Proceedings of ERAD (2002): 212–216 © Copernicus GmbH 2002

ERAD 2002

On the missing data problem in rass wind profiler measurements: an algorithm based on functional differential equations

A. Lay-Ekuakille, A. Cataldo, and A. O Trotta

Dipartimento d'Ingegneria dell'Innovazione, Università di Lecce Via Arnesano, 73100 Lecce, Italy

Abstract. RASS (Radio Acoustic Sounding System).is usually an option added to Wind Profiler to provide profiles of virtual temperature data. Rass-Wind Profiler belongs to meteorological radars. Virtual temperature is uncompensated for humidity or pressure. Recovered data must be processed in a given period and are accepted if only they belong to a consensus window, whose amplitude is in general 3 m/s and only for those data it is suitable to make an average. Many difficulties rise during Rass operating mode: first the unchanging of data with the increasing of height in some conditions and the data loss because of interruption. To recover data, we use Genetic Algorithms as prediction and retrieval technique.

1 Introduction

Doppler radar observes perturbations in the refractive index of Air created by high intensity acoustic pulses. The measurement of the velocity of acoustic propagation lets us calculate the temperature as a function of altitude. The acoustic wave is a local compression of the air followed by depression that is propagated through the atmosphere. The acoustic and electromagnetic wavefronts are approximately spherical. [a] When the acoustic source and the radar are situated at the same point, the electromagnetic energy returned by the gradient of the refractive index created by the acoustic wave is focused on the radar. Such a Radio Acoustic Sounding System (Rass) avoids the attenuation in r^{--2} (where r is the range) undergone by the backscattering from the natural fluctuations in the refractive index and also focuses the scattered energy and the spatial coherence of the peaks of acoustic waves. The scattered fields will arrive in phase at radar if the acoustic wave has a wavelenght λa equal to the half-wavelengths of the radar $\lambda_e/2$ [b].

The Radio Acoustic Sounding System (RASS) is an option that can be added to the Wind profiler [1] to provide to profiles of virtual temperature data. The RASS system, composed of four acoustic sources, one on each side of the profiler, transmits a vertically directed acoustic wave. RASS (Radio Acoustic Sounding System) uses acoustic wave emission to measure atmosphere virtual temperature profiles. Virtual temperature is that one must have dry air to make equal humid air density at the same pressure. This variable is widely used because it is possible to study variations of virtual temperature instead of density ones. Thus we define: $T_v = T(1 + 0.61Q)$ where T is absolute temperature and Q is specific humidity given by the ratio between water vapour mass and humid air mass containing water vapour $[Q = M_w/(M_W + M_d)]$.

Thus, a Wind Profiling Radar is a meteorological and atmospheric remote sensing instrument, gives information about a volume of the atmosphere at a distance without being physically located in the region.

The profiler uses the acoustic wave as a target, receiving and processing the resulting backscatter and measuring the speed of propagation. The profiler can compute virtual temperature profiles because the speed of sound is easily related to air temperature [2]. Raw temperature data are stored in the moment and spectral data files, but separated from wind data into consensus files. Table 1 shows an example of data file to obtain temperature profiles. The titles of each column are: HT height in km, T uncorrected temperature, T_c corrected temperature, W vertical component of wind, CNT consensus counting, and SNR signal-to-noise ratio. We see that there is no changing of T and T_c corrected with the increasing of height and the number of "good" data CNT decreases with height as well as SNR. The difficulty of recovering data after a given height associated with the choice of "good" data, represents a limitation and is a source of data errors.

1.1 Rass-Wind Profiler Measuring System

Data errors can affect the four measurements for moment as illustrated in Fig. 1; four quantities are calculated for each set of spectrally averaged data during the frequency-domain stage: the Doppler shift of the peak; the spectral width; the

Correspondence to: A. Lay-Ekuakille
(aime.lay.ekuakille@unile.it)

Table 1. Example of measured data from RASS operating mode

HT	T	T_c	W	CNT			SNR		
0.120	18.1	18.6	-0.2	29	30	28	6	5	-9
0.180	18.0	18.2	0.1	28	30	27	3	3	-10
0.480	16.8	15.6	0.5	28	31	30	-18	-18	-10
0.540	16.3	15.0	0.5	27	29	26	-22	-22	-12
0.600	16.2	15.1	0.6	22	27	22	-25	-26	-12
0.660	9999.0	9999.0	0.9	20	15	19	-28	-28	5
0.720	9999.0	9999.0	0.4	17	12	20	-29	-30	-10
0.780	9999.0	9999.0	1.0	15	9	21	-31	-31	16

noise power and the signal-to-noise ratio. The moments and spectra are considered "raw" data. The profiler creates a third type of data file, the consensus file. The data in this file are processed with a wind consensus averaging algorithm. The current algorithm uses two values to determine whether data are valid. One value is a range in which the samples must fit. The second value is a percentage number of samples taken during the consensus period that must fit within that range before the consensus is accepted as valid. Both values are chosen by the operator. The operator choice affects data processing, causing errors and less accuracy.

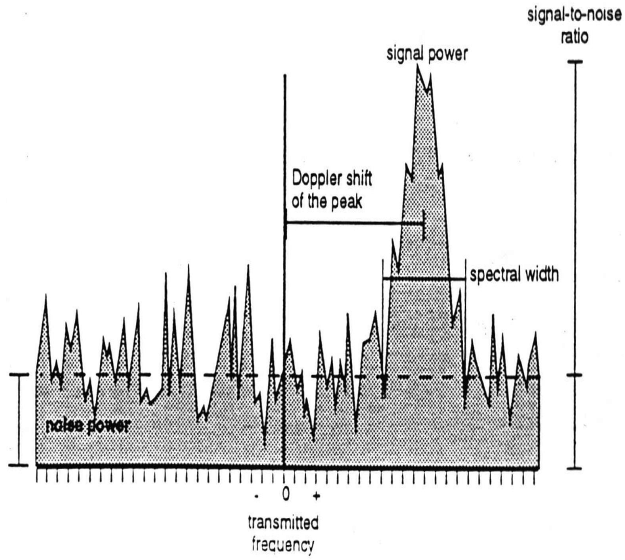

Fig. 1. Diagram of a spectrum showing the four measurements for moment data.

1.2 Theoretical aspects of Genetic Algorithms

There is a large class of interesting problems for which no reasonably fast algorithms have been developed. Many of these problems are optimization problems that frequently arise in applications. Given such a hard optimization problem, it is often possible to find an efficient algorithm whose solution is approximately optimal. For some hard optimization problems we can use probabilistic algorithms as well these algorithms do not guarantee the optimum value [3], but by randomly choosing sufficiently many "witnesses", the probability of error may be made as small as we like. There are a lot of important practical optimization problems for which such algorithms of high quality have become available [4]. For instance we can apply simulated annealing for wire routing and component placement problems in VLSI design or for the traveling salesman problem. Moreover, many other large-scale combinatorial optimization problems can be solved approximately on present-day computers by this kind of Monte Carlo Technique. In general, any abstract task to be accomplished can be thought of as solving a problem, which, in turn, can be perceived as a search through a space of potential solutions. Since we are after "the best" solution, we can view this task as an optimization process. For small spaces, classical exhaustive methods usually suffice; for larger spaces, special intelligence techniques must be employed. Genetic Algorithms (GA's) are among such techniques; they are stochastic algorithms whose search methods model some natural phenomena: genetic inheritance and Darwinian strife for survival. As stated [5]:

"... the metaphor underlying genetic algorithms is that of natural evolution. In evolution, the problem each species faces is one of searching for beneficial adaptations for a complicated and changing environment. The 'knowledge' that each species has gained is embodied in the makeup of the chromosomes of its members." The idea behind genetic algorithms is to do what nature does. Genetic algorithms use a vocabulary borrowed from natural genetics. We would talk about individuals (or *genotypes, structures*) in a population; quite often these individuals are called also *strings* or *chromosomes*. This might be a little bit misleading: each cell of every organism of a given species carries a certain number of chromosomes (man, for example, has 46 of them); however, in this paper we talk about one-chromosome individuals only not *diploidy* (pairs of chromosomes)[6]. Chromosomes are made of units, *genes* (also *features, characters, or decoders*), arranged in linear succession; every gene controls the inheritance of one or several characters. Genes of certain characters are located at certain places of the chromosomes, which are called *loci* (string positions). Any character of individuals (such as hair color) can manifest itself differently; the gene is said to be in several states, called *alleles* (feature values). Each genotype (in this paper a single chromosome) would represent a potential solution to a problem; an evolution process run on a population of chromosomes corresponds to search through a space of potential solutions. Such a search requires balancing two (apparently conflicting) objectives: exploiting the best solutions and exploring the search space [7].

GA's have been quite successfully applied to optimization problems like wire routing, scheduling, adaptive control, game playing, cognitive modeling, transportation problems, traveling salesman problems, for our case, in Forecasting [8]. The structure of a simple genetic algorithm is the same as the structure of any evolution program. During iteration t, a genetic algorithm maintains a population of potential solutions (chromosomes, vectors)

$$P(t) = \{x^t 1, \ldots, x^t n\} \tag{1}$$

Each solution x_i^t in Eq. (1) is evaluated to give some measure of its "fitness". Then, a new population (iteration $t+1$) is formed by selecting the more fit individuals. Some members of this new population undergo reproduction by means of crossover and mutation, to form new solutions. Crossover combines the features of two parent chromosomes to form two similar offspring by swapping corresponding segments of the parents. For example, if the parents are represented by five-dimensional vectors $(a_1, b_1, c_1, d_1, e_1)$ and $(a_2, b_2, c_2, d_2, e_2)$, then crossing the chromosomes after the second gene would produce the offspring $(a_1, b_1, c_2, d_2, e_2)$ and $(a_2, b_2, c_1, d_1, e_1)$. The intuition behind the applicability of the crossover operator is information exchange between different potential solutions. Mutation arbitrarily alters one or more genes of selected chromosome by a random change with a probability equal to the mutation rate. The intuition behind the mutation operator is the introduction of some extra variability into the population. A genetic algorithm (as any evolution program) for a particular problem must have the following five components:

- a genetic representation for potential solutions to the problem,
- a way to create an initial population of potential solutions,
- an evaluation function that plays the role of the environment, rating solutions in terms of their "fitness",
- genetic operators that alter the composition of children during reproduction,
- values for various parameters that the genetic algorithm uses (population size, probabilities of applying genetic operators, etc.).

2 Genetic algorithm approach

We propose in this paper a new algorithm that helps operator's choice in one hand and predicts the values of temperature with the increasing of height, in the other hand. Data analysis can be considered as a process in which starting from some given data sets, information about the respective application is generated. In this sense data analysis can be defined as a search for structure in data. Since in our problem there is the need of finding a link between the number of temperature data and the height, making use of some known data, that may be employed to predict the real range of data; the use of genetic algorithm program [9] techniques for symbolic regression and the use of genetic algorithms [10] and evolution strategies to solve the above problems [11]. The model we have used here [12] can be formulated in the following way: a set of observations of a given phenomenon (e.g. a physical system or a formal dynamic system) has a structure of a coupled set, $[(x^1, y^2), \ldots\ldots (x^N, y^N)]$ where $x^i = (x_1^i, \ldots, x_n^i)$ are independent variables and y^i (with $1 \le i \le N$) are dependent variables.

For example, in the forecasting field, dependent variables may be for instance humidity, barometric pressure, temperature, and the dependent variable may be for instance precipitation.

We use genetic algorithm to explore the space of condition sets on independent variables so that we can get good prediction on dependent variable.

To do so, we must find [13] the "predictability region" in times – series generated by Mackey-Glass equation [14].

A Packard's algorithm illustrated in [12] has been modified by us with the use of Non Retarded Functional Differential Equations (NRFDE's) and less iterations.

To solve the problem of RASS-Wind profiler measurement, we have to search the "predictability regions" of the time-series generated by the Mackey-Glass's equation (see [14]), by considering the RASS-Wind profiler data acquisition as a dynamic and chaotic system as a blood flow model, so:

$$\frac{dx}{dt} = \frac{ax(t-\tau)}{1+[x(t-\tau)]^c} - bx(t) \tag{2}$$

where $x(t)$ is a state variable, like temperature, to be recovered, t is the time and a, b, c and τ are constants to be retrieved. To form the data base, for each datum i, the independent variable $\boldsymbol{x}^i$ are 5 consecutive values of $x(t)$ (instead of 50 as suggested by Meyer Packard), one per second:

$$\boldsymbol{x}^i = (x_1^i, x_2^i, \ldots, x_5^i) \tag{3}$$

The dependent variable for the i-th datum, y^i, is the state variable after t' instants: that is $y^i = x_{5+t'}^i$. Each couple $(\boldsymbol{x}^i, y^i)$ is produced by iterating Eq. (2) with different initial condition, where an initial condition is a set of values for $(x_{1-\tau,\ldots}, x_0)$.

It is possible to use the following fitness function in order to measure the amount of information that is present in y distribution (y is the final temperature computed) for those points that satisfy C:

$$f(C) = -\log_2\left(\frac{\sigma}{\sigma_0}\right) - \frac{\alpha}{N_C} \tag{4}$$

where σ is the standard deviation of y set for data that satisfy C, σ_0 is the standard deviation of y distribution on the whole data set, N_C is the number of data that satisfy C condition and α is a constant.

For the goal of our research, we have transformed the RFDE of (2) in NRFDE [15] by getting optimal results as shown in the next section.

3 Solution optimization technique

We used data recovered by the RASS-Wind Profiler of University of L'Aquila , that concern 5 May 1997 between hour

Table 2. Fitness class results of simulations using NRFDE

a[m]	b[m]	C	t	T_c [°C]	T_m [°C]
0	120	0.260	150	24.6447	–
120	180	0.329	150	18.1653	18.1
180	240	0.327	150	18.0929	18.0
240	300	0.32	150	17.7146	17.5
300	360	0.315	150	17.4308	17.3
360	420	0.313	150	17.3052	17.3
420	480	0.305	150	16.8586	16.8
480	540	0.307	150	16.8960	16.8
540	600	0.295	150	16.3435	16.3
600	660	0.295	150	16.2725	16.2
660	720	0.285	150	15.7255	9999.0
720	780	0.275	150	15.1750	9999.0
780	840	0.265	150	14.6249	9999.0
840	900	0.255	150	14.0752	9999.0
900	960	0.245	150	13.5258	9999.0
960	1020	0.235	150	12.9768	9999.0
1020	1080	0.225	150	12.4281	9999.0
1080	1140	0.215	150	11.8796	9999.0
1140	1200	0.205	150	11.3313	9999.0
1200	1260	0.2025	150	11.1897	9999.0
1260	1320	0.20125	150	11.1162	9999.0
1320	1380	0.20105	150	11.1	9999.0
1380	1440	0.20045	150	11.0626	9999.0
1440	1500	0.20025	150	11.0472	9999.0

Table 3. Fitness class results of simulations using NRFDE

a[m]	b[m]	C	t	T_c [°C]	T_m [°C]
0	120	0.300	150	27.4884	–
120	180	0.320	150	17.6937	18.1
180	240	0.300	150	16.6656	18.0
240	300	0.310	150	17.1829	17.5
300	360	0.316	150	17.4842	17.3
360	420	0.312	150	17.2517	17.3
420	480	0.320	150	17.6634	16.8
480	540	0.320	150	17.6485	16.8
540	600	0.290	150	16.0206	16.3
600	660	0.290	150	16.0071	16.2
660	720	0.289	150	15.9413	9999.0
720	780	0.270	150	14.9051	9999.0
780	840	0.269	150	14.8410	9999.0
840	900	0.250	150	13.7962	9999.0
900	960	0.249	150	13.7421	9999.0
960	1020	0.230	150	12.8500	9999.0
1020	1080	0.229	150	12.5243	9999.0
1080	1140	0.219	150	11.4378	9999.0
1140	1200	0.205	150	11.2859	9999.0
1200	1260	0.2045	150	11.3621	9999.0
1260	1320	0.2019	150	11.2637	9999.0
1320	1380	0.2011	150	11.0275	9999.0
1380	1440	0.20089	150	10.8577	9999.0
1440	1500	0.2003	150	10.7592	9999.0

19:00 and hour 21:00. The interval between each acquisition is 30 minutes.

We have used the following parameters according the modified Eq. (2) to train the genetic algorithm:

a is the value in meter of the lower stratum, b is the upper one, c is a constant, and τ is 150. Table 2 summarizes the results.

The algorithm works as follows:
a) initializes population with a random set C
b) computes suitability of each C
c) classifies each population according to suitability d) discards a certain quantity of objects with less suitability,
and substitutes discarded objects by applying a mutation and crossing to remained C
f) returns to step b).

Now it is possible to see the graph that shows the RASS-Wind Profiler recovered temperature profile. The results we have got are basically interestint since they agree to the common aspects of RASS-Wind Profiler.

Remote Sensing. Meteorological Radars, such as RASS-Wind Profiler, have been replacing the widespread radiosondes in monitoring atmosphere for weather forecasting; since radiosondes do not give a correct vertical profile because they undergo wind interference, hence they constantly change direction and they present high cost of launching, RASS-Wind Profiler has only an initial high cost and can be supervised by one operator.

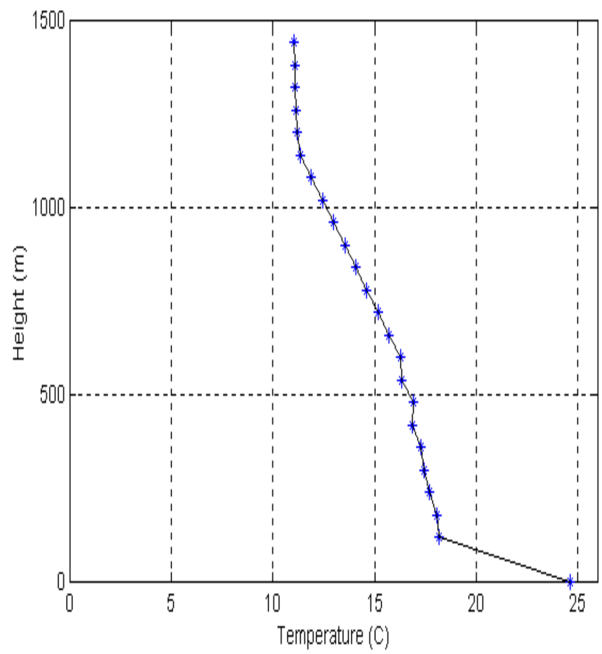

Fig. 2. Recovered temperature profile with fitness class 1.

RASS-Wind Profiler offers many advantages like vertical stability in sounding atmosphere searching targets. In Table 2, we have retrieved temperature data (T_c) with respect to the

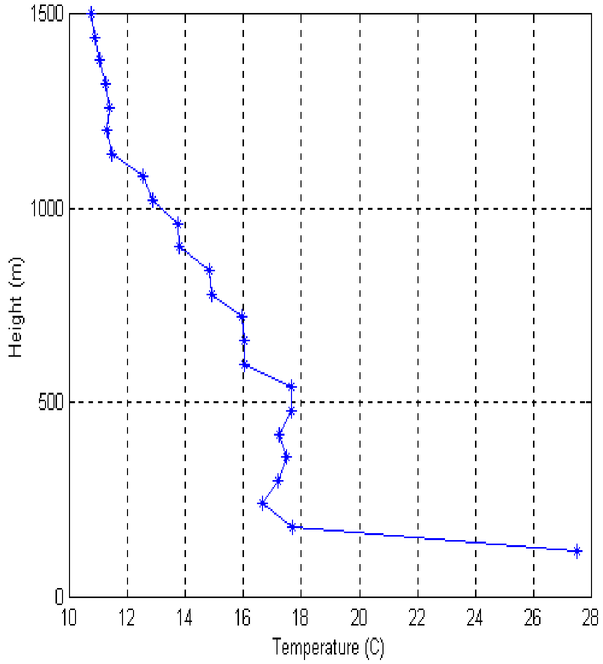

Fig. 3. Recovered temperature profile with fitness class 2.

measured temperature (T_m). We remark that from 720 m in vertical position up to 1440 m, it is not possible to get T_m. So, with this "dodge", it is possible to use these data without discarding them.

Table 4 shows the feasibility of using Chauvenet's criterion in terms of improving data processing by discarding those data that are presumably suspected. In this case only one temperature value can be considered. All data above considered, whose final graph is illustrated in Fig. 2, are coherent with the temperature trends observed in May 1997. All results can be evaluated as appreciable.

4 Summary and conclusions

In this paper we have demonstrated the applicability of GAs in weather forecasting, especially when recovering temperature profile.

We have also used Chauvenet's criterion in order to discard potential suspected data that would not be in accordance with "routine" trends in a given period. This application offers new opportunity in solving temporary limitations of RASS-Wind Profiler operating mode.

Acknowledgement. We thank Dr. Giovanni Cinque of University of L'Aquila for allowing us to get in touch not only with the RASS-Wind Profiler but to use data from it.

References

[a] H. Sauvageot, Radar meteorology, Artech House,Inc, Norwood, USA, 1992.

[b] A. D'Orazio, M. De Sario, A. Lay Ekuakille et al, Integrated Sensor Rass-Radiometer in Remote Sensing: Using a Supervising Unit in Joined Temperature Measurements, Proceedings of 15th International Conference on Applied Eletromagnetics and Communications, Dubrovnik, Croatia, October 11-13,1999.

[1] Radian Corporation Electronics Division, The LAP-3000 Wind Profiler Radar, Radian Corporation, Boulder, PH, 1996.

[2] M.A. Janssen, Atmospheric Remote Sensing by Microwave Radiometry, pp.177-178, John Wiley & Sons, New York, 1993.

[3] Z. Michalewicz, Genetic Algorithms+Data Structures=Evolution programs, Springer-Verlag, New York, 1992.

[4] L. Davis, Genetic Algorithms and Simulated Annealing, Morgan kaufmann Publishers, Inc., Los Altos,CA,1987.

[5] L. Davis & M. Steenstrup, Genetic Algorithms and Simulated annealing: An Overview, pp.1-11,Morgan kaufmann Publishers, Inc., Los Altos, CA,1987.

[6] D.E. Goldberg, K. Deb & Korb, B., Do not Worry, Be Messy, pp.24 - 30, Proceeding of the Fourth International Conference on Genetic Algorithms, Morgan kaufmann Publishers, Los Altos, CA, 1991.

[7] L.B. Booker, Improving Search in Genetic Algorithms, pp.61-73, Morgan kaufmann Publishers, Inc., Los Altos,CA,1987.

[8] M. Casdagli & S. Eubank, Nonlinear Modeling and Forecasting, Addison-Wesley Publishing Company, Inc., New York, 1992.

[9] L. Howard & D. D'Angelo, The GA-P: a genetic algorithm and genetic programming hybrid, IEEE Expert pp11-15, 1995.

[10] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.

[11] D. Quagliarella, J. Priaux, C. Poloni & G. Winter, Genetic Algorithms and Evolution Strategy in Engineering and Computer Science, John Wiley & Sons, New York, 1998.

[12] N.H. Packard, A genetic learning algorithm for the analysis of complex data. Complex Systems 4, n.5, pp 543-572, 1990.

[13] T.P. Meyer & N.H. Packard, Local forecasting of high-dimensional chaotic dynamics., Non linear Modeling and forecasting, Addison- Wesley, 1992.

[14] Mackey M.C. & L. Glass, Science 197, p 297, 1977.

[15] J. K. Hale, Asymptotic Behavior of Dissipative Systems, Mathematical Surveys and Monographs number 25, pp.61-71, American Mathematical Society, Rhode Island, 1988.

[16] H.W. Coleman & W.G. Steele, Experimentation and Uncertainty Analysis for Engineers, pp.34-37, John Wiley & Sons, Inc., New York, 1999.

[17] J. Taylor, An Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements, University Science Books, USA, 1982.

[18] A. D'Orazio, A. Lay Ekuakille et al., Architecture Design of an Integrated Remote Sensing Centre for Now Casting Weather Prediction in Mediterranean Area: an Overview, Proceeding of European Geophysical Society Conference, Maratea (Italy), October 1999.