▶IPS **Intelligent** Product Solutions

From: Robert Lieb
To: Christoph Gorder, Robert Lee, Joy Hung
        Charity Water
CC: Gianfranco Bonanome, Phil Salditt, Ralph Cassara, Matt Schreider, Matthew Leung
        IPS
Date: Dec 7, 2016

Subject: waterDetect.c Water Detection Algorithm

IPS designed a new waterDetect algorithm to improve the accuracy of the original AfriDev units. As before, the software measures the effective flow rate by measuring the height of the water within the well chamber. Greater accuracy was realized by taking more frequent samples, frequent re-calibrating, and estimating the percentage of each pad covered with water. The algorithm also detects and throws out outlier readings that appear to be the result of some unknown cause, but don't adversely affect the average accuracy of the measurement technique. Measurement accuracy under normal operating conditions is proven to be about +/- 2-4%, which is deemed acceptable for the intended application.

This algorithm can be deployed in the AfriDev prototype units without necessitating a complete redesign of the sensors and associated electronics. The new processor selected has the resources to include the existing AfriDev system functionality and the new algorithm. However, the existing units in the field would not be able to support this algorithm without freeing up RAM and Flash space.

1 Introduction

The AfriDev system uses a block of 6 capacitive sensor pads to measure the height of water in a water catch area within a well head. "waterDetect" software within the AfriDev unit uses these sensors to detect the presence of air or water covering each of the pads. When water is covering the pad, the measured capacitance goes down, and conversely when air is covering the pad the value goes up.

Changes in ambient temperature can change the measurements of the sensors. Additionally, the TI Capacitive Sensor library used to make capacitive measurements has been observed to generate incorrect data roughly every 20 samples.   The "waterDetect" algorithm has features to lessen the effects of these issues in the detection process.

The "waterDetect" software measures the state of the pads four times a second (every 250 mSec) and it trends this data once a second. The software determines each second how many pads are covered with water and to what extent they are covered.   Given the design of the well head, each 100% covered pad accounts for 61ml/second of water flow. The water flow estimate for partially covered pads is a percentage of this water flow rate.   Note that when the water level goes higher than the highest pad, then all the water that flows above this point is not detected.

Up to the publication of this document, the root cause of the recurring incorrect data from the

TI library has <u>not</u> been identified.   The "waterDetect" software identifies OUTLIER samples that do not appear to be valid and these samples are not used in the detection of water. Invalid samples are ones that are more than 1500 counts above or below the current mean for the pad.

Even with this OUTLIER removal criteria, occasional invalid samples are still seen.   The side effect of this is that a pad may be identified to be covered with air that is lower from another pad that detects water.   These are called "Unknown" cases in the initial software.   Given the 4 samples per second rate in the waterDetect software, it is unlikely that splashing is causing the "unknown" cases.

Tests were done at the charity: water offices in Tribeca and the accuracy of pour data was very good for pumping <u>at the rate of roughly one stroke per second</u>. The pumping was done doing 5 strokes within 5 to 6 seconds, and the rest of the time the water drained out of the well's enclosure.

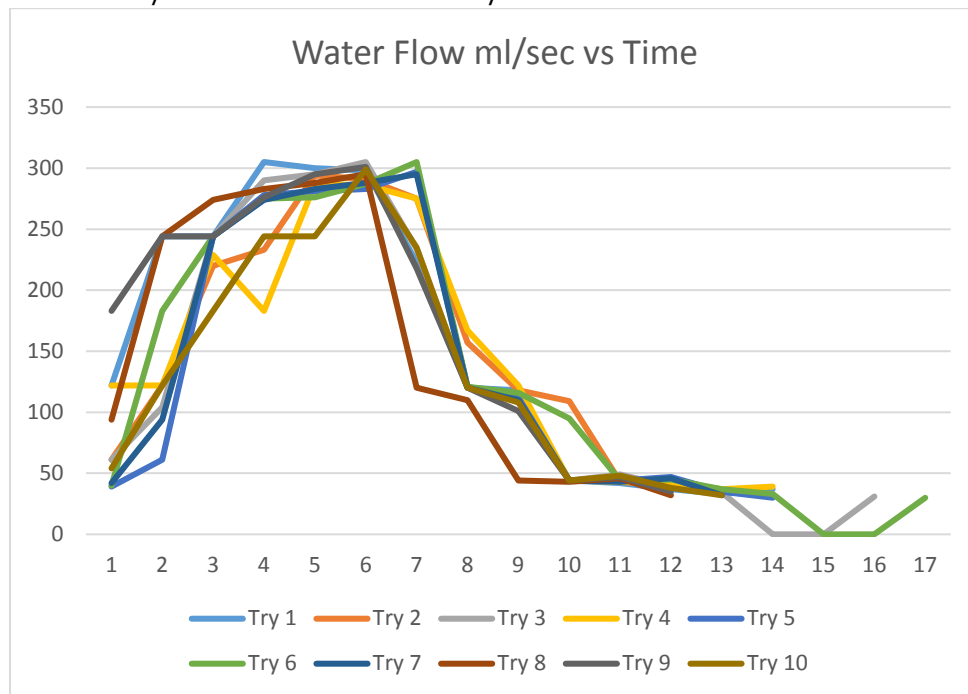Figure 1 shows the ml/second flow rates read by the "waterDetect" software in these tests

Figure 1 Water flow over time

With 10 trials to measure 2 liters of water from the well, 7 out of 10 times, the accuracy was within +/- 2% of the actual (which is +/-40 ml of water). The least accurate attempts fell within -6.8% to 5.4% accuracy. None of the attempts measured water above all 6 sensors, so no overflow condition occurred.

Figure 2 shows the results of the first round of tests.

Trial	Measured Flow	Actual Flow	Error %
1	2167 ml	2100 ml	3.2%
2	1998 ml	2000 ml	-0.1%
3	1962 ml	2000 ml	-1.9%
4	1998 ml	2000 ml	-0.1%
5	1916 ml	1900 ml	0.8%
6	2131 ml	2100 ml	1.5%
7	1910 ml	1900 ml	0.5%
8	1873 ml	1900 ml	-1.4%
9	2108 ml	2000 ml	5.4%
10	1771 ml	1900 ml	-6.8%
Overall	19834 ml	19800 ml	0.2%

Figure 2 Pour data from first set of trials

Tests were also done at a faster pumping rate, doing roughly 17 strokes in 10 seconds. At the faster rate, two problems became evident. First, the water was detected to rise above the highest pad on the detector. All of this water would not be detected, so the measured flow was off from the actual figures. Second, the well head was observed to rock on its housing, to the point where water in the collection area could conceivably register too much water flow. These two scenarios operating together yielded some interesting data. All but two of the trials where the water went above the highest pan registered "short" measurements (off by -4.5% to -12%). These trials are shaded in orange.

In these ten trials, three of the times the well was "held in place" as best as possible while the pump handle was quickly engaged. The well head still shook on the base, just not as much. Unfortunately not much could be concluded with this data. Figure 3 shows the data collected for these trials.

Trial	Measured Flow	Actual Flow	Error %	Pad 0 Overflow	Movement Restricted
1	3519 ml	4000 ml	-12.0%	Yes	Yes
2	3915 ml	4100 ml	-4.5%	Yes	No
3	3817 ml	4100 ml	-6.9%	No	Yes
4	3803 ml	4000 ml	-4.9%	Yes	No
5	4158 ml	4000 ml	4.0%	No	No
6	4072 ml	3900 ml	4.4%	No	No
7	4104 ml	4000 ml	2.6%	Yes	No
8	3700 ml	3900 ml	-5.1%	Yes	No
9	4399 ml	3800 ml	15.8%	No	No
10	4098 ml	4100 ml	0.0%	Yes	Yes
Overall	39585 ml	39900 ml	-0.8%		

Figure 3 Fast pumping trial data

2 waterDetect algorithm details

2.1 Measurement Frequency

The waterDetect algorithm takes samples four times a second.   Each second, the software takes the mean of the samples.   In a steady state, the mean would fall into one of 2 target areas: water or air. Figure 4 shows real data over time for a single pad.
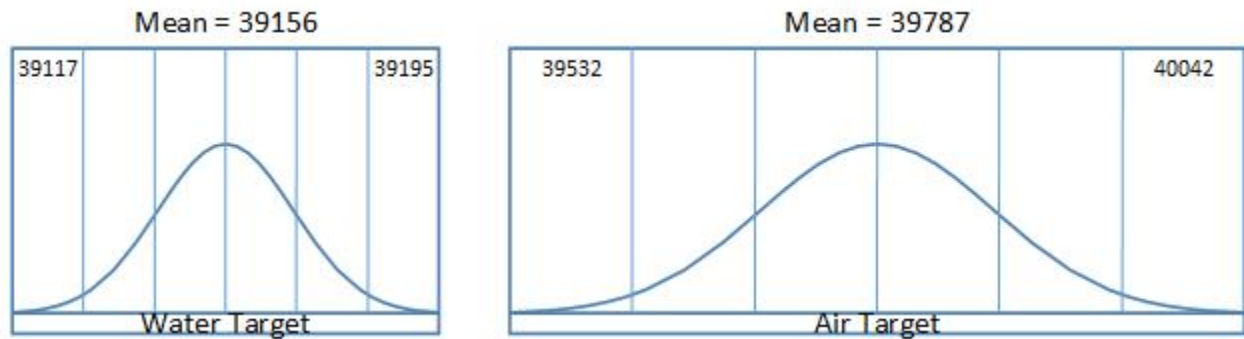

Figure 4: Normal distributions of "water" and "air" for a single Pad

The software takes four samples every second on each pad. Figure 5 shows actual data for one pad:

- In the first second, the values are all within the "water" mean.
- In the second second, there is a mix of water and air samples.
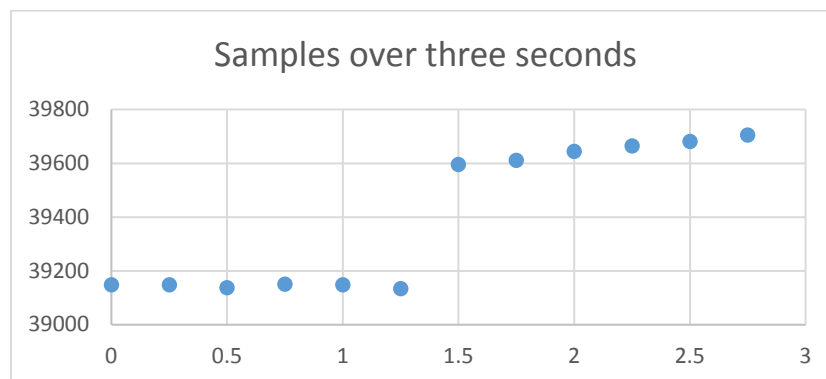- In the last second, the data is all within the "air" mean.

Figure 5: Sample values over three seconds showing jump in mean

2.2 Continuous re-calibration of targets

From one second to the next, the software looks for a "jump" in mean up or down more than 500 counts. In Figure 6, the black line shows the calculated mean for a given second of time. The orange and blue bars show changes in sample both up and down. When a negative delta greater than 500 is seen, this is detected as a "jump" to water. The blue bars show water

samples until a positive delta is seen that is greater than 500, this is detected as a "jump" to air. The orange bars show air samples.
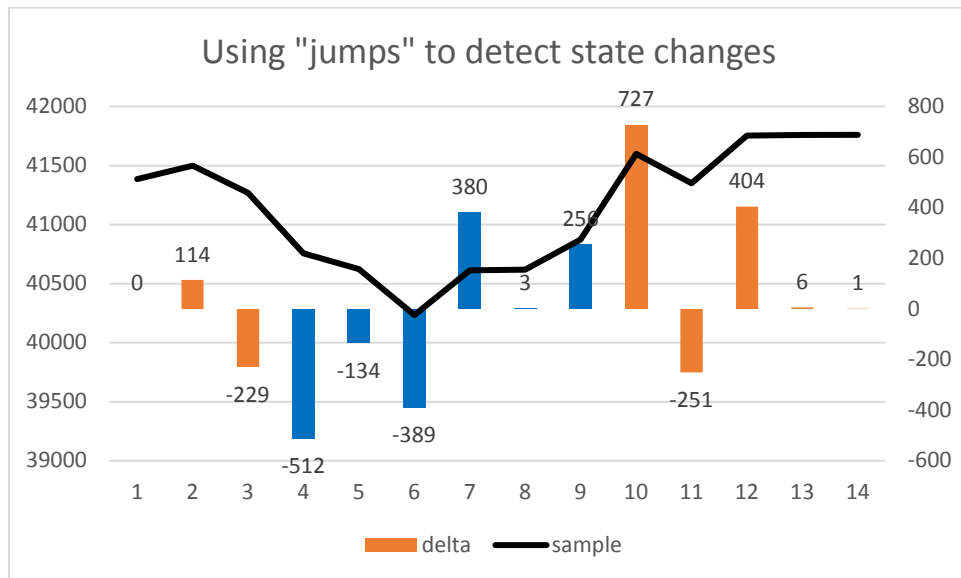

Figure 6 Using "jumps" to detect state changes

If the jump is seen, but it is within the existing target range for "air" or "water", then a new target is set only if it is within 250 counts of the existing "air" or "water" target. This kind of a jump is uncommon, usually it is a result of processing an invalid sample of data.

Using the jumps to establish new target means for "air" or "water" causes the targets to be "re-calibrated" every time the pad transitions from "water" to "air" and vice versa.   So when the ambient temperature changes this process will normalize the data for future calculations.

2.3 Midpoint based water detection

Across many seconds, the maximum target for "air" and the minimum target for "water" is tracked. A midpoint is calculated between these target values.   For every second, if the sample mean goes above the midpoint, then the current state of the pad is "air", if the sample mean goes below the midpoint, then the current state of the pad is "water" In Figure 5, if the current second's mean is above the midpoint 39472, the pad is covered with "air", and if it is below the midpoint it is covered with "water".

This is necessary for the software to handle "turbulent" water that is constantly moving up and down over the pad surfaces.

2.4 Flow rate calculation
Once a second, the software estimates the amount of water that is flowing around the pads and out of the pipe.   The flow rate is calculated by multiplying the number of pads completely covered with water by 61 ml/sec and adding the proportional flow level for the highest pad touching water (described in the next section).

Figure 7 shows the approximate flow rate for water out of the pipe related to the number of pads completely covered with water.
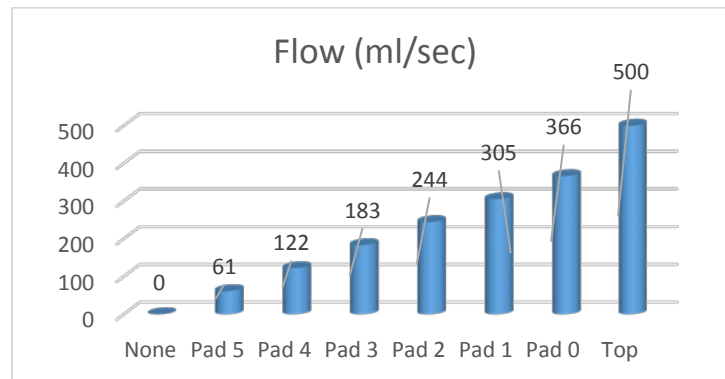

Figure 7 Cumulative Flow rates

The highest pad covered with water may only be partially covered. A proportional pad coverage (Prop%) calculation needs to be made. Figure 8 shows a scenario where the water level is up to 77% of pad 3.   For this example, the total flow for the second is 169 ml/sec.

Pad	Flow	Prop%	AdjFlow
Pad 3	61 ml/sec	77%	47 ml/sec
Pad 4	61 ml/sec	100%	61 ml/sec
Pad 5	61 ml/sec	100%	61 ml/sec

Figure 8 Adding up Flow rate using Proportional Percentage

To calculate the proportional pad coverage (Prop%) for a given pad, the current sample mean (the average of the last 4 samples read from the capacitive sensor) is compared to the range of values from the current "target water" mean to the current "target air" mean.

	Sample Mean	Prop%
target air	41764	0%
current	40643	77%
target water	40323	100%

Figure 9 Calculating Proportional Percentage

Figure 9 shows the current data for a single pad. The proportional percentage is calculated by following the steps in Figure 10:

Step	Calculation
Find the difference between target air and the current mean:	AirDiff = 1121
Find the difference between target air and target water:	PadDiff = 1441
Divide AirDIff by PadDiff:	Prop% = 77%

Figure 10 Steps to Calculate Prop%

Multiplying he Prop% figure with the pad's flow rate (61 ml/sec) yields a flow rate of 47 ml/sec as shown in Figure 8

2.5 Debug trace data format

Figure 11 shows some debug trace data, with lines added for readability

Time	Pad 0	Pad 1	Pad 2	Pad 3	Pad 4	Pad 5	Level/Flow	Prop%
T0167	0(?)9f86	1(a)a212	2(a)9ce2	3(a)a33b	4(a)abc9	5(a)b3a2	L0F000	P000
T0168	0(?)9f85	1(a)a172	2(a)9e31	3(a)a1a6	4(a)abc9*	5(a)b3a2*	L0F000	P000
T05a0	t=027C							
T0169	0(?)9f90	1(a)a211	2(a)9ce8	3(a)a343	4(a)abca	5(A)b3a6	L0F000	P000
T016a	0(?)9f6b	1(a)a1f0	2(a)9cd2	3(a)a252	4(W)a955	5(W)b0f4	L2F122	P100
T016b	0(?)9f50	1(a)a1cb	2(a)9c05	3(a)a112	4(W)a6ea	5(W)ae48	L2F122	P100
T016c	0(?)9f22	1(a)a0cd	2(w)9a2c	3(W)9ebe	4(W)a6d2	5(W)ae40	L4F229	P076
T016d	0(?)9eef	1(W)9f46	2(a)9ba3	3(W)9ea3*	4(W)a6d0*	5(W)ac63	L3F183u	P100u
T016e	0(?)9e4a	1(w)a010	2(W)9945	3(w)9ea9	4(W)a6cb	5(w)ae3c	L5F287	P072
T016f	0(?)9e4c	1(w)a024	2(w)995b	3(w)9ea4	4(W)a6cb	5(w)ae3d	L5F286	P069
T0170	0(?)9e4d	1(w)a0ae	2(w)998e	3(w)9eab	4(w)a6cc	5(w)ae3c	L5F275	P051
T0171	0(?)9f4d	1(a)a1c4	2(A)9bff	3(w)9fc1	4(w)a6d8	5(w)ae3d	L3F167	P075
T0172	0(?)9f6c	1(a)a151	2(A)9cd6*	3(a)a160	4(W)a61d	5(w)acad	L2F122	P100
T0173	0(?)9f76	1(a)a1fb	2(A)9ce0	3(a)a327	4(A)a959	5(w)ae54	L1F044	P073
T0174	0(?)9f7b	1(a)a201	2(A)9ce5	3(a)a329	4(A)ab20	5(w)ae63	L1F043	P072
T0175	0(?)9f7e	1(a)a205	2(A)9ce8	3(a)a329	4(A)ab9e	5(w)ae9d	L1F042	P069
T0176	0(?)9f80	1(a)a207	2(a)9ce7	3(a)a330	4(A)abb4	5(w)af2b	L1F037	P061
T0177	0(?)9f82	1(a)a166	2(A)9e34	3(a)a1a2	4(A)abbc*	5(w)aef4	L1F039	P064
T0178	p=01998ml							
T0178	0(?)9f87	1(a)a20e	2(a)9ced	3(a)a338	4(A)abbe	5(a)b0d6	L0F000	P000

Figure 11 Sample Debug Trace Information

Each pad has the current state in parenthesis (a) = air, (w) = water, (?) unknown; followed by the current sample mean in hexadecimal.

The detected water level is the number of pads covered in water. Pads are covered in water from Pad 5 down to Pad 0 and they are marked in the debug data with a "w" or "W".

The last column of data in the debug trace is the Prop% figure for the highest pad covered with water. This is an estimate of what percentage of the pad is covered with water.

Once a minute the internal air temperature is calculated and displayed in the "Pad 0" column starting with "t=" , this is shown in Figure 11 shaded in yellow

When water is detected, the flow measurements for each second are added together. When the flow of water stops, the total flow is displayed in the "Pad 0" column starting with "p=". They are shown in Figure 11 shaded in dark blue.

An asterisk "*" is shown next to pads where one or more data points were erased as an OUTLIER. They are shown in Figure 11 shaded in green.

The debug trace will show a capital "W" or "A" when the mean is at the target for water or air respectively. The trace will show lowercase "w" or "a" when the mean is below or above the midpoint, but not at the target. In the start of the data collection the trace will show a question mark "?" as the pad's state. The software does not know whether the first value is "air" or

"water" until a jump in mean is seen.

2.6 Filtering of "outlier" data

In the waterDetect software, we see a recurring problem where the capacitance measurement is not read for the full 16 mSec period. This yields incorrect sample measurements. They usually happen every 20 times the capacitive measurements are done. There is no time constraint on this, it happens the same way whether the group of 4 samples are repeated every second or every 20 seconds.

The waterDetect software filters the collected sample data in two ways:

1. In our investigations, the sample values for this board design are observed to be between 39,000 and 48,000 counts. To be safe, before a mean is calculated, the code "erases" outlier samples whose values are not between 37000 and 50000.

2. The current software will drop any sample value that is more than 1500 counts away from the last mean calculated.

Since the code takes 4 samples a second, occasionally erasing one sample per second does not affect the accuracy of the measurements. However, "outlier" data that is not detected will occasionally cause "Unknown" cases to be seen.

2.7 Low and High Frequency Sampling

The system will calculate a new water level once a second as long as water is covering at least 1 pad. If 5 minutes elapses with no water seen, then the system will stop measuring samples for 20 seconds. At this point one second's worth of data (4 samples) are measured and if water is seen then, then the system will resume higher frequency sampling.