variability to the retrieval error) is 3-4 times lower than over OW. Nevertheless, these are the best basin-wide WVC estimates among currently available datasets for the Arctic region. The validation of the two alternative satellite WVC products, based on the AMSU-B and MODIS measurements against the ERA-Interim data for 2008 along with the validation of AMSR-E retrievals showed the advantages of the new method. We found that both alternative products showed high errors for comparatively wet atmosphere, mostly during summer months, when the retrieval errors started to be of the order or exceeded the value of WVC natural variability. Although the errors for the MODIS product and the AMSR-E estimates in winter are similar, the underestimation of WVC_{Era-I} by the MODIS data is notable throughout the year. Moreover, MODIS WVC retrievals are limited to cloud free regions. As for the AMSU-B based WVC, we found a significant overestimation and underestimation over SI and OW respectively as compared to WVC_{Era-I}.

The strength of the conclusions, based on the validation with the Era-Interim reanalysis data may be called into question since reanalysis data assimilate in particular, satellite passive microwave measurements. But more accurate WVC data over the vast Arctic sea ice areas are apparently lacking. The new algorithm provides new measures to study the Arctic atmospheric water vapor distribution and variability and may suggest a new stage of satellite passive microwave data exploration.

[1] K. E. Trenberth, Y. Zhang, J. T. Fasullo, and S. Taguchi, "Climate variability and relationships between top-of-atmosphere radiation and temperatures on Earth," *J. Geophys. Res. Atmospheres*, vol. 120, no. 9, pp. 3642–3659, May 2015.

[2] M. Schröder, M. Lockhoff, F. Fell, J. Forsythe, T. Trent, R. Bennartz, E. Borbas, M. G. Bosilovich, E. Castelli, and H. Hersbach, "The GEWEX Water Vapor Assessment archive of water vapour products from satellite observations and reanalyses," *Earth Syst. Sci. Data*, vol. 10, pp. 1093–1117, 2018.

[3] D. Alraddawi, P. Keckhut, A. Sarkissian, O. Bock, A. Irbah, S. Bekki, C. Claud, and M. Meftah, "Enhanced MODIS Atmospheric Total Water Vapour Content Trends in Response to Arctic Amplification," *Atmosphere*, vol. 8, no. 12, p. 241, 2017, doi: 10.3390/atmos8120241

[4] F. Alshawaf, G. Dick, S. Heise, T. Simeonov, S. Vey, T. Schmidt, and J. Wickert, "Decadal variations in atmospheric water vapor time series estimated using ground-based GNSS," *Atmos. Meas. Tech. Discuss.*, 2017, doi: 10.5194/amt-2016-151.

[5] S.-P. Ho, L. Peng, C. Mears, and R. A. Anthes, "Comparison of global observations and trends of total precipitable water derived from microwave radiometers and COSMIC radio occultation from 2006 to 2013.," *Atmospheric Chem. Phys.*, vol. 18, no. 1, 2018.

[6] R. Lindstrot, M. Stengel, M. Schröder, J. Fischer, R. Preusker, N. Schneider, T. Steenbergen, and B. R. Bojkov, "A global climatology of total columnar water vapour from SSM/I and MERIS," *Earth Syst. Sci. Data*, vol. 6, pp. 221-233, 2014.

[7] S. Mieruch, S. Noël, H. Bovensmann, and J. P. Burrows, "Analysis of global water vapour trends from satellite measurements in the visible spectral range," *Atmospheric Chem. Phys.*, vol. 8, no. 3, pp. 491–504, 2008.

[8] T. Nilsson and G. Elgered, "Long-term trends in the atmospheric water vapor content estimated from ground-based GPS data," *J. Geophys. Res. Atmospheres*, vol. 113, no. D19, 2008, doi: 10.1029/2008JD010110.

[9] W. Peng, X. Tongchuan, D. Jiageng, S. Jingmin, W. Yanling, S. Qingli, D. Xin, Y. Hongliang, S. Dejun, and Z. Jinrong, "Trends and Variability in Precipitable Water Vapor throughout North China from 1979 to 2015," *Advances in Meteorology*, vol. 2017, Article ID 7804823, 10 pages, 2017. doi: 10.1155/2017/7804823.

[10] K. E. Trenberth, J. Fasullo, and L. Smith, "Trends and variability in column-integrated atmospheric water vapor," *Clim. Dyn.*, vol. 24, no. 7–8, pp. 741–758, 2005.

[11] J. Wang, A. Dai, and C. Mears, "Global Water Vapor Trend from 1988 to 2011 and Its Diurnal Asymmetry Based on GPS, Radiosonde, and Microwave Satellite Measurements," *J. Clim.*, vol. 29, no. 14, pp. 5205–5222, Apr. 2016.

[12] Y. Zhang, J. Xu, N. Yang, and P. Lan, "Variability and Trends in Global Precipitable Water Vapor Retrieved from COSMIC Radio Occultation and Radiosonde Observations," *Atmosphere*, vol. 9, no. 5, p. 174, 2018.

[13] T. Vihma, "Effects of Arctic sea ice decline on weather and climate: A review," *Surv. Geophys.*, vol. 35, no. 5, pp. 1175–1214, 2014.

[14] L. M. Miloshevich, H. Vömel, D. N. Whiteman, and T. Leblanc, "Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements," *J. Geophys. Res.*, vol. 114, no. D11, p. D11305, 2009.

[15] A. P. Ferreira, R. Nieto, and L. Gimeno, "Completeness of radiosonde humidity observations based on the IGRA," *Earth Syst. Sci. Data Discuss.*, 2018, doi: 10.5194/essd-2018-95.

[16] A. Dai, J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang, "A new approach to homogenize daily radiosonde humidity data," *J. Clim.*, vol. 24, no. 4, pp. 965–991, 2011.

[17] I. Moradi, B. Soden, R. Ferraro, P. Arkin, and H. Vömel, "Assessing the quality of humidity measurements from global operational radiosonde sensors," *J. Geophys. Res. Atmospheres*, vol. 118, no. 14, pp. 8040–8053, 2013.

[18] L. P. Bobylev, E. V. Zabolotskikh, L. M. Mitnik, and M. L. Mitnik, "Atmospheric water vapor and cloud liquid water retrieval over the Arctic Ocean using satellite passive microwave sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 283–294, 2010.

[19] B.-C. Gao and Y. J. Kaufman, "Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels," *J. Geophys. Res. Atmospheres*, vol. 108, no. D13, p. 4389, Jul. 2003.

[20] J. Miao, K. Kunzi, G. Heygster, T. A. Lachlan-Cope, and J. Turner, "Atmospheric water vapor over Antarctica derived from Special Sensor Microwave/Temperature 2 data," *J. Geophys. Res. Atmospheres*, vol. 106, no. D10, pp. 10187–10203, 2001.

[21] S. W. Seemann, J. Li, W. P. Menzel, and L. E. Gumley, "Operational retrieval of atmospheric temperature, moisture, and ozone from MODIS infrared radiances," *J. Appl. Meteorol.*, vol. 42, no. 8, pp. 1072–1091, 2003.

[22] I. Durre, C. N. Williams, X. Yin, and R. S. Vose, "Radiosonde-based trends in precipitable water over the Northern Hemisphere: An update," *J. Geophys. Res. Atmospheres*, vol. 114, no. D5, 2009.

[23] S. Businger, S. R. Chiswell, M. Bevis, J. Duan, R. A. Anthes, C. Rocken, R. H. Ware, M. Exner, T. VanHove, and F. S. Solheim, "The promise of GPS in atmospheric monitoring," *Bull. Am. Meteorol. Soc.*, vol. 77, no. 1, pp. 5–18, 1996.

[24] J. Berckmans, R. V. Malderen, E. Pottiaux, R. Pacione, and R. Hamdi, "Validating the water vapour content from a reanalysis product and a regional climate model over Europe based on GNSS observations," *Atmos. Chem. Phys.*, 2019, doi: 10.5194/acp-2018-1097.

[25] K. de Jong, M. Goode, X. Liu, and M. Stone, "Precise GNSS positioning in Arctic regions," in *OTC Arctic Technology Conference*, 2014, doi: 10.4043/24651-MS.

[26] M. Zygmuntowska, T. Mauritsen, J. Quaas, and L. Kaleschke, "Arctic Clouds and Surface Radiation–a critical comparison of satellite retrievals and the ERA-Interim reanalysis," *Atmospheric Chem. Phys.*, vol. 12, no. 14, pp. 6667–6677, 2012.

[27] H. Ren, C. Du, R. Liu, Q. Qin, G. Yan, Z. Li, J. Meng, "Atmospheric water vapor retrieval from Landsat 8 thermal infrared images," *J. Geophys. Res. Atmospheres*, vol. 120, no. 5, pp. 1723–1738, 2015.

[28] S. Mieruch, M. Schröder, S. Noël, and J. Schulz, "Comparison of monthly means of global total column water vapor retrieved from independent satellite observations," *J. Geophys. Res. Atmospheres*, vol. 115, no. D23, 2010, doi: 10.1029/2010JD013946.

[29] M. Shangguan, S. Heise, M. Bender, G. Dick, M. Ramatschi, and J. Wicert, "Validation of GPS atmospheric water vapor with WVR data in satellite tracking mode," in *Annales of Geophysics*, 2015, vol. 33, pp. 55–61.

[30] X. Yu, F. Xie, and C. O. Ao, "Evaluating the lower-tropospheric COSMIC GPS radio occultation sounding quality over the Arctic," *Atmospheric Meas. Tech.*, vol. 11, no. 4, pp. 2051–2066, Apr. 2018.

[31] F. J. Wentz, "A model function for ocean microwave brightness temperatures," *J Geophys Res*, vol. 88, no. C3, pp. 1892–1908, 1983.

[32] E. V. Zabolotskikh and B. Chapron, "Improvements in Atmospheric Water Vapor Content Retrievals Over Open Oceans From Satellite Passive Microwave Radiometers," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 7, pp. 3125–3133, 2017.

[33] J. C. Alishouse, S. A. Snyder, J. Vongsathorn, and R. R. Ferraro, "Determination of oceanic total precipitable water from the SSM/I," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 5, pp. 811–816, 1990.

[34] M. Kazumori, "A retrieval algorithm of atmospheric water vapor and cloud liquid water for AMSR-E," *Eur. J. Remote Sens.*, vol. 45, pp. 63–74, Mar. 2012.

[35] Y. Wang, Y. Fu, G. Liu, Q. Liu, and L. Sun, "A new water vapor algorithm for TRMM Microwave Imager (TMI) measurements based on a log linear relationship," *J. Geophys. Res. Atmospheres*, vol. 114, no. D21, p. D21304, Nov. 2009.

[36] F. J. Wentz, "A well-calibrated ocean algorithm for special sensor microwave/imager," *J Geophys Res*, vol. 102, pp. 8703–8718, 1997.

[37] M. N. Deeter, "A new satellite retrieval method for precipitable water vapor over land and ocean," *Geophys. Res. Lett.*, vol. 34, no. 2, p. L02815, Jan. 2007.

[38] F. Aires, C. Prigent, W. B. Rossow, and M. Rothstein, "A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations," *J. Geophys. Res.*, vol. 106, no. D14, pp. 14887–14907, 2001.

[39] J. Du, J. S. Kimball, and L. A. Jones, "Satellite Microwave Retrieval of Total Precipitable Water Vapor and Surface Air Temperature Over Land From AMSR2," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2520–2531, May 2015.

[40] F. J. Wentz and T. Meissner, "Algorithm Theoretical Basis Document (ATBD), Version 2, AMSR Ocean Algorithm, RSS Tech. Proposal 121599A-1, Remote Sensing Systems," *Algorithm Theor. Basis Doc. ATBD Version 2 AMSR Ocean Algorithm RSS Tech Propos. 121599A-1 Remote Sens. Syst.*, 2000.

[41] R. C. Scarlat, C. Melsheimer, and G. Heygster, "Retrieval of total water vapour in the Arctic using microwave humidity sounders," *Atmospheric Meas. Tech.*, vol. 11, no. 4, pp. 2067–2084, 2018.

[42] C. Melsheimer and G. Heygster, "Improved retrieval of total water vapor over polar regions from AMSU-B microwave radiometer data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2307–2322, 2008.

[43] D. P. Dee, S. M. Uppala, A. J. Simmons, Paul Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, *et al.*, "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system," *Q. J. R. Meteorol. Soc.*, vol. 137, no. 656, pp. 553–597, 2011.

[44] E. V. Zabolotskikh, L. M. Mitnik, and B. Chapron, "New approach for severe marine weather study using satellite passive microwave sensing," *Geophys. Res. Lett.*, vol. 40, no. 13, pp. 3347–3350, 2013.

[45] K. Imaoka, M. Kachi, M. Kasahara, N. Ito, K. Nakagawa, and T. Oki, "Instrument performance and calibration of AMSR-E and AMSR2," *Int. Arch. Photogramm. Remote Sens. Spec. Inf. Sci.*, vol. 38, no. Part 8, pp. 13–18, 2010.

[46] N. P. Romanov, "A new formula for saturated water steam pressure within the temperature range -25 to 220°C," *Izvestiya. Atmospheric and Oceanic Physics*, vol. 45, no. 5, pp. 799–804, 2009.

[47] G. Spreen, L. Kaleschke, and G. Heygster, "Sea ice remote sensing using AMSR-E 89-GHz channels," *J. Geophys. Res. Oceans 1978–2012*, vol. 113, no. C2, 2008.

[48] S. Yu. Matrosov and E. M. Shulgina, "Scattering and attenuation of microwave radiation by precipitation," *MGO Trans.*, vol. 448, pp. 85–94, 1982. (in Russian)

[49] E. V. Zabolotskikh and B. Chapron, "New Geophysical Model Function for Ocean Emissivity at 89 GHz Over Arctic Waters," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 573–576, 2018.

[50] H. J. Liebe and D. H. Layton, "Millimeter-wave properties of the atmosphere: Laboratory studies and propagation modeling," Nat. Tech. Inf. Service Boulder, CO, NTIA Rep. 87-24, 1987.

[51] D. D. Turner, M. P. Cadeddu, U. Lohnert, S. Crewell, and A. M. Vogelmann, "Modifications to the water vapor continuum in the microwave suggested by ground-based 150-GHz observations," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 47, no. 10, pp. 3326–3337, 2009.

[52] T. Meissner and F. J. Wentz, "The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and earth incidence angles," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3004–3026, 2012.

[53] Q. Liu, F. Weng, and S. J. English, "An Improved Fast Microwave Water Emissivity Model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1238–1250, 2011.

[54] NORSEX Group, "Norwegian remote sensing experiment in a marginal ice zone," *Science*, vol. 220, no. 4599, pp. 781–787, 1983.

[55] S. Willmes, M. Nicolaus, and C. Haas, "The microwave emissivity variability of snow covered first-year sea ice from late winter to early summer: a model study," *The Cryosphere*, vol. 8, no. 3, pp. 891–904, 2014.

[56] M. Li, J. Liu, Z. Wang, H. Wang, Z. Zhang, L. Zhang, and Q. Yang, "Assessment of Sea Surface Wind from NWP Reanalyses and Satellites in the Southern Ocean," *J. Atmospheric Ocean. Technol.*, vol. 30, no. 8, pp. 1842–1853, May 2013.

[57] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.

[58] V. M. Krasnopolsky, "Neural network emulations for complex multidimensional geophysical mappings: Applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling," *Rev. Geophys.*, vol. 45, no. 3, p. RG3009, Sep. 2007.

[59] C. Mallet, E. Moreau, L. Casagrande, and C. Klapisz, "Determination of integrated cloud liquid water path and total precipitable water from SSM/I data using a neural network algorithm," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 661–674, 2002.

[60] P. M. Atkinson and A. R. L. Tatnall, "Introduction Neural networks in remote sensing," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 699–709, 1997.

[61] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.

[62] Y. Xu and R. Goodacre, "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning," *J. Anal. Test.*, vol. 2, no. 3, pp. 249–262, 2018.

[63] E. V. Zabolotskikh, L. M. Mitnik, and B. Chapron, "An Updated Geophysical Model for AMSR-E and SSMIS Brightness Temperature Simulations over Oceans," *Remote Sens.*, vol. 6, no. 3, pp. 2317–2342, 2014.

[64] M. G. Bosilovich, J. Kennedy, D. Dee, R. Allan, and A. O'Neill, "On the reprocessing and reanalysis of observations for climate," in *Climate Science for Serving Society*, Springer, 2013, pp. 51–71.

[65] W. S. Parker, "Reanalyses and observations: What's the difference?," *Bull. Am. Meteorol. Soc.*, vol. 97, no. 9, pp. 1565–1572, 2016.

[66] M. Peña and Z. Toth, "Estimation of analysis and forecast error variances," *Tellus Dyn. Meteorol. Oceanogr.*, vol. 66, no. 1, p. 21767, 2014.

[67] Q. Zhang, J. Ye, S. Zhang, and F. Han, "Precipitable Water Vapor Retrieval and Analysis by Multiple Data Sources: Ground-Based GNSS, Radio Occultation, Radiosonde, Microwave Satellite, and NWP Reanalysis Data," *Journal of Sensors*, 2018. [Online]. Available: https://www.hindawi.com/journals/js/2018/3428303/. [Accessed: 11-Feb-2019].

[68] E. Jakobson, T. Vihma, T. Palo, L. Jakobson, H. Keernik, and J. Jaagus, "Validation of atmospheric reanalyses over the central Arctic Ocean," *Geophys. Res. Lett.*, vol. 39, no. 10, 2012.

Figures.
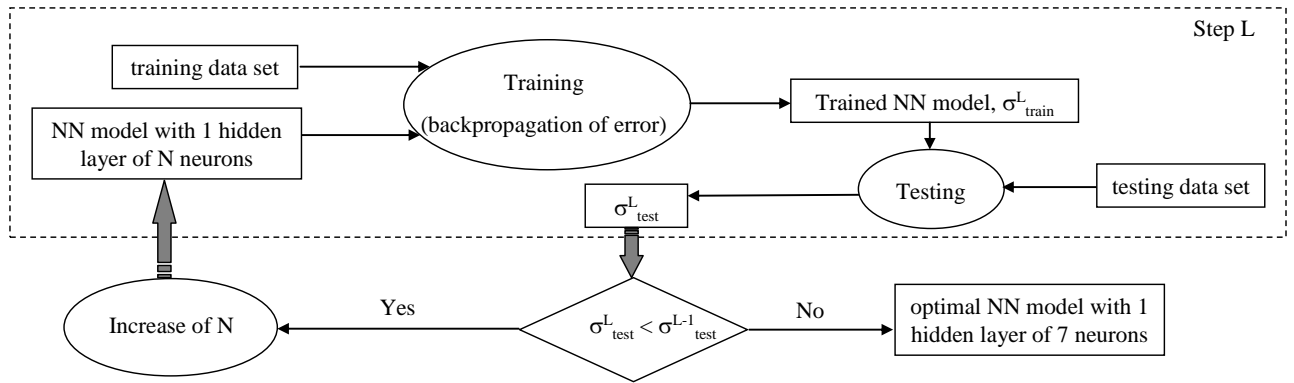
Fig. 1. A flow diagram illustrating the process of an optimal NN development.
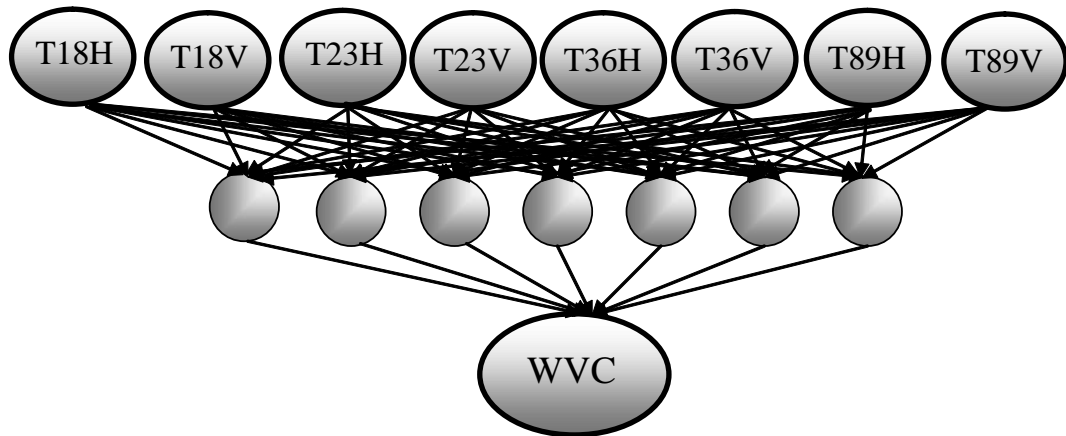
Fig. 2. Optimal Neural Network for WVC retrieval over the Arctic SI and OW surfaces. T18H, T18V, T23H, T23V, T36H, T36V, T89H, T89V denote brightness temperatures, measured at 18.7, 23.8, 36.5 and 89.0 GHz on horizontal (H) and vertical (V) polarizations.
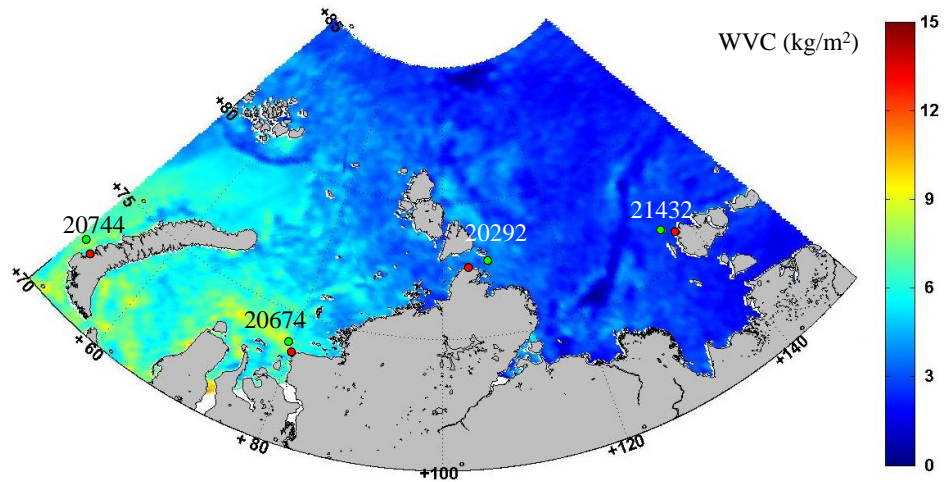
Fig. 3. Example of WVC field, retrieved with the developed algorithm from the AMSR2 Level 1R T_B in the area of the Russian coastal r/s stations on March 11, 2014. The AMSR2 measurement points are marked with green circles, the station positions are marked with red circles.
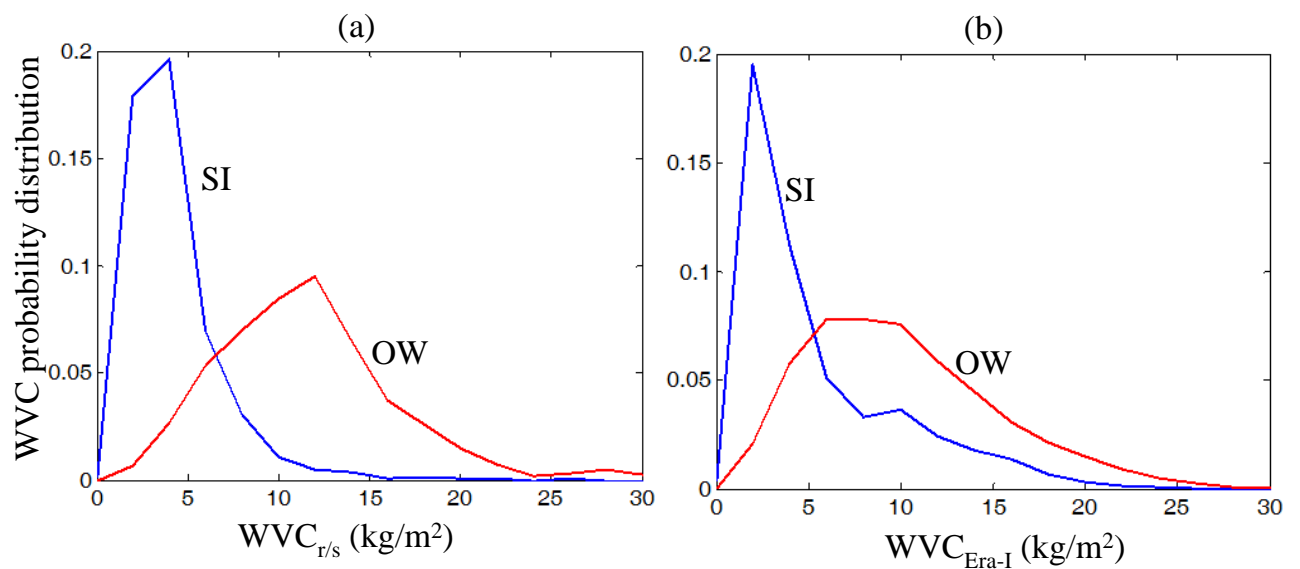
Fig. 4. WVC distribution in the Arctic: (a) from the 4 r/s station measurements during 2014-2017; (b) from the Era-Interim reanalysis data over the Arctic region for the period of 2015. Blue and red lines refer to WVC over SI and over OW surfaces respectively.
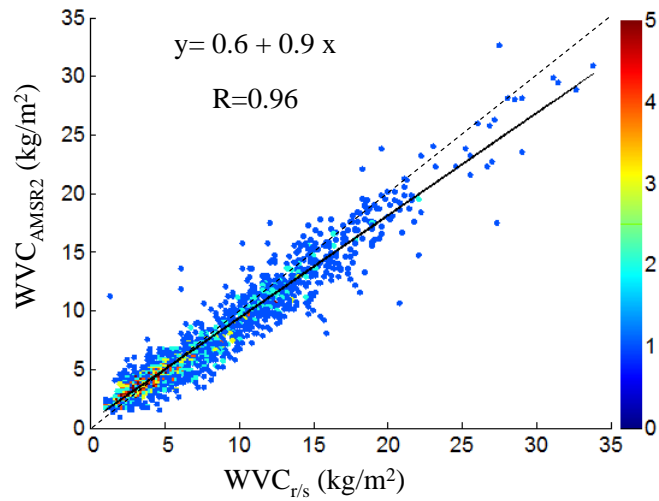
Fig. 5. The scatter plot of WVC, retrieved from the AMSR2 data, versus WVC, measured by radiosondes. Color shows the number of WVC data, falling into the bin of $\Delta_{WVC} = 0.175$ kg/m². The solid line represents the linear regression and the dashed line indicates 1:1 function.
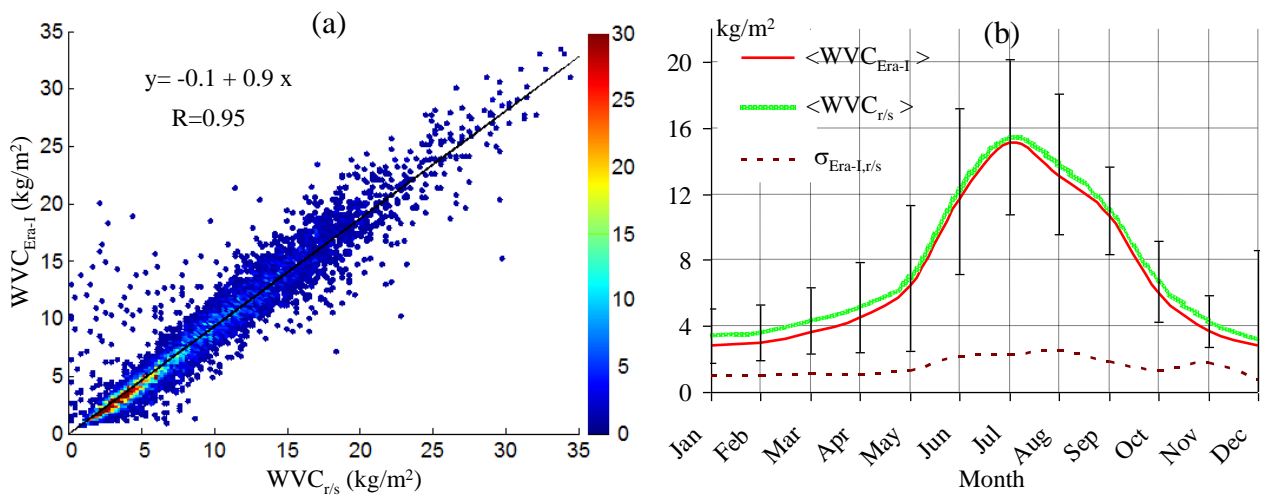
Fig. 6. (a) The scatter plot of WVC_{Era-I} versus $WVC_{r/s}$. Color shows the number of WVC data, falling into the bin of $\Delta_{WVC} = 0.175$ kg/m², the black line represents the linear regression; (b) Root mean square error $\sigma_{Era-I,r/s}$ and monthly means $<WVC_{Era-I}>$ and $<WVC_{r/s}>$. Error bars relate to $WVC_{r/s}$.
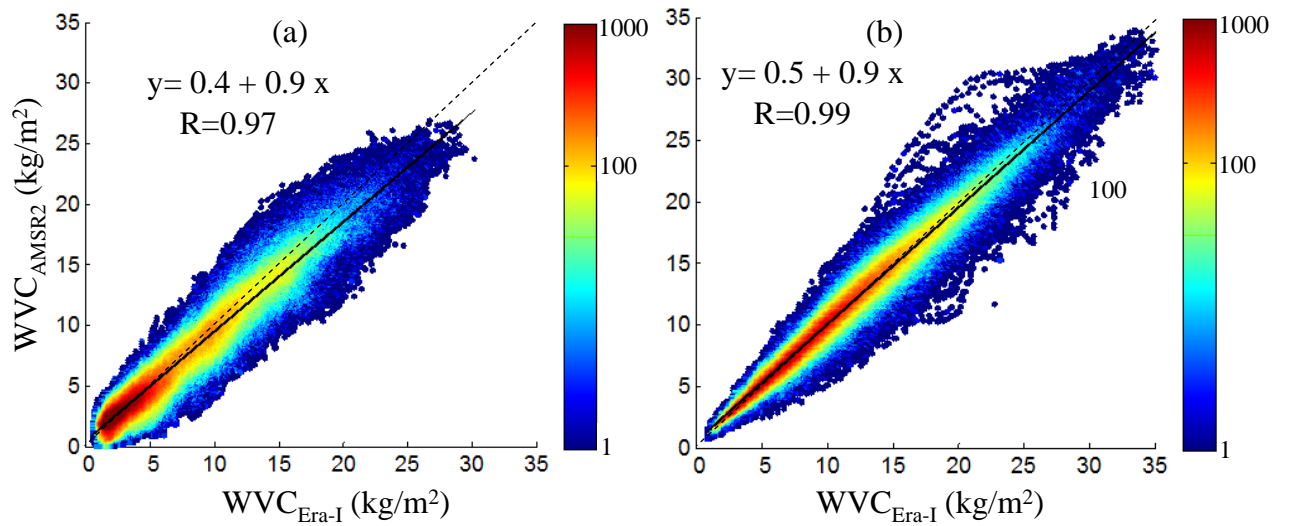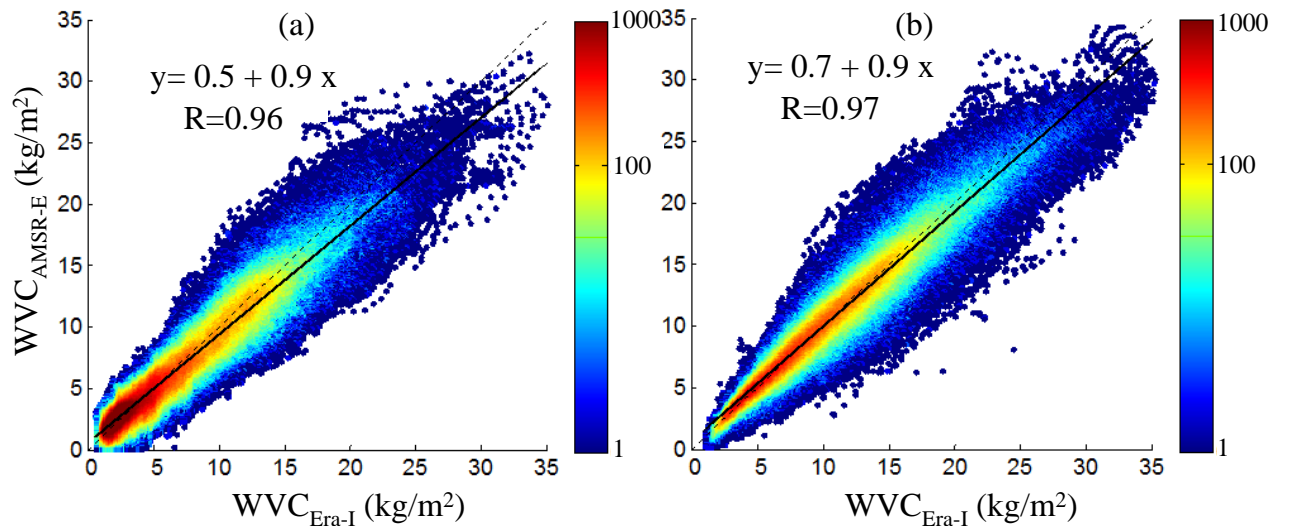
Fig. 7. (a) The scatter plot of WVC_{AMSR2} versus WVC_{Era-I} for the full domain of daily averaged data for the period of the year 2015 over SI; (b) the same over OW. Color shows the number of WVC data, falling into the bin of $\Delta_{WVC} = 0.117 \, kg/m^2$. The solid lines represent the linear regression and the dashed lines indicate 1:1 functions.
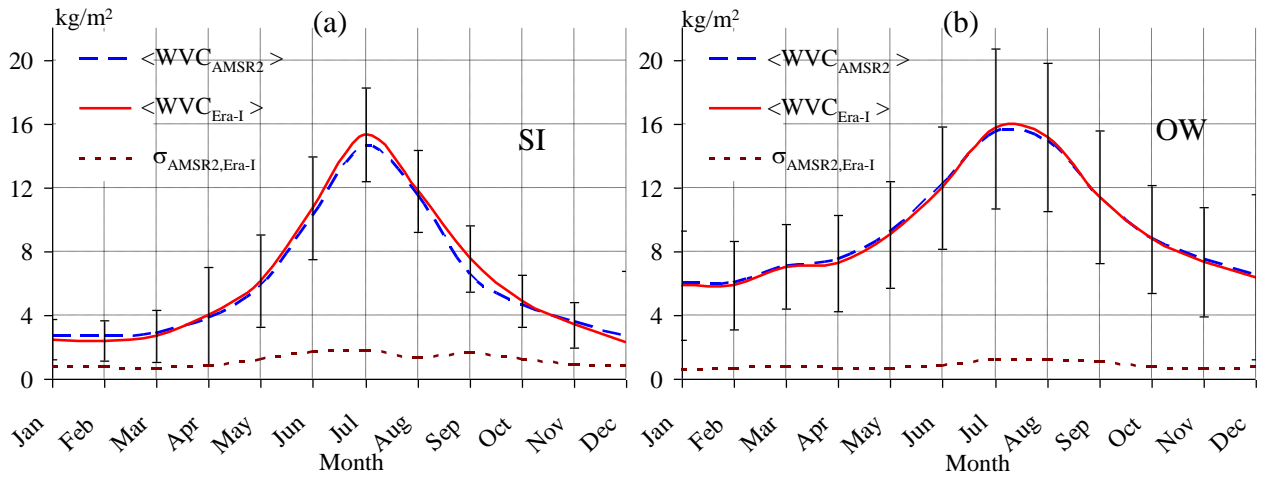

Fig. 8. (a) The scatter plot of WVC_{AMSR-E} versus WVC_{Era-I} for the full domain of daily averaged data for the period of the year 2008 over SI; (b) the same over OW. Color shows the number of WVC data, falling into the bin of $\Delta_{WVC} = 0.117 \, kg/m^2$. The solid lines represent the linear regression and the dashed lines indicate 1:1 functions.

Fig. 9. Monthly averaged WVC derived from the AMSR2 data ($<WVC_{AMSR2}>$) and Era-Interim reanalysis ($<WVC_{Era-I}>$), and the root mean square error $\sigma_{AMSR2,Era-I}$, for monthly datasets of the year 2015: (a) over SI; (b) over OW. Error bars relate to WVC_{Era-I}.
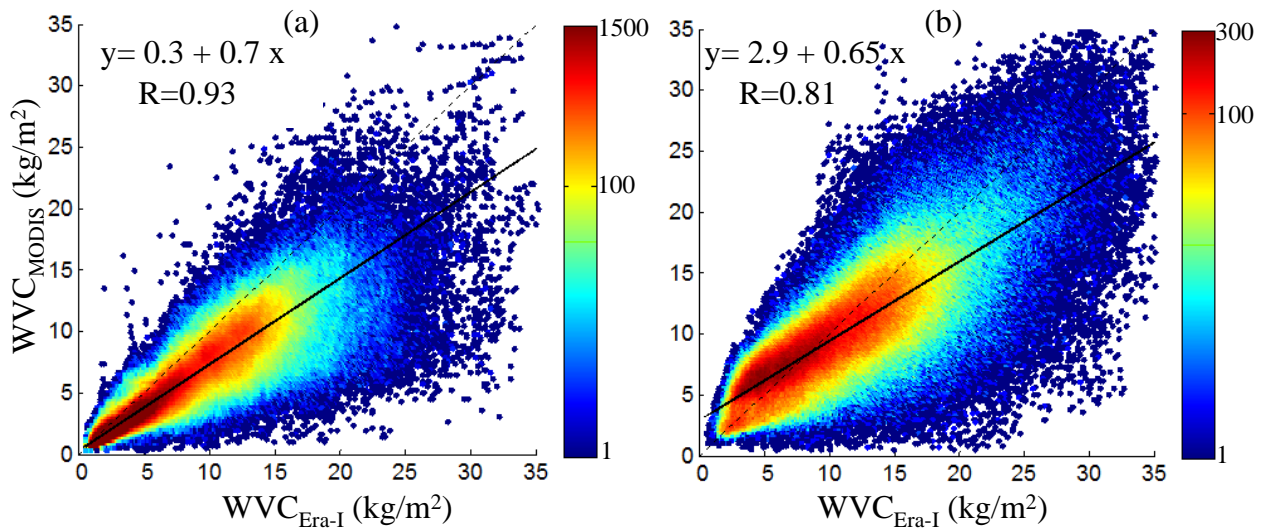

Fig. 10. (a) The scatter plot of WVC_{MODIS} versus WVC_{Era-I} for the full domain of daily averaged data for the period of 2008 over SI; (b) the same over OW. Color shows the number of WVC data, falling into the bin of $\Delta_{WVC} = 0.117$ kg/m^2. The black lines represent the data linear regression. The dashed lines indicate 1:1 functions.
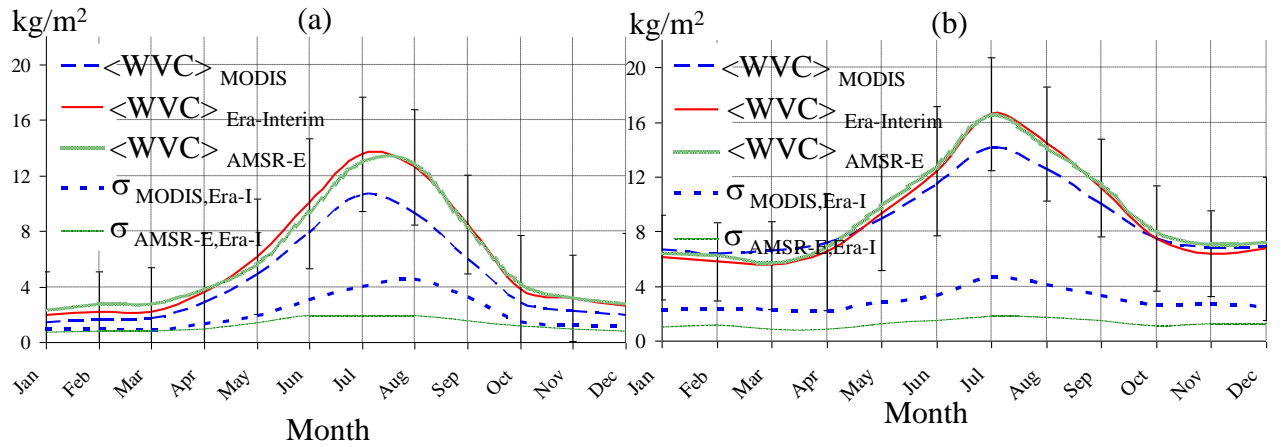
Fig. 11. Monthly averaged WVC derived from MODIS product ($<$WVC$_{MODIS}$$>$), AMSR-E data ($<WVC_{AMSR-E}$$>$) and Era-Interim reanalysis ($<$WVC$_{Era-I}$$>$), and the root mean square errors $\sigma_{MODIS,Era-I}$ and $\sigma_{AMSR-E,Era-I}$ calculated for monthly datasets of the year 2018: (a) over SI; (b) over OW. Error bars relate to WVC$_{Era-I}$.
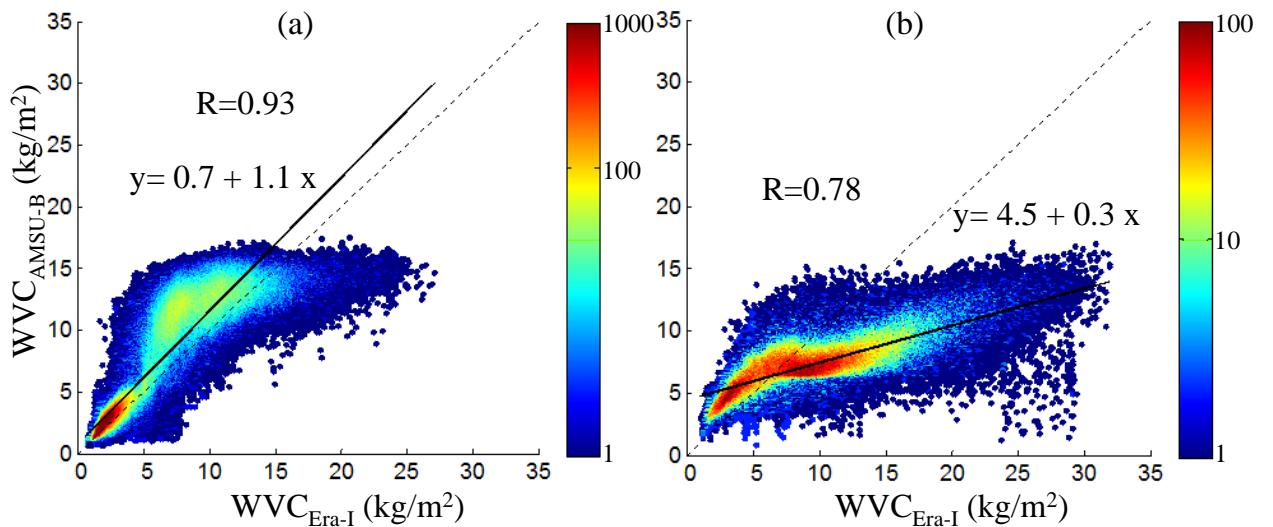

Fig. 12.   (a) The scatter plot of WVC$_{AMSU-B}$ versus WVC$_{Era-I}$ for the full domain of daily averaged data for the period of 2008 over SI; (b) the same over OW. Color shows the number of WVC data, falling into the bin of $\Delta_{WVC} = 0.117$ kg/m^2. The solid lines represent the data linear regression and the dashed lines indicate 1:1 functions.
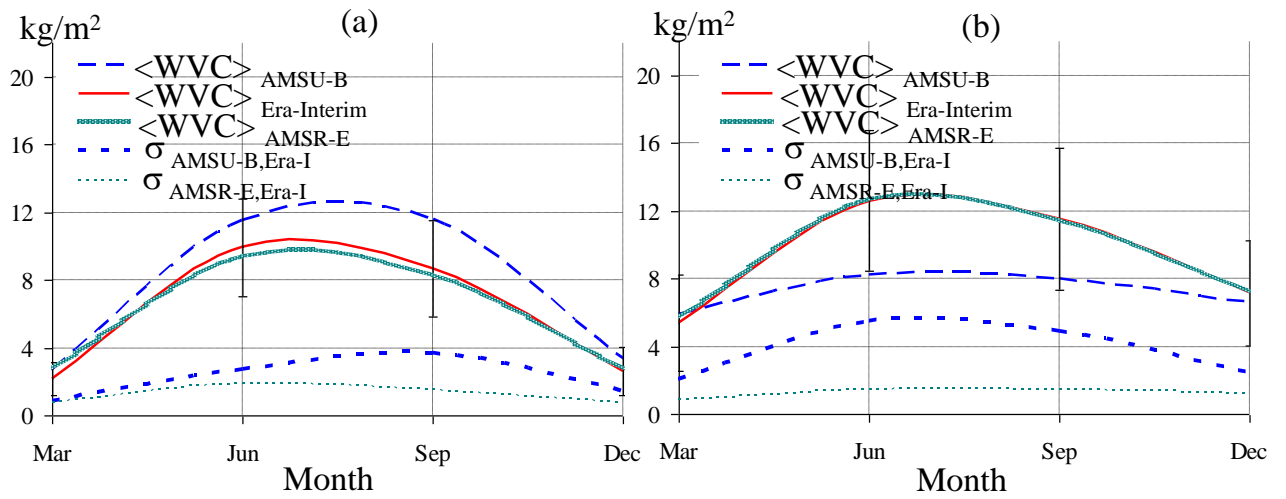
Fig. 13.   Monthly averaged WVC derived from the AMSU-B product (<WVC$_{AMSU-B}$>), the AMSR-E data (<WVC$_{AMSR-E}$>) and Era-Interim reanalysis (<WVC$_{Era-I}$>), and the root mean square errors $\sigma_{AMSU-B,Era-I}$ and $\sigma_{AMSR-E,Era-I}$ for March, June, September and December of 2008: (a) over SI; (b) over OW.

Fig. 14. WVC saptial distribution from (a) Era-Interim reanalysis(WVC$_{Era-I}$); (b) WVC$_{Era-I}$ - WVC$_{AMSR-E}$; (c) WVC$_{Era-I}$ - WVC$_{MODIS}$; (d) WVC$_{Era-I}$ - WVC$_{AMSU-B}$ on March 13, 2008.
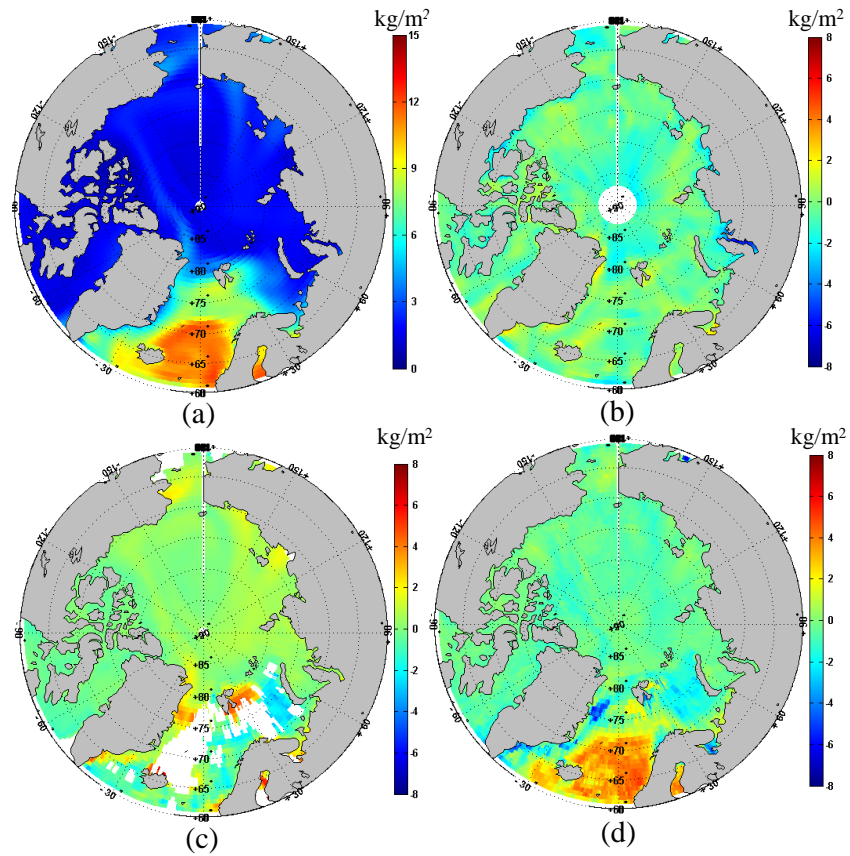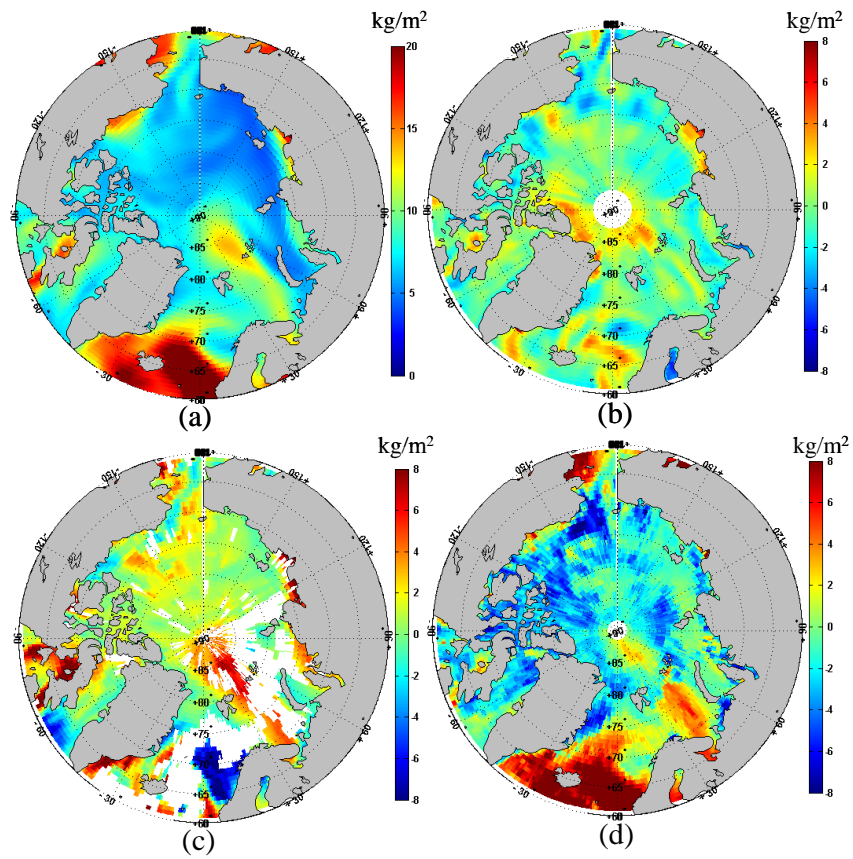
Fig. 15. WVC saptial distribution from (a) the Era-Interim reanalysis (WVC$_{Era-I}$); (b) WVC$_{Era-I}$ - WVC$_{AMSR-E}$; (c) WVC$_{Era-I}$ - WVC$_{MODIS}$; (d) WVC$_{Era-I}$ - WVC$_{AMSU-B}$ on June 3, 2008.

Tables.

Table I. Main characteristics of the AMSR-E and AMSR2 instruments.

AMSR-E and AMSR2 Channel Set				
Center frequency,GHz	Band width, MHz	Polarization	Ground resolution, km	Sampling interval, k m
6.925 7.3 – AMSR2 only	350	Vertical and Horizontal	35×61	10
10.65	100		24×41	
18.7	200		13×22	
23.8	400		15×26	
36.5	1000		7×12	
89.0(A&B)	3000		3×5	5

Table II. NN weights for the optimal NN model. T_{Bi} =T18H, T18V, T23H, T23V, T36H, T36V, T89H, T89V

n	w_{2in}								w_{1n}	b_{2n}
	i									
	1	2	3	4	5	6	7	8		
1	-0.015668	-0.025567	0.018162	0.000008	0.007004	0.023214	-0.004581	0.000150	-1.000000	1.000000
2	-0.000317	0.000471	0.013752	0.039928	-0.025111	-0.031677	0.010623	-0.004112	-1.000000	1.000000
3	0.000516	0.000774	0.000163	-0.001377	-0.003342	-0.003810	-0.002599	-0.002214	0.007264	0.628167
4	0.015318	0.011721	-0.037771	-0.023029	0.030968	0.005835	-0.012890	0.015067	-0.463866	0.652968
5	-0.009629	-0.035304	0.007139	0.050979	-0.025413	-0.007038	0.021316	0.000047	-0.432594	1.000000
6	0.000577	0.024602	0.005757	-0.026490	-0.006607	-0.017917	-0.004449	0.017089	0.139035	1.000000
7	0.000013	-0.000002	0.012980	0.003089	0.003489	-0.000075	-0.003979	-0.013136	-0.832870	1.000000

$b_1 = 14.053956$; $w_0 = 48.606963$; $b_0 = 45.027934$

Table III. The statistical characteristics of the WVC retrieval algorithm, applied to the AMSR2 and AMSR-E data over SI and OW, estimated with the Era-Interim WVC data for the whole year 2015 and 2008 respectively.

		σ, kg/m^2	σ_{rel}, %	R	bias, kg/m^2
AMSR2	SI	1.13	20	0.97	-0.10
	OW	0.90	9	0.99	-0.01
AMSR-E	SI	1.19	23	0.96	-0.07
	OW	1.10	11	0.97	-0.05
ZC$_{OW}$	OW	0.9	9	0.99	0.003

Table IV. The results of the comparison of satellite WVC products with the Era-Interim WVC data for the year 2008.

		σ, kg/m^2	σ_{rel}, %	R	bias, kg/m^2
MODIS	all data	2.4	38	0.91	-1.1
	SI	2.1	41	0.93	-1.2
	OW	3.2	32	0.81	-0.7
AMSU-B	all data	2.7	46	0.80	0.3
	SI	2.0	43	0.93	1.1
	OW	4.0	44	0.78	-2.0