

Urban sensing and Weather Prediction: Can IOT devices be used to improve weather prediction in cities?

N. Garg^a, **M. Prakash**^a, **M. Thatcher**^b and **R. Sankaran**^c

^aCSIRO Data61, Goods Shed, 34 Village Street, Docklands, VIC 3008

^bCSIRO Oceans and Atmosphere, Station Street, Aspendale, VIC 3195

^cArgonne National Lab, Chicago, USA

Email: nikhil.garg@data61.csiro.au

Abstract: Data assimilation (DA) in numerical weather prediction (NWP) has relied on observations sourced from variety of sources like radiosondes, dropwindsondes, meteorological sensors and remote sensing through satellites. The sensors used for collecting meteorological variables such as wind speed, wind direction, temperature, relative humidity are designed to provide high quality observations. Additionally, a set of defined instructions must be followed for installation of such systems, with the presumption that the sensors will behave as defined in the user manual. Due to the aforementioned reasons, firstly, these systems are expensive to install and operate, and secondly, they cannot be installed at a large scale within cities, thus requiring alternative ways of sensing environmental conditions within such areas.

In recent times, Internet of Things (IOT), Big Data and Cloud computing has been attracting considerable attention. This has been made possible by the availability of high speed connectivity, ease of access to high quality computing resources at pay-per-use basis and the increased emphasis on informed decision making. Due to such developments, there has been increased diffusion of smart systems equipped with miniaturized sensors allowing such systems to adapt to their environmental conditions *e.g.*, smart air-conditioners controlling the indoor conditions in response to the changing ambient environment. These disruptive technologies have led to the development of platforms such as Tulip¹, Array of Things² and Dryp³, that rely on alternative sensing methods and technologies.

Historically, weather and climate predictions has been carried out by the national weather agencies. These agencies use weather models together with the data from their observation networks to provide weather forecasts for public use on time scales ranging from daily, yearly to decadal. Despite the considerable advancements in modelling and computing systems, the weather forecasts from agencies can only attain spatial resolution of 1-10km. Due to the cost and computational systems needed to run such models, there has been a growing interest in combining big data and machine learning to aid in localizing the predictions performed by these models.

In the current study, we investigate the implications of combining observations from Array of Things (Catlett et al., 2017) network installed in City of Chicago with the model outputs from Conformal Cubic Atmosphere Model (CCAM). To that end, the study utilizes modelled and observed air temperature over a one month period. The preliminary results showed a good correlation between the modelled and observed air temperature. However, the sensors mounted on AOT node use different sensing techniques to measure air temperature. Due to these differences, a considerable spread exists in the air temperature observed by the different sensors mounted on a single AOT node. This observed spread in the air temperature underlines the need of caution when using data from IOT devices. Further evaluation against the data from a co-located meteorological sensor may elucidate the implications of this spread when data from IOT devices is combined with the models.

Keywords: *Numerical weather prediction, Internet of things, Array of things, Bias correction*

¹<https://www.tulipnetwork.org>

²<https://arrayofthings.github.io/>

³<https://www.dryp.global>

1 INTRODUCTION

In last two decades, the skill of operational forecasts from NWP models have improved considerably owing to the widespread use of ensembles of deterministic forecasts and DA. The improvements in forecast skill has also been due to improvements in model physics, increased grid resolution and adoption of coupled models *e.g.*, coupling atmosphere model with a land surface model and/or ocean model, thus allowing for better representation of small scale processes. However, the contribution of improvements in model physics have been lower compared to the improvements due to higher grid resolution and DA. For a detailed review on the progress in regional climate modelling, interested readers are referred to Wang *et al.* (2004). Moreover, the operational weather forecasts currently produced by majority of national weather services are still limited to a horizontal resolution of 1-10km, but there are on-going efforts at various institutions like Bureau of Meteorology (BOM) in Australia and UK MetOffice (UKMO) in United Kingdom to increase the spatial resolution to 200-500m over major cities. Although, spatial resolution of models that are currently operational is appropriate for producing forecasts at national and regional scale, they have limited skill in modelling localized microscale flow prevalent in urban areas. An illustration of various scales related to mesoscale-microscale is shown in Figure 1. The limitations of NWP models in resolving microscale flow have led to the development of integrated modelling systems such as Chen *et al.* (2011), where an urban canopy model is embedded within a NWP model. However, such models are computationally expensive to run and, have only been limited to research applications.

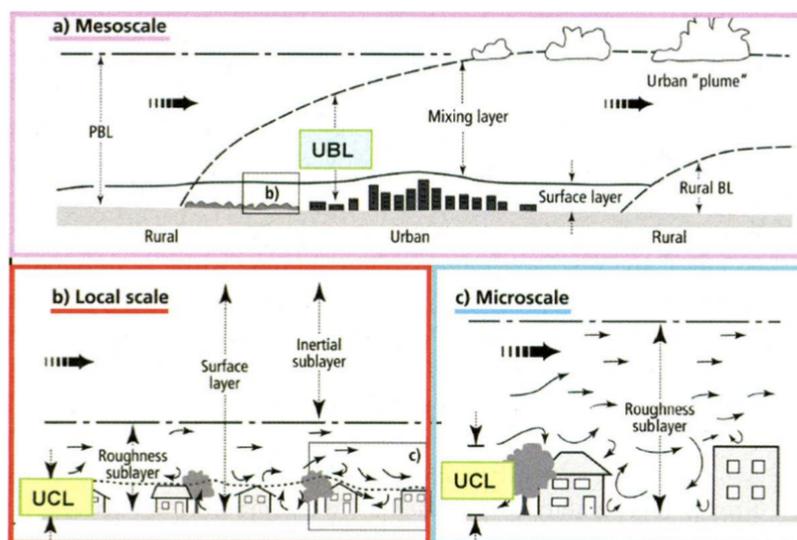


Figure 1. An illustration of urban boundary layer processes, from mesoscale to microscale. Source: Pullen *et al.* (2008)

Furthermore, the amount of observations and the diversity in sources of observations for use in NWP has grown considerably due to the improvements in connectivity. In addition to the conventional sources of observations *e.g.*, weather station networks, satellites, buoys and weather radars, data is also being collected from un-conventional sources *e.g.*, private weather stations (PWS; Wundergroud (2019)) and mobile devices (Netatmo SAS, 2019). Unlike these un-conventional sources, the conventional methods of observations have to adhere to the standards and methods laid out by the World Meteorological Organization (WMO, 2014), thus limiting where and how many of such systems can be installed. Additionally, over the last few years, sensor networks like Tulip and Array of Things (AOT; Catlett *et al.* (2017)) have been developed. In contrast to PWS and mobile phones, these sensor networks use low cost sensors for increasing the density of devices measuring environment conditions. The installation of such networks are either led by or carried out in collaboration with the city councils or administrations, where the local government agencies are interested in improving the liveability of the cities. For the sake of brevity, hereafter, data from the non-conventional sources of observations will be referred to as 'crowd-sourced' data. The unprecedented growth in 'crowd-sourced' data have also generated interest from the national weather agencies as evidenced by Hintz *et al.* (2019). A number of preliminary studies like Madaus and Mass (2017); Hintz *et al.* (2019) have used the barometric pressure data from the mobile phones and this was assimilated in NWP models. These studies have demonstrated the potential of 'crowd-sourced' data in increasing model skill, however, they have also pointed out the various

steps needed for screening and quality control. In light of the foregoing discussion, we aim to investigate the implications of combining meteorological observations collected from the AOT network installed within the City of Chicago with the forecast generated by the CCAM model. To this end, we will examine the potential of AOT network for correcting bias in modelled temperature field within an urban location.

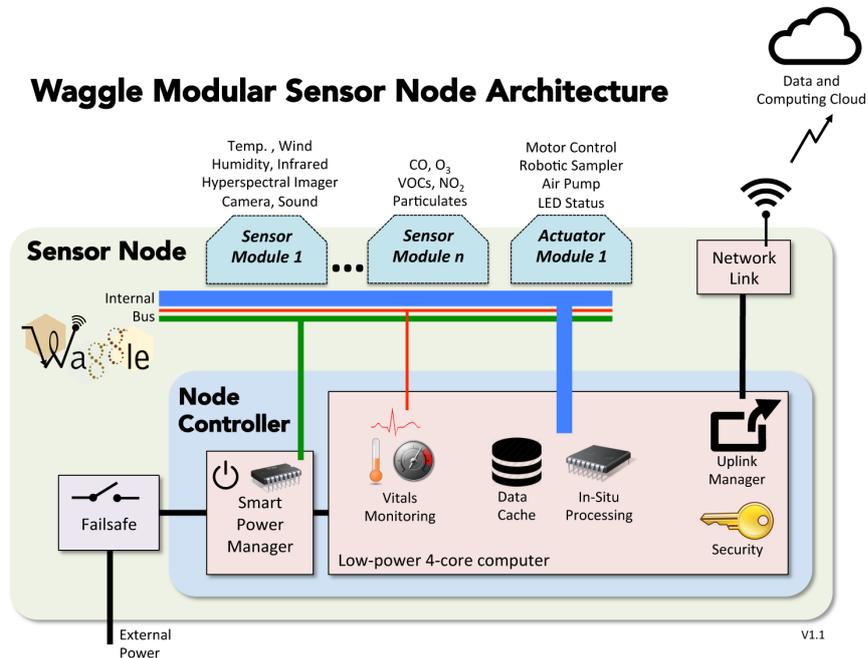


Figure 2. Architecture of Array of things sensor node. Source: <http://wa8.gl/architecture/architecture/>

2 DATA AND METHODS

2.1 Modelling System

The CCAM model (McGregor, 2005; McGregor and Dix, 2008) is developed primarily for the purpose of atmospheric modelling with the capability of modelling climate at global scale and for downscaling climate over a regional domain. The CCAM model employs conformal cubic grid and uses Schmidt transformation (Schmidt, 1977) for increasing horizontal resolution over a specified regional domain for downscaling. Grid stretching used in CCAM avoids the need for lateral boundary conditions needed by the majority of the other regional climate models *e.g.*, Advanced Research Weather and Forecasting model (WRF-ARW; Skamarock et al. (2008)) and Consortium for small scale modelling (COSMO; see online at <http://www.cosmo-model.org/>).

In the present study, long-wave and short-wave radiation are parametrized using Schwarzkopf and Ramaswamy (1999); Freidenreich and Ramaswamy (1999), and cloud microphysics was modelled using the scheme described in Rotstayn (1997) and Lin et al. (1983). The planetary boundary layer was modelled using a stability-dependent boundary layer scheme based on the Monin-Obukhov similarity theory (McGregor, 1993) with the non-local treatment following Holtslag and Boville (1993). The cumulus convection scheme was based on the mass flux closure described in McGregor (2005). It includes downdrafts, entrainment and detrainment. For the land surface scheme, a simple canopy scheme described in Kowalczyk et al. (1994) was used. This simple canopy scheme uses six layers for soil temperature and moisture respectively and three layers for snow. The soil moisture is obtained by solving Richard's equation. Lastly, the scale selective digital filter (Thatcher and McGregor, 2009) was used to force the temperature, sea level pressure and large scale winds from the 'parent' model (*i.e.*, reanalysis/ CCAM) onto each subsequent 'child' model.

The model setup uses three domains in one-way nesting mode, where the domains are centered at 87.68° West, 41.84° North. The first two model domains (*i.e.*, d01, d02) use C144 grid, whereas the third domain (d03) uses C192, resulting in 144×144 and 192×192 grid cells on each panel respectively. The coarsest domain d01 has a horizontal resolution of 60km, the nested domain d02 has a horizontal resolution of 10km, and the subsequently nested highest resolution domain d03 has a spatial resolution of 2km. It is worth highlighting

that the larger jumps in resolution (from d01-d03) are possible because of the stretched global grid used in CCAM. Furthermore, the aforementioned spatial resolution are only applicable within the output region (*i.e.*, the stretched part of the model grid), whereas, the model uses a lower resolution grid outside the output region. The model simulations were initialized with the ECMWF (European Center for Medium Weather Forecast) Intermediate reanalysis (ERA-Interim) (Simmons *et al.*, 2007). Herein, the ERA-Interim has a horizontal resolution of $0.75^\circ \times 0.75^\circ$ ($\approx 80\text{km}$). The start date for the model simulation was 1 January, 2019 at 00 UTC and the model simulation was carried out for 30 days.

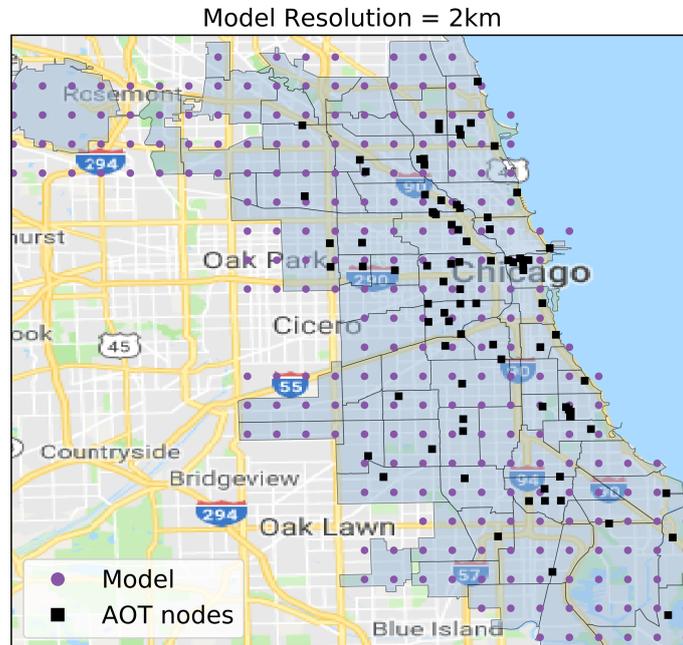


Figure 3. Illustration of CCAM model grid and the AOT sensor network within the Greater Chicago Area.

2.2 Observations

Over the last five years, the University of Chicago and the Argonne National Laboratory in collaboration with the City of Chicago has installed a network of more than 100 Array of things (AOT; Catlett *et al.* (2017)) nodes within the Greater Chicago Area. The main purpose of these nodes is to collect a variety of real-time data on urban environment, infrastructure.

The AOT nodes are equipped with sensors for measuring meteorological variables like air temperature, barometric pressure, humidity, air quality variables like concentration of carbon monoxide, nitrogen dioxide, sulphur dioxide and light intensity among others. For this study, we will restrict ourselves to the data collected from the sensors deployed for collecting observations on meteorological variables. An exhaustive list of data collected and the sensors used for collecting the data can be found at <https://arrayofthings.github.io/node.html> and <https://github.com/waggle-sensor/sensors>.

In addition to the sensors for collecting data, the nodes also contain two on-board computers, where one computer is used for node operations and the other one for edge computing. The former computer is used for performing in-situ operations like packing data and transferring it to a server referred to as "beehive", whereas the latter computer designated for edge computing comes with a graphical processing unit (GPU), thus enabling the deployment of tools for performing operations related to image processing and machine learning on the node itself before transferring processed data to the server. An illustration of the waggle modular sensor node architecture is given in Figure 2.

2.3 Crowd-sourced data for bias correction

The climate and weather forecast generated by NWP models suffer from bias. Therefore, a number of methods for correcting forecast bias have been developed. The bias in model forecasts is alleviated by using the observed data collected through conventional methods. However, due to lack of such observations within an urban area, we will use the data collected through the AOT network in the City of Chicago to bias correct the

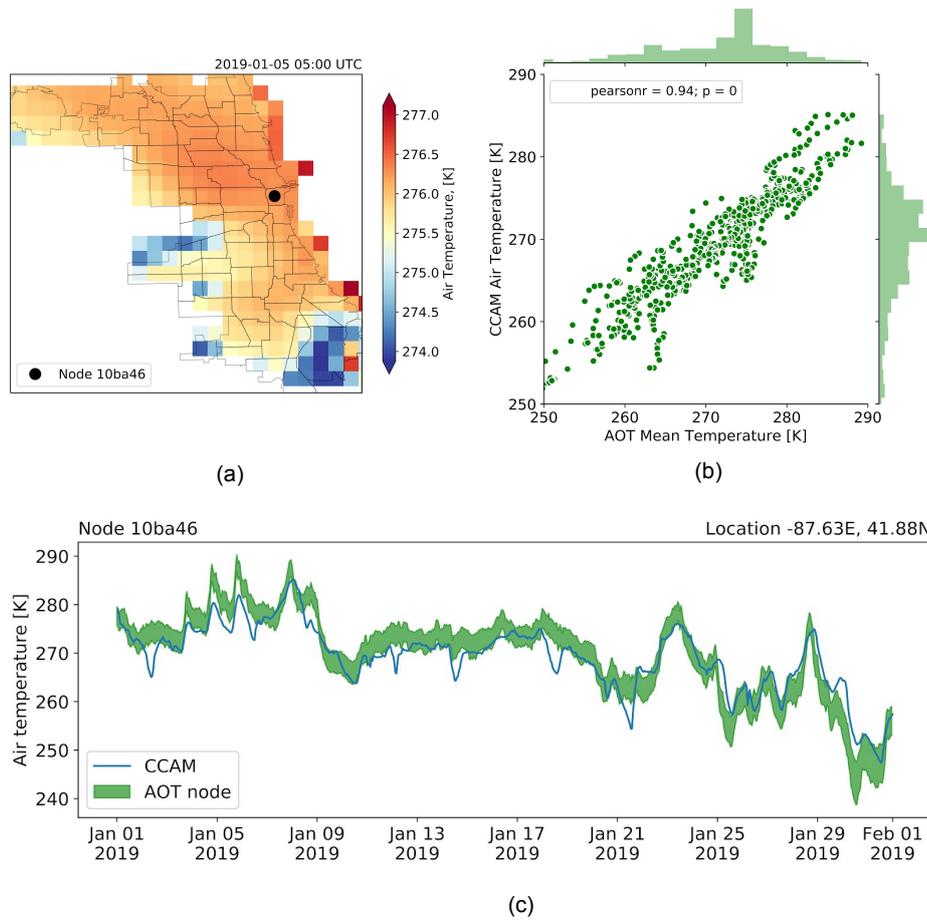


Figure 4. (a) The CCAM model temperature on 5 January, 2019 at 05UTC, (b) a scatter plot of mean air temperature from six temperature sensors mounted on the AOT node 10ba46 located at -87.63°E , 41.88°N and, (c) a comparison of hourly observed and modelled air temperature at the location of node 10ba46. The observed air temperature in (c) also shows the spread in air temperature observed by the 6 sensors used for observing temperature.

air temperature output from CCAM model. The various methods range from simple additive bias correction, quantile mapping (Piani et al., 2010) to kalman filter (Monache et al., 2008) based bias correction. Here, we have described the two widely used bias correction methods.

Additive bias correction. Following Hawkins et al. (2013), the bias in raw model output $x_m(t)$ can be corrected by simply shifting the mean of the model output \bar{x}_m by the mean of the observed data \bar{x}_o .

$$x_m^{corr}(t) = x_m(t) - (\bar{x}_o - \bar{x}_m) \quad (1)$$

In addition, it is also possible to not only shift the mean as per eq. (1), but also the temporal variability using eq. (2).

$$x_m^{corr}(t) = \bar{x}_o + \frac{\sigma_o}{\sigma_m}(x_m(t) - \bar{x}_m) \quad (2)$$

Here, the bias correction method uses the standard deviation of model output σ_m and the observations σ_o in addition to the mean values.

Quantile Mapping. In addition to the simple additive bias correction methods, removal of bias in model output can be done using quantile-quantile mapping (Piani et al., 2010). In quantile mapping, it is possible to account for bias in all statistical moments.

$$x_m^{corr} = F_o^{-1}(F_m(x_m)) \quad (3)$$

In this method, the quantile mapping is carried out as per eq. (3), where the cumulative distribution function (CDF) of the observations F_o and model output F_m are used. For CDF, it is possible to either use an empirical

cumulative distribution function, or the CDF of a fitted continuous distribution, with the assumption that both the model output and observations follow a continuous distribution *e.g.*, normal distribution.

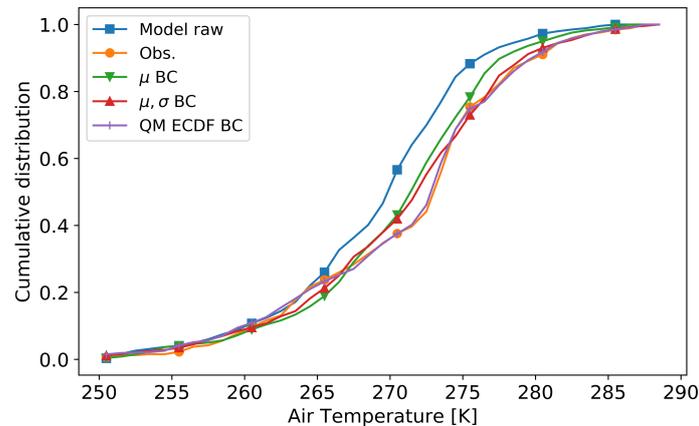


Figure 5. A comparison of empirical cumulative distribution function (CDF) obtained from observed and modelled air temperature. The CDF of modelled air temperature after bias correction (BC) using the two additive BC methods and quantile-quantile mapping BC is also shown.

3 RESULTS AND DISCUSSION

The model grid points from the domain d03 and the locations of AOT nodes within the Greater Chicago area are shown in Figure 2. It is evident that the network of AOT nodes has increased the density of observations collected. Figure 4 shows the spatial variation of air temperature from CCAM model within the Greater Chicago area on 5 January, 2019 at 05UTC. Additionally, a comparison of air temperature from 6 different sensors mounted on an AOT node located at -87.63°E , 41.88°N is also shown (see Figure 4(c)). It can be noted that the model output compares quite well with observed air temperature. However, it must be pointed out that the air temperature observed by different sensors deployed on AOT node vary considerably (see Figure 4(c)). A comparison of the empirical cumulative distribution of observed, raw modelled air temperature and bias corrected modelled air temperature is shown in Figure 5. From the preliminary results presented here, it is evident that observations collected through AOT nodes can be used for NWP modelling, however, the spread of temperature measured by the various sensors mounted on each node does raise questions on both the reliability and fidelity of such observation networks. It must also be kept in mind that the preliminary results presented here compare observations from a single node and for a single variable. To derive any robust conclusions, the procedure presented here will be expanded onto other nodes, both to evaluate the spatial pattern of observed temperature and bias in model outputs.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Array of Things team, the Argonne National Laboratory (ANL), the University of Chicago and the City of Chicago for providing Array of Things data collected from the City of Chicago in public domain. The first author would also like to acknowledge Rajesh Sankaran from Mathematics and Computer Sciences (MCS) division for hosting him at ANL.

REFERENCES

- Catlett, C. E., P. H. Beckman, R. Sankaran, and K. K. Galvin (2017). Array of things: a scientific research instrument in the public way: platform design and early lessons learned. In *Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering - SCOPE '17*, Pittsburgh, Pennsylvania, pp. 26–33. ACM Press.
- Chen, F., H. Kusaka, R. Bornstein, J. Ching, C. S. B. Grimmond, S. Grossman-Clarke, T. Loridan, K. W. Manning, A. Martilli, S. Miao, D. Sailor, F. P. Salamanca, H. Taha, M. Tewari, X. Wang, A. A. Wyszogrodzki, and C. Zhang (2011). The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *International Journal of Climatology* 31(2), 273–288.
- Freidenreich, S. M. and V. Ramaswamy (1999). A new multiple-band solar radiative parameterization for general circulation models. *Journal of Geophysical Research: Atmospheres* 104(D24), 31389–31409.

- Hawkins, E., T. M. Osborne, C. K. Ho, and A. J. Challinor (2013). Calibration and bias correction of climate projections for crop modelling: An idealised case study over Europe. *Agricultural and Forest Meteorology* 170, 19–31.
- Hintz, K. S., K. O’Boyle, S. L. Dance, S. Al-Ali, I. Ansper, D. Blaauboer, M. Clark, A. Cress, M. Dahoui, R. Darcy, J. Hyrkkänen, L. Isaksen, E. Kaas, U. S. Korsholm, M. Lavanant, G. Le Bloa, E. Mallet, C. McNicholas, J. Onvlee-Hooimeijer, B. Sass, V. Siirand, H. Vedel, J. A. Waller, and X. Yang (2019). Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4–5 December 2018. *Atmospheric Science Letters* 20(7), e921.
- Hintz, K. S., H. Vedel, and E. Kaas (2019). Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation. *Meteorological Applications*, 1–14.
- Holtzlag, A. A. M. and B. A. Boville (1993). Local Versus Nonlocal Boundary-Layer Diffusion in a Global Climate Model. *Journal of Climate* 6(10), 1825–1842.
- Kowalczyk, E. A., J. R. Garrat, and P. B. Krummel (1994). Implementation of a soil canopy scheme into the CSIRO GCM-regional aspects of the model response. Technical Report 32, Melbourne, Australia.
- Lin, Y.-L., R. D. Farley, and H. D. Orville (1983). Bulk Parameterization of the Snow Field in a Cloud Model. *Journal of Climate and Applied Meteorology* 22(6), 1065–1092.
- Madaus, L. E. and C. F. Mass (2017). Evaluating smartphone pressure observations for mesoscale analyses and forecasts. *Weather and Forecasting* 32(2), 511–531.
- McGregor, J. L. (1993). The CSIRO 9-level atmospheric general circulation model. Technical Report 26, Melbourne. OCLC: 246798611.
- McGregor, J. L. (2005). C-CAM: Geometrical Aspects and Dynamical Formulation.pdf. CSIRO Marine and Atmospheric Research Tech. Paper 70, CSIRO Marine and Atmosphere, Victoria, Australia.
- McGregor, J. L. and M. Dix (2008). A updated description of conformal cubic atmospheric model. In K. Hamilton and W. Ohfuchi (Eds.), *High Resolution Numerical Modelling of the Atmosphere and Ocean*, pp. 51–75. New York, NY: Springer New York.
- Monache, L. D., J. Wilczak, S. Mckeen, G. Grell, M. Pagowski, S. Peckham, R. Stull, J. Mchenry, and J. Mcqueen (2008). A kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone. *Tellus B: Chemical and Physical Meteorology* 60(2), 238–249.
- Netatmo SAS (2019). Netatmo weathermap. <https://weathermap.netatmo.com/>. Accessed on 29 July, 2019.
- Piani, C., J. O. Haerter, and E. Coppola (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology* 99(1), 187–192.
- Pullen, J., J. Ching, D. Sailor, W. Thompson, B. Bornstein, and D. Koracin (2008). Progress toward meeting the challenges of our coastal urban future. *Bulletin of the American Meteorological Society* 89(11), 1727–1734.
- Rotstayn, L. D. (1997). A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of the microphysical processes. *Quarterly Journal of the Royal Meteorological Society* 123(541), 1227–1282.
- Schmidt, F. (1977). Variable fine mesh in the spectral global models. *Beitraege zur Physik der Atmosphaere* 50, 211–217.
- Schwarzkopf, M. D. and V. Ramaswamy (1999). Radiative effects of CH₄, N₂O, halocarbons and the foreign-broadened H₂O continuum: A GCM experiment. *Journal of Geophysical Research: Atmospheres* 104(D8), 9467–9488.
- Simmons, A., S. Uppala, D. Dee, and S. Kobayashi (2007). ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF newsletter* 110(110), 25–35.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. Gill, D. M. Barker, M. Duda, X. Y. Huang, W. Wang, and J. G. Powers (2008). A Description of the Advanced Research WRF Version 3. Technical report, National Center for Atmospheric Research (NCAR). NCAR/TN-475+STR.
- Thatcher, M. and J. L. McGregor (2009). Using a Scale-Selective Filter for Dynamical Downscaling with the Conformal Cubic Atmospheric Model. *Monthly Weather Review* 137(6), 1742–1752.
- Wang, Y., L. R. Leung, J. L. McGregor, D. K. Lee, W. C. Wang, Y. Ding, and F. Kimura (2004). Regional Climate Modeling: Progress, Challenges, and Prospects. *Journal of the Meteorological Society of Japan. Ser. II* 82(6), 1599–1628.
- WMO (2014). Guide to Meteorological Instruments and Methods of Observation. Technical report, World Meteorological Organization, Geneva.
- Wunderground (2019). Wunderground PWS. <https://www.wunderground.com/weatherstation/overview.asp>. Accessed on 29 July, 2019.