Surface and Airborne Monitoring Technology for Detecting Geologic Leakage in a CO2-Enhanced Oil Recovery Pilot, Anadarko Basin, Texas

Final Technical Report

Reporting Period Start Date – October 1, 2013
Reporting Period End Date – March 31, 2017

Principal Author(s): Peter E. Clark, Jamey Jacob, M. Tyler Ley, Nicholas F. Materer, and Jack Pashin

Date Report was Issued (June 30, 2017)

DOE Award Number: DE-FE0012173

Oklahoma State University
201 Advanced Technology Research Center
Stillwater, OK 74078

Abstract

Surface and airborne gas monitoring programs are becoming an important part of environmental protection in areas favorable for subsurface storage of carbon dioxide. Understanding structural architecture and its effects on the flux of fluids, specifically CO_2 and CH_4, in the shallow subsurface and atmosphere is helping design and implement next generation monitoring technologies, including unmanned aerial vehicles (UAVs). An important aspect of this research is using subsurface fracture data to inform the design of flight pathways for UAVs in the Farnsworth Oil Unit of the Anadarko Basin. The target zone for CO_2 storage and enhanced oil recovery in the Farnsworth Oil Unit is in the upper Morrow sandstone at subsurface depths greater than 2,000 m. Field study reveals that sandstone and chert in the High Plains Aquifer contain numerous joints that provide crucial insight into aquifer architecture and subsurface flow pathways. Properties of more than 1,700 joints were measured in the field and in high-resolution satellite images. Two distinctive joint systems interpreted as a conjugate pair were identified in the study area. Joint spacing follows a log-normal statistical scaling rule. These fractures appear to be the product of an E-NE regional compressive stress and may have a significant effect on flow in the High Plains Aquifer system. Based on the results of this research, design of UAV flight paths should be oblique to fractures in a way that maximizes the likelihood of CO_2 and CH_4 flux from of systematic joints and cross joints.

Detection of leaks either from subsurface seal failures or surface equipment problems is important to insure that stored CO_2 is not being lost to the atmosphere. In the case of surface leaks, costs go up and injection efficiency goes down. Surface leaks are often short term caused by seal leaks, pop off valves, or other equipment failure modes. In either case, CO_2 is released to the atmosphere. This projects primary goal was to explore the use of low cost sensors coupled with the appropriate hardware and software to produce a low cost system capable of detecting both long and short term leaks. In addition, an aerial system was designed to measure the CO_2 and CH_4 distributions above the surface. Both the ground and UAV based systems were designed, tested and deployed in the Farnsworth field.

Contents

Executive Summary

The overall goal of this project was to develop and field test a leak detection system constructed from low-cost sensors to measure carbon dioxide and methane levels and to show that these sensors would provide adequate performance to detect leaks over a long period of time and weather conditions. The original proposal described a system that stored data onboard each sensor node. Using this design it would have been necessary to periodically travel to Farnsworth to collect the data. During the design and evaluation phase of the project, it was determined that for a nominal extra cost the units could be connected, using a WiFi and cellular based system, to communication nodes that would send the data each evening to a computer on the OSU campus. This allowed the planned onboard sensor node memory to serve as a backup incase the communication system failed. The cost savings in travel and manpower more than paid for the extra electronics.

As a design goal, the sensor input subsystem was to be flexible enough to change or upgrade the sensor package as new sensors become available. There is nothing in the design that limits the sensors to carbon dioxide and methane. Any number of other gases could be monitored with the addition of the proper sensors. When the proposal was under development, the plan was to put the sensors on short poles at interval around the site. Farming methods used in the area precluded this because the local farmers tend to farm right to the edge of the roads. A mounting bracket was designed and permission to mount the units on utility poles in the area was secured. This proved to be quite successful until a late winter ice storm hit the region. Most of the sensor units survived, but many of the power poles did not.

As part of the project, an airborne monitoring system was developed. This unit was tested in a number of different situations besides the injection site. It was used over a controlled burns conducted in the Stillwater area and at. the OSU small craft airport. Land-based units were also deployed at the OSU airport. This was the site where controlled release testing was accomplished.

Massive amounts of data were collected. Over eighteen thousand data points a day from the Farnsworth site alone. Data for CO_2 and CH_4 , temperature, pressure, and humidity were measured at fifteen minute intervals for several months. The reliability of the units was outstanding. Failure rates of the units was one percent or less. Battery failure proved to be the biggest problem. The communications nodes used a more expensive sensor system in addition to the sensor set used in the sensor nodes. Unfortunately, the more expensive sensor system had an unexplained high failure rate. Data was still acquired from these nodes using the sensor node sensors. As mentioned earlier, approximately 95% the units survived an ice storm when many of the power poles they were mounted on did not. The two units that were lost sustained major damage to the cases, but not the sensors or electronics. All aspects of the project were successful. The aerial system proved to be rugged and reliable. Using inexpensive sensors in the ground based sensor units worked well for leak detection. A number of events at the Farnsworth site were recorded during the study. These, however,

were short term in nature, which would indicate that the CO_2 came from non-geologic sources. One would expect a geologic leak to produce a baseline level shift in at least one sensor node. No baseline shifts were detected in any of the sensor nodes. When taken as a whole, the project was successful. The sensors and associated electronics were stable and reliable. Costs for each of the sensor nodes including communications capability were an order of magnitude less than one of the commonly used commercial sensors.

1 Geological Considerations For Leak Detection at Farnsworth

1.1 Introduction

Ensuring safe, permanent storage of CO_2 is vital for the success of subsurface geologic CO_2 storage projects. The development of robust monitoring technology is vital for validating storage permanence, as well as for ensuring the integrity of storage operations. Accordingly, monitoring programs are considered essential for meeting the goals of CO_2 emissions reduction, environmental protection, and human health and safety (NETL, 2012). This study is a product of an investigation sponsored by the U.S. Department of Energy, National Energy Technology Laboratory that aims to advance the state of the art of surface and airborne monitoring and includes the deployment of low-flying unmanned aerial vehicles (UAVs) for near-surface detection of CO_2 and CH_4 plumes emanating from the land surface.

The study area is Farnsworth Oil Unit in the Anadarko Basin of the northeastern Texas Panhandle (Fig.1.1). The Southwest Regional Carbon Sequestration Partnership (SWP) and Chaparral Energy, LLC, are conducting CO_2-enhanced oil recovery operations in the Farnsworth Oil Unit. Recovery operations are being carried out at a depth $>$ 2,000 m in the Pennsylvanian-age upper Morrow sandstone. Aquifer protection is the primary concern of underground injection regulations in the United States. The High Plains Aquifer is the primary underground source of drinking water (USDW) in the study area and consists primarily of the Ogallala Formation. Characterizing the geologic framework and fracture architecture in the Farnsworth area will aid in the design and implementation of the surface and airborne monitoring program, as well as the interpretation of the monitoring results. Although several publications have discussed the general structural framework of the Texas Panhandle (Seni, 1980 ,Gustavson and R.T.Budnik, 1985 Gustavson, 1986), a need remains to examine the surface and shallow subsurface fracture architecture in the study area. To the authors knowledge, there has been no detailed description or quantitative analysis of the fractures in the High Plains Aquifer or the related strata in the northeastern Texas Panhandle. To help fill this gap, this study characterizes the stratigraphic framework, structural framework, and fracture architecture of the High Plains Aquifer and related strata in the Farnsworth area. This analysis will provide insight into the ways that fracture architecture may affect the movement of gas and water, and the results will be used to design and optimize surface

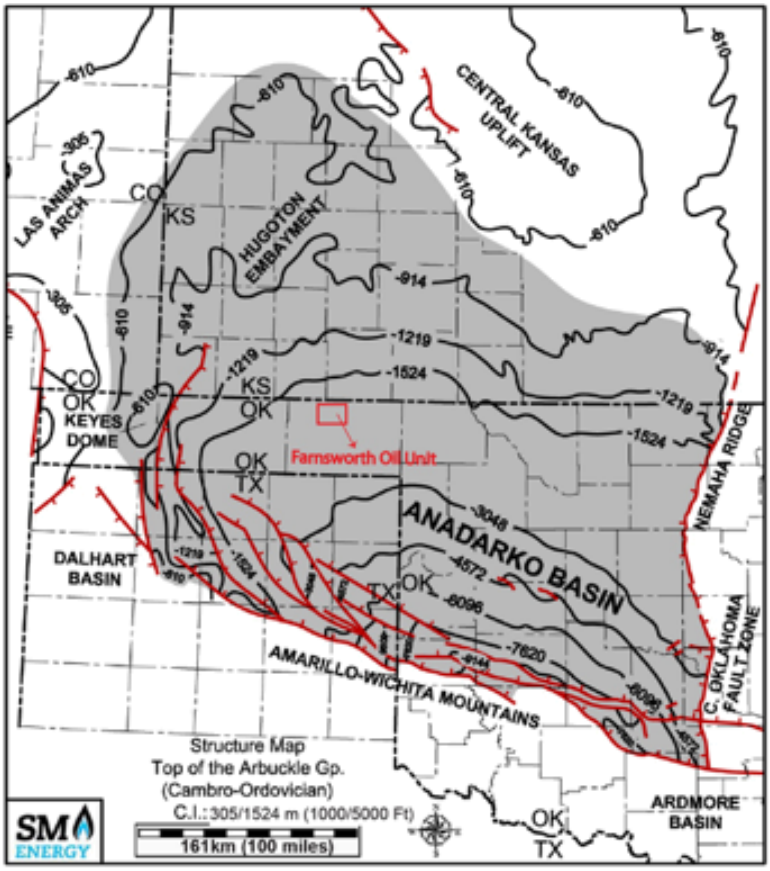


Figure 1.1: Arbuckle structure map of the Anadarko basin. Red box shows location of the study area. Contour interval variable (305-1524 m; 1000-5000 ft) in deeper part of basin (modified from Davis and Northcutt, 1989).

and airborne monitoring programs.

This study focuses on identifying faults, fractures, and other geologic discontinuities that could serve as shallow subsurface flow paths that could affect the flux of water and gas, in and around the Farnsworth Oil Unit. Basic geologic data were compiled and used to characterize and analyze the stratigraphic and structural framework of the High Plains Aquifer in the test area. Surface strata were described and measured in outcrops near the Farnsworth Oil Unit. Subsurface analysis focused on characterizing the subsurface geologic framework and identifying any discontinuities that may be indicative of enhanced fracturing and small-scale faulting. Statistical methods were used to characterize the natural fractures at the surface. To support the surface and airborne monitoring program, optimal flight pathways for UAVs were designed based on the subsurface fracture architecture characteristics. To facilitate this objective, two conceptual models have been developed to explain how shallow subsurface processes affect the flux of CO_2 and CH_4.

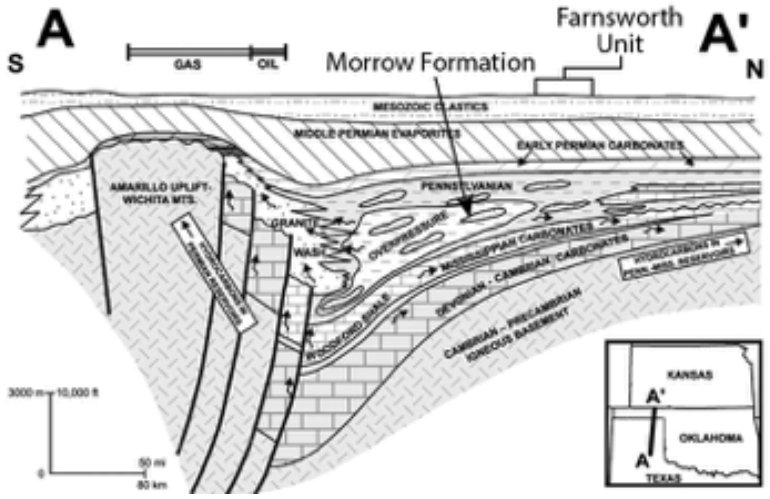


Figure 1.2: Generalized subsurface cross section showing the geologic setting of Morrow sandstone in the Anadarko Basin

1.2 Geological setting

The Anadarko Basin is one of the deepest and most productive petroliferous basins in the United States. The basin is a northwest-elongate, asymmetrical synclinorium extending from west-central Oklahoma into the Panhandle region of Oklahoma and northern Texas (Wang and Philp, 1997, Fig. 1.1). It is bordered by the Nemaha Uplift in the east, the Amarillo-Wichita Mountains in the south, the Cimarron Arch in the west, and the central Kansas uplift in the north (Wang and Philp, 1997, Carter et al., 1998). The Farnsworth Field is developed in the southern part of the Anadarko shelf (Fig. 1.2), where the strata dip homoclinally southward at about 0.7 degree (12 m/km), and lack any major folds and faults. Also, the Farnsworth Oil Unit is north of the major overpressure region in the heart of the basin.

The Farnsworth Oil Unit is located in west-central Ochiltree County, Texas, approximately 11 km south and 16 km west of the town of Perryton. The oil unit has produced more than 36 million barrels of oil and 27 billion cubic feet of gas since 1955 and is historically the most productive upper Morrow oil field in the Anadarko Basin (Munson, 1990). The thickness of the upper Morrow sandstone in the Farnsworth Oil Unit is 10 to 20 m, and the reservoir is composed of quartzarenitic sandstone and conglomerate. The sandstone is enveloped by shale containing siltstone laminae and this shale forms the topseal for the upper Morrow oil accumulation and also is the principal confining unit for CO_2 injection.

Paleozoic sedimentary rocks in the basin range in thickness from 3,000 m to about 12,000 m (Kennedy et al., 1982). The Paleozoic basin fill is at its thickest along the southwestern margin of the basin. Pre-Pennsylvanian strata of the Anadarko Basin consist primarily of shallow marine carbonate rocks interbedded with lesser amounts of sandstone and shale (Fig. 1.3). Numerous shale units with thickness in the order of 100 m are present in the

Middle and Upper Pennsylvanian sections and provide secondary confining units that help ensure the containment of injected CO_2. Pennsylvanian and Permian strata compose most of the overburden in the Farnsworth Field. In the Permian section, moreover, the Red Cave evaporites form a regionally extensive sealing stratum that helps isolate the High Plains Aquifer from development activities in the deep subsurface (Hill, 1984). Chaparral Energy began CO_2-enhanced oil recovery operations in the Farnsworth Unit in 2010, and this effort includes the large-scale Phase III field test being performed by the SWP. Phase III tests aim to verify safe, permanent capture, transportation, injection, and storage of CO_2 at scales on the order of 10^5 to 10^6 tons (NETL, 2013). Neogene strata disconformably overlie Permian strata in the study area and are assigned mainly to the Ogallala Formation (Fig. 1.3), which is thought to be of Miocene to Pliocene age (Gustavson and D.A.Winkler, 1988). The Ogallala is the principal formation of the High Plains Aquifer, also called the Ogallala Aquifer, which is composed of poorly to moderately indurated calcareous sandstone (Gustavson and D.A.Winkler, 1988, Johnson et al., 1989,Gustavson and V.T.Holliday, 1999). The High Plains Aquifer extends from South Dakota into the Texas Panhandle; it is the primary underground source of drinking water (USDW) in the southern High Plains (Mehta, Fryar, and Banner, 2000) and thus a resource that needs to be protected. The Ogallala Formation thickens westward from 100 m to about 275 m in the northeastern Texas Panhandle and covers the western third of the Anadarko Basin (Cunningham, 1961, Seni, 1980). Reeves (Reeves, 1972) and Gustavson and Winkler Gustavson and D.A.Winkler, 1988 described lower Ogallala Formation as consisting primarily of laterally extensive, heterogeneous, vertically stacked successions of gravel and sand. A thick succession of very fine sand and loamy sand, silt, and clay compose the upper part of the Ogallala. Ripple cross-laminae and cross-beds have been reported in some outcrops (Gustavson and D.A.Winkler, 1988). Siliceous and calcareous root tubules (rhizoconcretions) and calcareous nodules (glaebules) are common in the Ogallala Formation. Quartz pebbles have been found locally in the Ogallala (Bretz and Horberg, 1949).

Pliocene and Quaternary strata in the area disconformably overlie the Ogallala Formation (Fig. 1.3). Locally, the base of the Quaternary section contains loamy strata that can be difficult to distinguish from Ogallala strata and may be assignable to the Blackwater Draw Formation and other units (Gustavson and D.A.Winkler, 1988, Gustavson and V.T.Holliday, 1999). A widespread caliche zone is present at the top of the section throughout the study area (Gustavson and D.A.Winkler, 1988). The caliche locally contains chert as well as some gypsum. The caliche ranges in age from Pliocene to Holocene, and the well-lithified Pliocene caliche section is commonly referred to as the caprock (Reeves, 1970). The modern soil profile consists of a dark, loamy epipedon (upper soil layer) that grades downward into caliche, and supports a range of agricultural activities in the Farnsworth area. Surface topography in the northeastern Texas Panhandle, is nearly flat, except for areas where streams dissect the caprock and the upper Ogallala Formation. The land surface slopes imperceptibly east-southeast and may have been controlled primarily by the elevation of the pre-Ogallala surface or local subsidence due to the dissolution of the Permian evaporites (Seni, 1980, Gustavson,

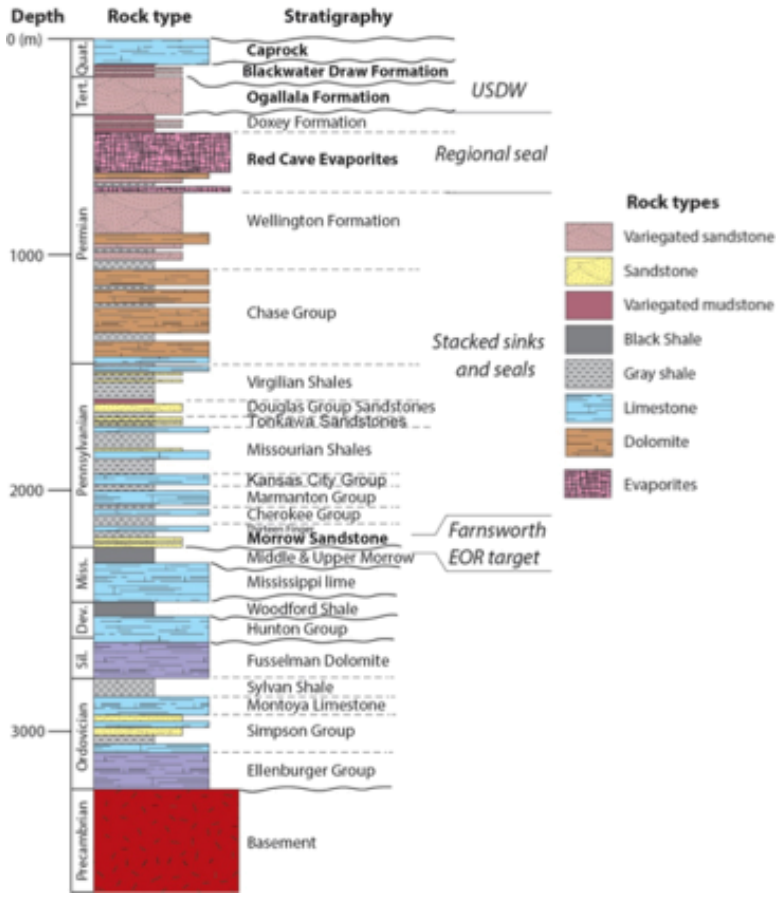


Figure 1.3: Stratigraphic column of the Anadarko Basin in the Farnsworth Oil Unit showing the position of the Morrow injection target relative to the shale and evaporite seals and the USDW in the Ogallala Formation.

1986). Geologic structures in the Paleozoic fill of the Anadarko Basin do not appear to be expressed in the post-Permian section. Little information on fracture networks is available in the northeastern Texas Panhandle. Farther west, however, fractures strike northeast and northwest and may be related to regional tectonic stresses (Gustavson and R.T.Budnik, 1985).

1.3 Methodology

The analytical approach for this research consists of two main components: (1) stratigraphic analysis and (2) structural analysis. This approach is designed to characterize the geologic framework of the Farnsworth area from reservoir depth to the surface. The results of this analysis were then used to formulate conceptual models that help guide the design and implementation of surface and airborne monitoring technologies in the Farnsworth Oil Unit.

1.3.1 Stratigraphic and Sedimentologic Analysis

Outcrop and laboratory analyses were performed to help characterize the High Plains Aquifer and associated strata in the Farnsworth area. No outcrops are available within the oil unit, where the topography is effectively flat and the bedrock is concealed below an agricultural landscape. However, the upper part of the Ogallala Formation and caprock strata are well exposed where streams have dissected the topography along Palo Duro Creek (northwest of the oil unit) and along Wolf Creek (east of the oil unit), and the Canadian River (south of the oil unit; Fig. 1.4). Outcrops were described and sampled in the summer of 2014. Access to outcrops along the Canadian River is limited, and so most detailed work focused on Palo Duro Creek and Wolf Creek. Detailed descriptions and measured sections were made using standard stratigraphic and sedimentologic procedures, and photographs and notebook sketches were made to characterize the strata. Color, grain size, texture, thickness, and bedding, as well as physical, biological, and diagenetic sedimentary structures were recorded. Seventeen hand samples were collected for petrographic thin section analysis. Hand samples and thin sections were described to determine the color, grain size, texture, framework composition, cementation, and porosity. The standard Chayes point count method (Chayes, 1949, Chayes, 1956) was employed for thin section analysis on a polarizing microscope. At least 300 points per thin section were counted to characterize framework composition and porosity. The results were then plotted on ternary diagrams to classify the sandstone.

The shallow subsurface stratigraphic framework of the study area was investigated by

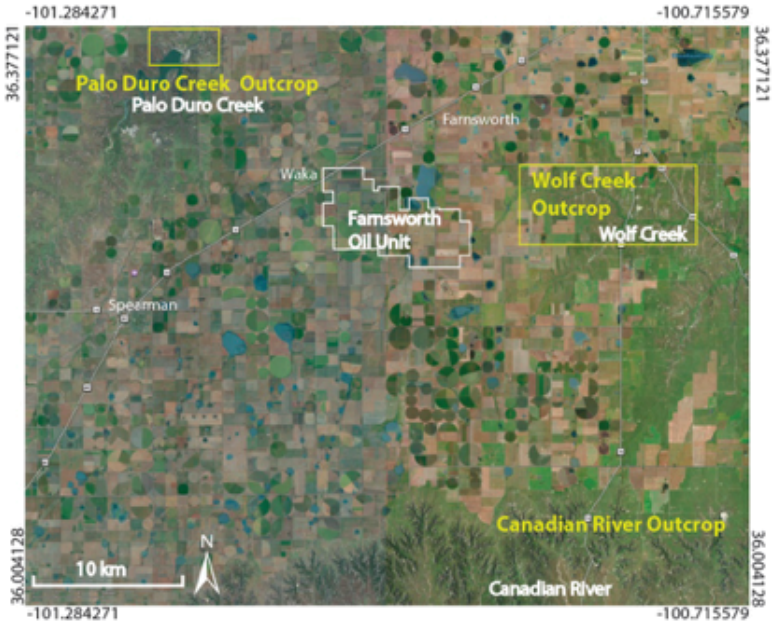


Figure 1.4: Landsat Thematic Mapper (TM) image of the study area showing the location of the outcrop areas relative to that of the Farnsworth Oil Unit.

studying water well records. Records from 147 water wells in the Farnsworth area, including the adjacent outcrop belts, were provided by the SWP. The format and quality of these records vary substantially, and many record basic lithologic information, which was used to identify Permian redbeds, the Ogallala Formation, and caprock strata. Isopach maps of the Ogallala Formation and caprock strata were contoured to determine aquifer and caprock thickness based on interpretation of the water well records. The maps were constructed using ArcGIS and refined in Adobe Illustrator.

1.3.2 Structural Analysis

Structural contour maps and a surface elevation map were constructed to identify any flexures or discontinuities that may be indicative of enhanced fracturing and small-scale faulting in the study area. Surface elevation data were derived from a digital elevation model (DEM) and mapped using ArcGIS software. A DEM from the U.S. Geological Survey National Elevation Dataset (10 m grid spacing) was used. Structural contour maps of the base and top of the Ogallala Formation also were made using the interpreted water well records.

Numerous fractures were observed in the field and in Landsat Thematic Mapper (TM) imagery with a location accuracy of <30 m. Due to limitations of outcrop quality and accessibility, most fracture attribute data were collected from satellite imagery. Orientation and spacing of more than 1,700 joints were measured, and cross-cutting relationships were analyzed using high-resolution (up to 0.5m) Texas Orthoimagery Program satellite imagery in Google Earth software. Fracture traces were measured where imaged clearly, and the orientation and spacing data were recorded in a spreadsheet. Fracture orientation, length, and spacing were recorded using measurement tools in Google Earth. Fractures in the study area tend to be vertical, and so orientation was defined by the azimuth of the surface trace of each fracture.

Once the fracture data were recorded, they were analyzed using basic population statistics to characterize orientation and spacing. Structural analysis of fracture orientation data was accomplished using Stereonet software (Allmendinger et al., 2012) to generate rose diagrams and calculate directional statistics, including vector mean azimuth, vector magnitude, and angular standard deviation following the methods of Krause and Geijer (Krause and Geijer, 1987). Spacing of the joints was analyzed statistically using Microsoft Excel to identify the governing scaling rules. Due to incomplete exposure of long fractures, only the partial length of systematic joints could be observed. Accordingly, no rigorous assessment of systematic fracture length could be made. the variability of fracture strike provides some constraint because fractures of the same set do not cross.

1.4 Results

1.4.1 Stratigraphic framework

Outcrops near the Farnsworth unit include exposures of Ogallala sandstone and caprock strata, including chert and caliche (Figs. 1.5 -1.7). The best exposures, which offer panoramic views of these strata, are along the Palo Duro Creek in eastern Hansford County and along the Canadian River in Roberts County (Fig. 1.7). The eastern outcrops in the Wolf Creek drainage system have lower relief and are more weathered but also provide instructive exposures. Also, chert was only observed along Palo Duro Creek, whereas caliche rests directly on Ogallala sandstone in the Wolf Creek area.

Three lithofacies were identified in the Ogallala Formation: 1) massive sandstone; 2) argillaceous sandstone; and 3) indurated sandstone. The massive sandstone lithofacies is best exposed in the lower part of the section along Palo Duro Creek (Fig. 1.5). The facies contains thickly bedded, pink to reddish yellow (7.5YR 8/4 to 7.5 YR 7/8), medium to fine grained sandstone. Grain size typically fines upward with the color of the sandstone becoming lighter. The sandstone is friable, poorly sorted and calcareous, and commonly contains pebble-size caliche nodules (Fig. 1.6A). Also, the sandstone is massive; sedimentary structures including cross-beds and horizontal laminae were observed locally. Biogenic structures including root tubules and sparse meniscate burrows were identified in this facies. The massive sandstone lithofacies contains few natural fractures.

The argillaceous sandstone lithofacies is widespread in the Wolf Creek area and consists primarily of thickly bedded, pinkish white to pink (7.5YR 8/2 to 7.5 YR 8/4), fine grained, and silty sandstone. The sandstone is argillaceous and calcareous, friable and poorly to moderately sorted. Abundant calcite cemented rhizoconcretions (Fig. 1.6B) and pebble- to cobble-size caliche nodules are common in the sandstone. The argillaceous sandstone lithofacies contains few natural fractures. The indurated sandstone lithofacies is resistant to weathering and forms distinctive ledges and pavements in the Wolf Creek area. The sandstone is thickly bedded ($\approx$ 1-2 m), white to pinkish white (7.5YR 8/1 to 7.5 YR 8/2), and fine to very fine grained. The sandstone is moderately indurated, calcareous, and moderately to well sorted. Pebble to cobble-sized caliche nodules and rhizoconcretions are locally abundant in this lithofacies. Abundant jointing is a salient feature of the indurated sandstone (Fig. 1.7A). The fractures are linear to curvilinear in plan view. In profile, the joints tend to terminate at the top of the indurated sandstone and extend downward into the more friable sandstone of massive and argillaceous sandstone lithofacies.

Thin section analysis demonstrates that all three lithofacies of Ogallala Formation consist of medium- to very fine-grained sandstone (Fig. 1.8a). Framework grains range in size from coarse sand to silt, and the sand grains are poorly to moderately sorted, angular to well

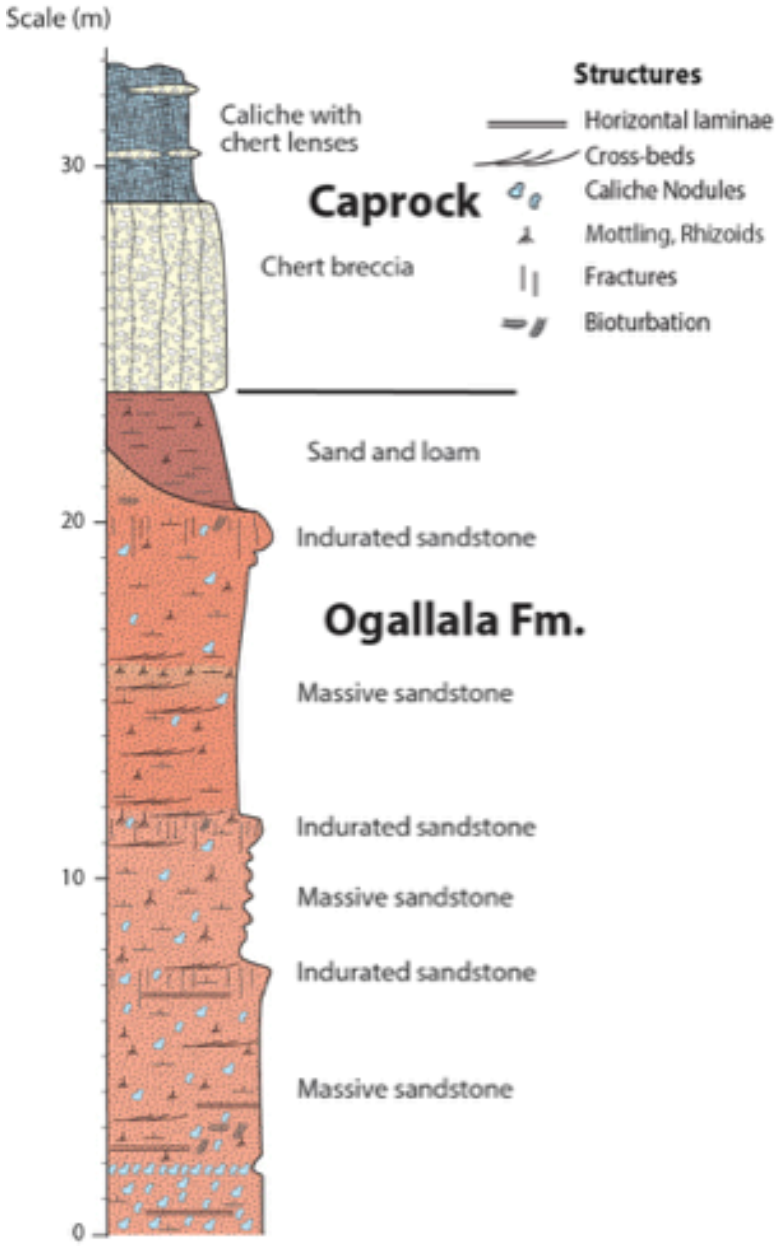


Figure 1.5: Composite stratigraphic column showing the rock types and Ogallala sandstone lithofacies along Palo Duro Creek

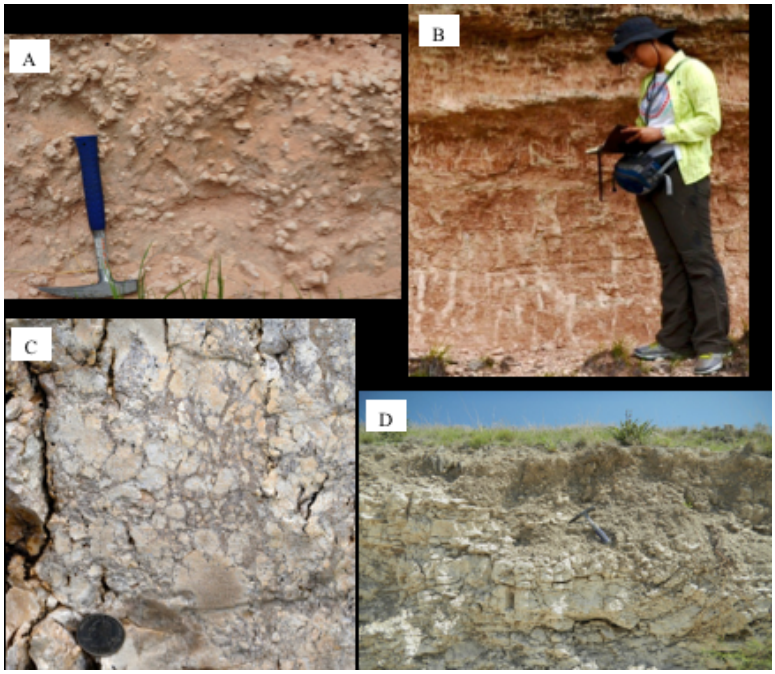


Figure 1.6: Field photographs of Ogallala and caprock strata in the field area. A) Caliche nodules in the massive sandstone facies of the Ogallalla Formation, Palo Duro Creek. B) Subvertical, branching root tubules in the argillaceous sandstone facies of the Ogallala Formation, Wolf Creek. C) Chert breccia (silicified collapse breccia) exposed along Palo Duro Creek. Open fractures are preserved in the breccia. D) Caliche overlain by modern argillaceous soil profile in the Wolf Creek outcrop area.

rounded, and have variable sphericity. Monocrystalline quartz is the dominant framework constituent, forming up to 95% of the sandstone. Feldspar generally accounts for less than 5% of the framework grains, but locally forms 17% of the sandstone. Sedimentary rock fragments generally constitute less than 1% of the framework grains. Plotting sandstone composition on the QFL ternary diagram of Folk (Folk, 1980) indicates that sandstone is mainly subarkose with some quartzarenite (Fig. 1.9). The sandstone is texturally quite immature, suggesting recycling from nearby sources of sediment and minimal diagenetic alteration of the framework grains. The sandstone is cemented by a mixture of clay and micritic carbonate, and cement content ranges from 13-41%. The clay-micrite mixture is typically expressed as grain coatings and interparticle cement. Interparticle porosity predominates in the sandstone, and point counting indicates that porosity ranges from 11% to 30%. Cement content is highest in caliche nodules, where framework grains commonly float in micritic matrix.

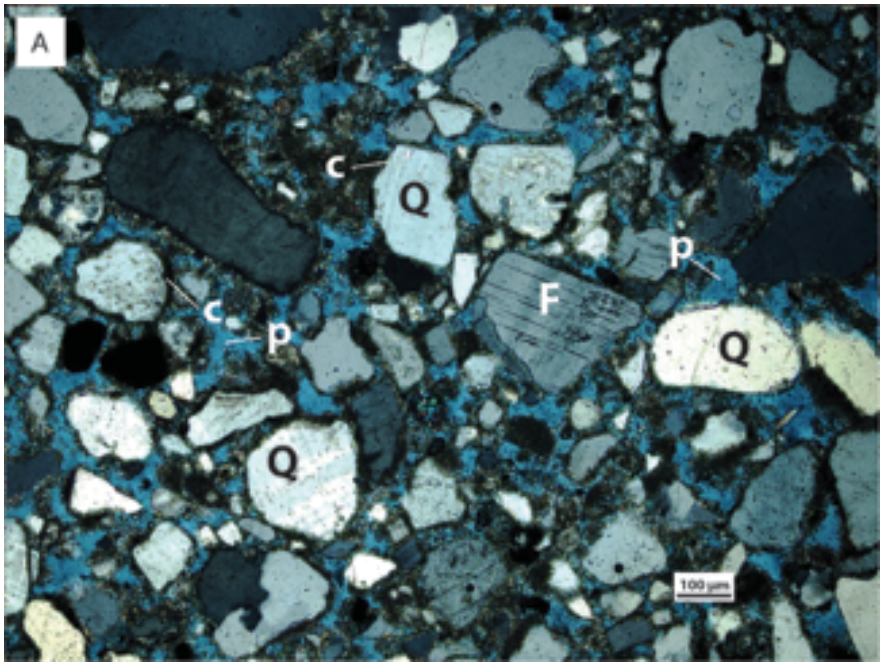
Some strata of indefinite affinity have been identified in the western outcrop area between the Ogallala Formation and the caprock succession. Light gray to red loamy deposits were identified that may be assignable to other stratigraphic units above Ogallala Formation

(a)

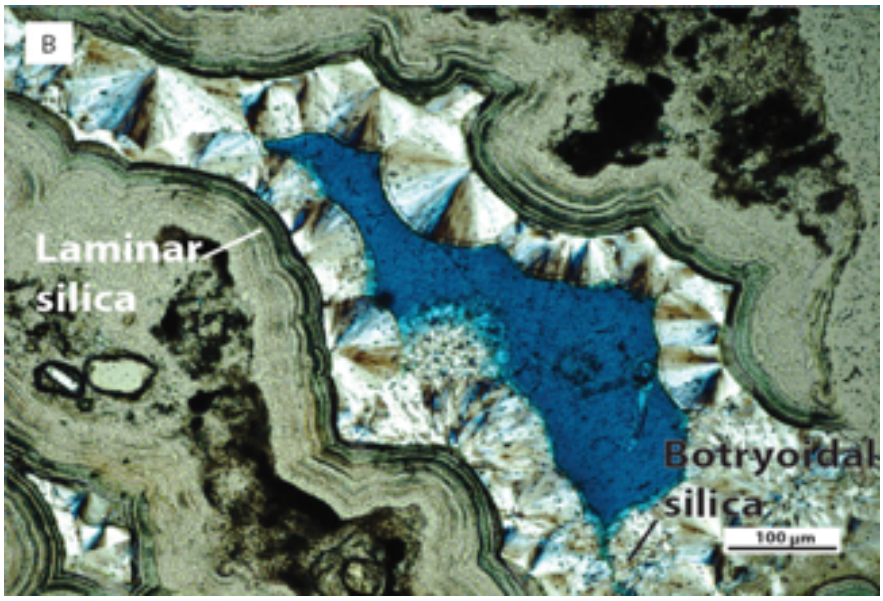
(b)

Figure 1.7: Ogallala and caprock exposures in the Farnsworth area, northeastern Texas Panhandle. a) Chert caprock overlying the Ogallala Formation along Palo Duro Creek. Note closely spaced joints in the indurated sandstone lithofacies and more widely spaced joints in the chert caprock. b) Caliche caprock overlying the Ogallala Formation along the Canadian River valley south of the Farnsworth Oil Unit.

(Fig. 1.5). These strata are generally thinner than 4 m and appear to be preserved as broad, shallow channel fills. Caprock strata were subdivided into two lithofacies: 1) chert caprock and 2) caliche caprock. About 5 m of chert breccia sharply overlies the Ogallala Formation



(a)



(b)

Figure 1.8: Photomicrographs of Ogallala sandstone and siliceous caprock in cross-polarized light. a) Thin section of the massive sandstone lithofacies at Palo Duro Creek showing abundant primary porosity (p) and grain coatings and interparticle cement composed of a mixture of clay and carbonate (c). Note grains of quartz (Q), feldspar (F), primary porosity (p). b) Thin section of the chert breccia at Palo Duro Creek showing the diagenetic features in a void fill.

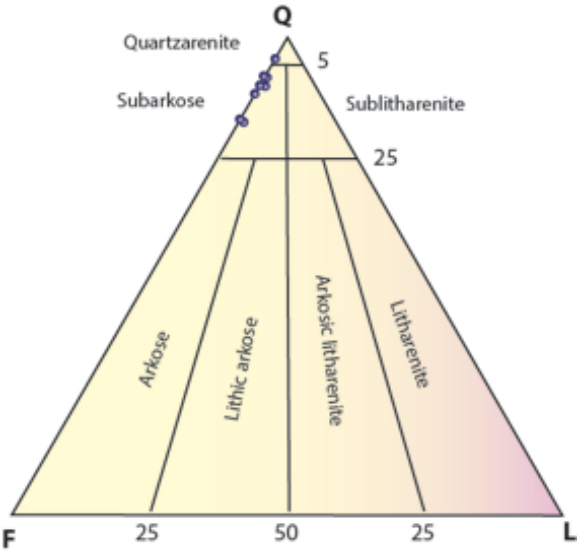


Figure 1.9: Ternary diagram showing subarkosic to quartzarenitic composition of Ogallala sandstone

and forms a resistant layer that forms pronounced ledges in the western outcrop area along the Palo Duro Creek (Fig. 1.6C). The thickly bedded chert breccia is light grey to very pale brown (10YR 7/1 to 10 YR 7/3). Clasts in the breccia are angular, are of pebble to cobble size, and are cemented by laminar to botryoidal microcrystalline silica (Fig. 1.8b). Abundant joints were observed in the chert breccia, and the fractures are curvilinear in profile (Fig. 1.7a).

Thick caliche overlies the chert caprock along Palo Duro Creek, and overlies the Ogallala Formation along Wolf Creek (Fig. 1.7b). Exposures in the study area reveal the characteristics of the basal 8 to 15 m of the caliche, and drilling records in the area indicate that the caliche section is in places thicker than 60 m. The caliche is pale yellow to reddish yellow (2.5YR 8/2 to 7.5YR 8/6), thickly bedded, and micritic. The caliche section along the Wolf Creek is intensely cracked (Fig. 1.6D). Chalky and friable caliche containing some gypsum was observed along Palo Duro Creek; two discontinuous medium to thick beds of chert occur in the caliche section. Sedimentary structures include horizontal laminae and desiccation cracks. Biogenic structures include abundant root tubules. The modern soil profile is developed above the caliche and has a gradational basal contact (Fig. 1.6D). The soil is gray and loamy and appears to be a residuum derived from weathering of the underlying caliche. The soil supports a desert-like flora where streams dissect the caliche.

Shallow subsurface and surface structure and aquifer thickness Drilling records from 147 water wells in the Farnsworth area provide a basis for characterizing subsurface structure and aquifer thickness. Table 1.1 shows the criteria that were used to pick the top and base of the Ogallala Formation. Caliche is commonly noted in driller's records, although picking the base of the caprock succession is complicated by inconsistent terminology. Strata assigned

to the chert caprock is typically described by drillers as hard sand rock or hard rock. The Ogallala Formation is distinctive in that it is composed primarily of sandstone and forms a fining-upward succession with conglomeratic strata at the base. Most of the water wells were drilled through the Ogallala into the Permian redbeds. Permian rocks beneath the Ogallala Formation are described as red, red clay or as redbeds, and drillers commonly note the relative hardness of these strata, which were instrumental for recognizing the base of the Ogallala Formation.

Based on the water well records, the thickness of the Ogallala Formation ranges from 30 to 260 m (Fig. 1.10). Changes in thickness are controlled primarily by variations in the depth and elevation of the pre-Ogallala disconformity surface; the Farnsworth Oil unit sits atop a paleotopographic high separating incised Ogallala paleovalleys (Fig. 1.11). The siliceous caprock is distributed mainly along Palo Duro Creek in the northwestern part of the study area, and a small patch of chert caprock appears to be present in the western part of the Farnsworth Oil unit (Fig. 1.12). The thickness of the caliche caprock varies from 3 m to 80 m in the study area and is anomalously thick near the town of Farnsworth (Fig. 1.13). Within the oil unit, however, the caliche is about 30 m thick, which is more typical of the region. Changes in thickness of the caliche caprock generally correspond with the structure of the top of the Ogallala Formation (Fig. 1.14).

Interval	Depth (ft)	Description (interpretation)
Caprock and soil	1 - 2.5	soil
	2.5 - 3.5	Caliche (caliche caprock)
	3.5 - 50	Hard sand rock (siliceous caprock)
Ogallala Formation	50 - 65	Rock, sand, gravel (caliche nodules?)
	65 - 70	Hard sand rock
	70 - 80	Sand and gravel
	80 - 85	Hard sand rock
	85 - 95	Sand
	95 - 172	Sand and gravel
Permiam undifferentiated	172	Red clay

Table 1.1: Typical water well record describing geology in the Farnsworth Area.

[1] Well number: 03-39-601

[2] Driller: Buschman Drilling Company

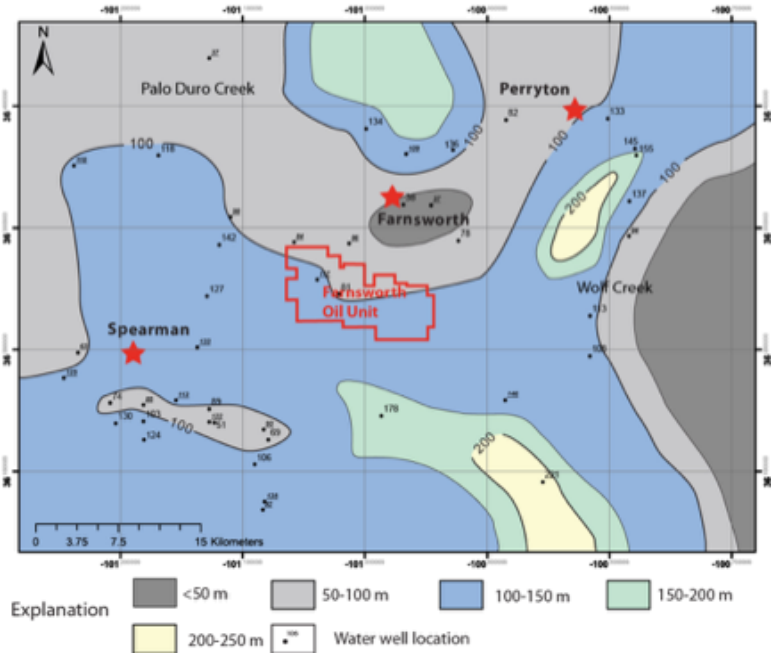


Figure 1.10: Isopach map of the Ogallala Formation. Contour interval = 50 m.

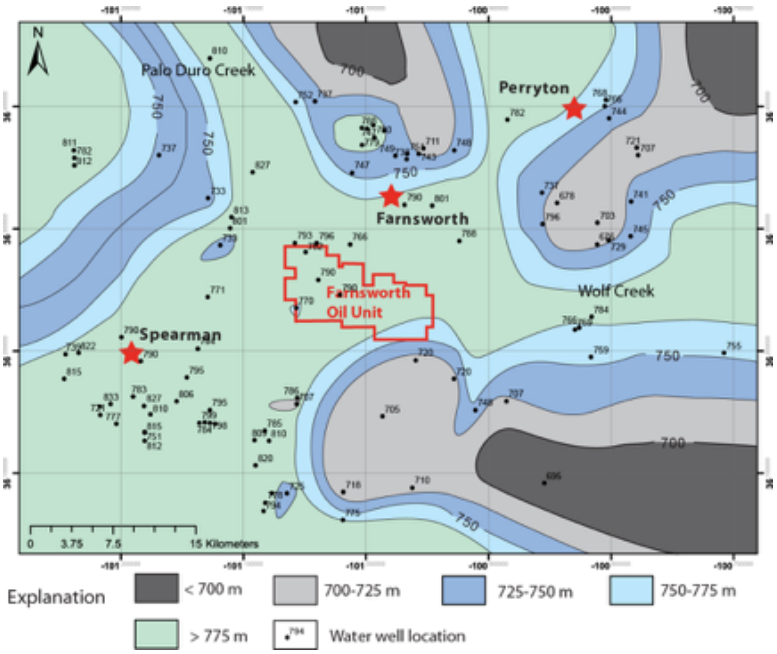


Figure 1.11: Structural contour map of the base of the Ogallala Formation. Contour interval = 25 m. Structural relief reflects paleotopography at the Permian-Miocene disconformity surface.

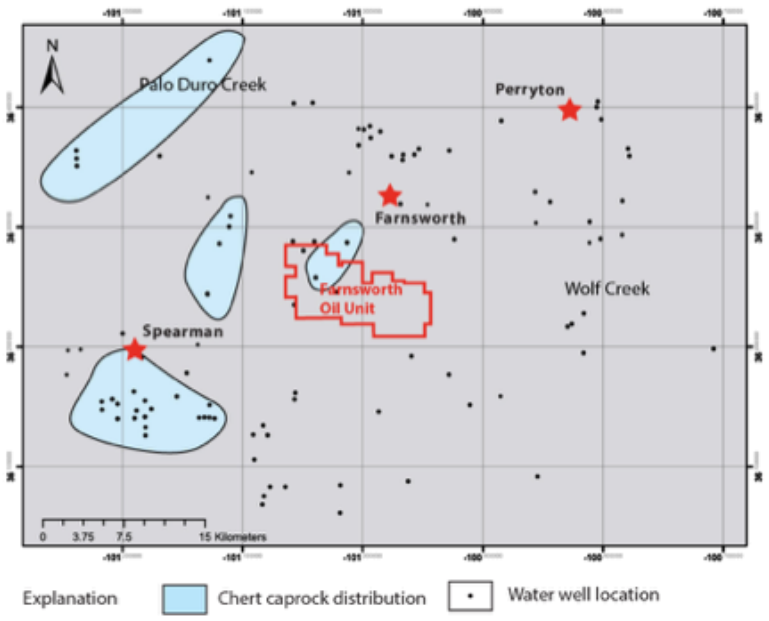


Figure 1.12: Chert distribution showing the siliceous caprock in the Farnsworth Oil Unit.

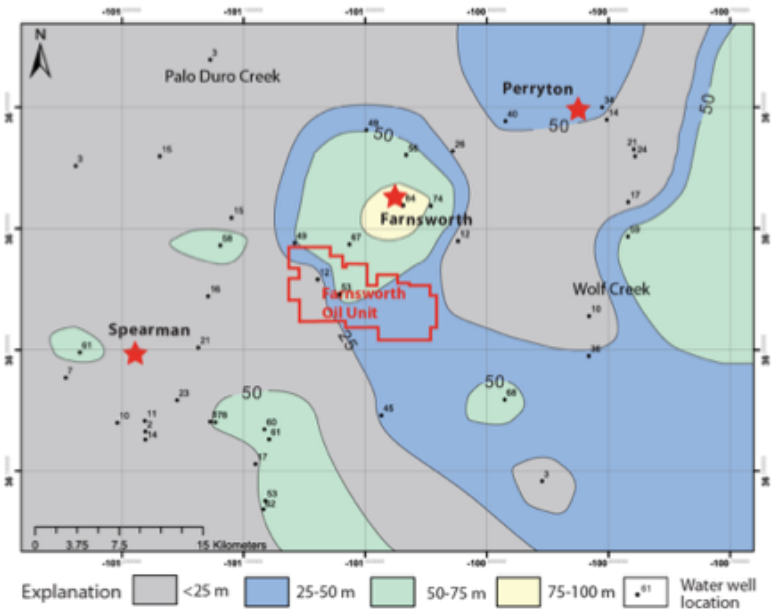


Figure 1.13: Isopach map of the caliche caprock. Contour interval = 25 m.

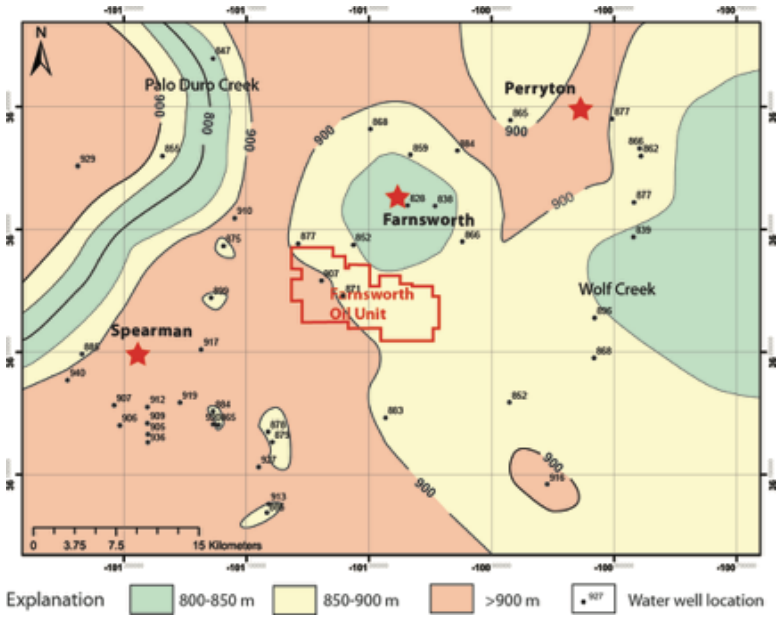


Figure 1.14: Structural contour map of the top of the Ogallala Formation. Contour interval = 50 m.

Topography is characterized regionally by a gentle slope of about 0.06 toward the northeast containing relief of about 70 m (Fig. 1.15). This area of gentle slope defines the elevated plain containing the Farnsworth Oil Unit. The plain is deeply dissected northwest of the oil unit by Palo Duro Creek and east of the oil unit by a dendritic stream system containing Wolf Creek. The plain is held up by the caprock succession, whereas the Ogallala Formation is exposed in the dissected terrain.

Based on the topographic map and field observations, there are no major faults or folds in the study area. However, some localized folds and rotational slumps were observed along the edges of some valleys in the study area. Figure 1.16, for example, shows a dipping chert layer overlain by Quaternary fill along Palo Duro Creek. Nearby, the chert caprock is gently folded along the valley margin, with strata dipping gently toward the axis of the valley. Similar structures were also observed locally along Wolf Creek. These structures appear to have formed by localized slumping and flow of poorly consolidated Ogallala sand along the steep slopes of modern stream valleys. No evidence for similar structures was found in the subsurface below the plain containing the Farnsworth Oil Unit.

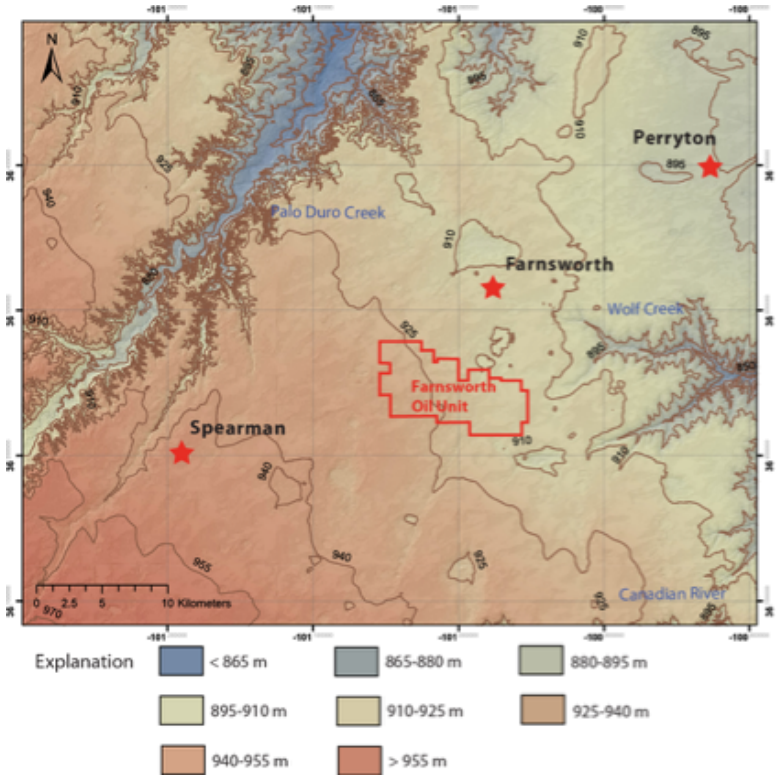


Figure 1.15: Hill-shade map showing topography of the Farnsworth area. Contour interval = 15 m.

Figure 1.16: Dipping chert layer along Palo Duro Creek.

1.5 Fracture Analysis

Outcrops and Landsat Thematic Mapper (TM) images in Google Earth reveal that numerous joints crop out in the field area and form significant fracture pavements on bedding plane exposures in the Ogallala Formation and in the chert caprock (Fig. 1.17). In vertical section, the joints are typically curvilinear and effectively strata-bound, cutting well-cemented, indurated sandstone and caprock, terminating either at bedding contacts or within the adjacent friable sandstone (Fig. 1.8a). On bedding planes, the exposed joints are weathered and host vegetation. Indeed, it is the vegetation that has colonized the fractures rather than the fractures themselves that define the joint pathways in satellite imagery (Fig. 1.17a).

The satellite imagery reveals that fracture networks consist of well-developed systematic joints and cross joints (Fig. 1.17b). Systematic joints are strongly aligned and have average length that typically exceeds 60 m. Average length of the joints is hard to calculate due to the incomplete exposure of the long fractures. Systematic joint length typically exceeds the dimensions of the exposures. Cross-joints tend to terminate at systematic joints, and so cross-joint length is effectively equal to systematic joint spacing.

Data collected from 21 outcrop bedding plane exposures using Google Earth resulted in more than 1,700 azimuth measurements. Two main joint systems are apparent in rose diagrams (Fig. 1.18). In the Wolf Creek area, two suborthogonal joint networks are well exposed in the Ogallala Formation. Joint System 1 is the dominant (i.e., most abundant) joint set striking southeast with a vector mean azimuth of 115 and a subordinate joint set striking northeast with a vector mean azimuth of 31. System 2 joints, by contrast, contain a dominant set striking northeast with a vector mean azimuth of 67 and a subordinate set striking southeast at 146. Cross-cutting relationships among outcrops are highly inconsistent, with systematic joints in one system belonging to the dominant joint set in one outcrop and to the subordinate set in another. In addition to distinctive regional joint systems, non-systematic joints that follow topography and are thus probably related to hillslope stresses also are developed in places.

Northwest of the Farnsworth Oil Unit along Palo Duro Creek, extensive fracture pavements are developed atop the siliceous caprock and are dominated by System 2 joints. System 2 in this area contains a well-developed systematic joint set striking northeast accompanied by SE-striking cross joints abutting the systematic joints. The vector mean azimuth of the systematic joints in System 2 is 67, and the cross joints strike with a vector mean azimuth of 146; these azimuths are identical to System 2 in the Wolf Creek area.

Spacing data from 1,288 joints were collected using the measurement tool in Google Earth. Cumulative probability plots of spacing data indicate that joint spacing in the study area is highly organized and follows log-normal spacing rules (Fig. 1.19). Mean, log-normal mean (geometric mean), and the limits of the log-normal standard deviation of joint spacing are shown in Table 1.2. For the Ogallala Formation sandstone along Wolf Creek, the general

(a)

(b)

Figure 1.17: Exposures of joint networks in the field area. a). Outcrop image showing bedding plane exposure of joints marked by vegetation in the Ogallala Formation along Wolf Creek. b) Satellite image showing well-developed joints in the siliceous caprock along Palo Duro Creek.

spacing of the System 1 joints are 2.0 m of southeast mode and 1.9 m of northeast mode. Joint System 2 contains the general spacing of 1.7 m at northeast mode and 1.8 m at southeast mode. For the chert caprock along the Palo Duro Creek area, the general spacing of the System 2 joints is 3.3 m in the northeast mode and 3.4 m in the southeast mode.

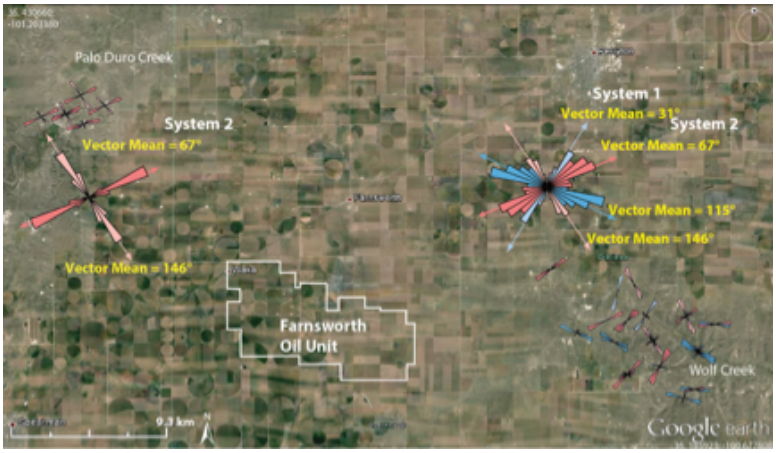


Figure 1.18: Rose diagram showing join systems and joint orientation in the study area.

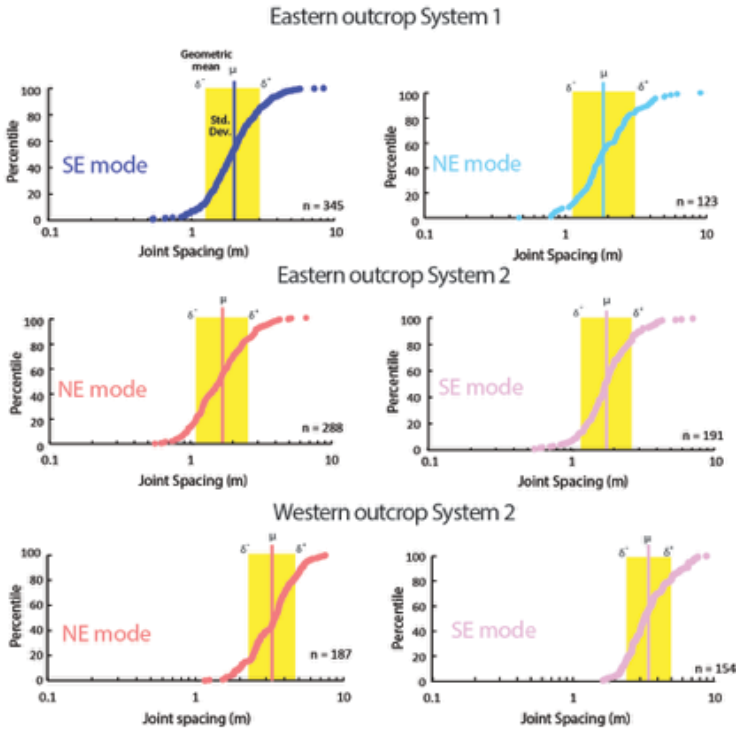


Figure 1.19: Percentile plots showing the log-normal distribution of joint spacing in the study area.

Lithology	System	Strike	Number of Readings	Distribution	Mean of joint spacing (m)	Log-normal Mean (m)	Limits of the standard deviation (m)
	System 1	115°	345	Log-normal	2.2	2.0	(1.3,3.0)
Ogallala	System 1	31°	123	Log-normal	2.2	1.9	(1.2,3.1)
Formation	System 2	67°	288	Log-normal	1.8	1.7	(1.1,2.6)
	System 2	146°	191	Log-normal	2.0	1.8	(1.2,2.7)
Chert	System 2	67°	187	Log-normal	3.3	3.3	(2.3,4.8)
Caprock	System 2	146°	154	Log-normal	3.4	3.4	(3.4,5.0)

Table 1.2: Analysis of joint properties and statistical distributions.

1.6 Discussion

1.6.1 Origin of the Fractures

Cross-cutting relationships of the two joint systems among outcrops are highly inconsistent along the Wolf Creek. No bedding plane exposures are found in the Ogallala Formation along Palo Duro Creek. Accordingly, it is unclear whether the sandstone beds host joints of both systems as is the case in the eastern outcrop area. However, the absence of System 1 joints in the chert caprock suggests that System 1 joints may have formed earlier than System 2 joints. The inconclusive abutting relationships in the Ogallalla Formation suggest that aquifer strata have experienced a complex breaking history and that the poorly consolidated sandstone was a poor propagator of stress in comparison to the brittle chert caprock.

Orientation and spacing analysis indicate that joint networks in the study area are well organized. Also, the jagged shape of the probability curves reveals that the dataset from the Palo Duro Creek area is not as well organized as that from the Wolf Creek area. Different level of organization and average spacing in both outcrops may related to the geomechanics of different caprock types. The systematic joints indicate the orientation of regional tectonic stresses. Those joints may be the tensile fractures that strike parallel to tectonic stress. This interpretation appears to be consistent with the active stress regime according to the World Stress Map (Heidbach et al., 2008). The World Stress Map (Fig. 1.20) reveals that joint orientation in the Farnsworth area is consistent with the modern lithospheric stress in the western Texas Panhandle as determined by hydraulic fracturing. The two systematic joint sets appear to reflect contemporaneous stress fields and thus may form a conjugate pair related to subtle shear stresses about a maximum horizontal stress. The probable orientation of the maximum horizontal stress is the bisectrix of the dominant fracture directions, or N. 80 E (Fig. 1.20). The subordinate joint sets may be related to local stress release.

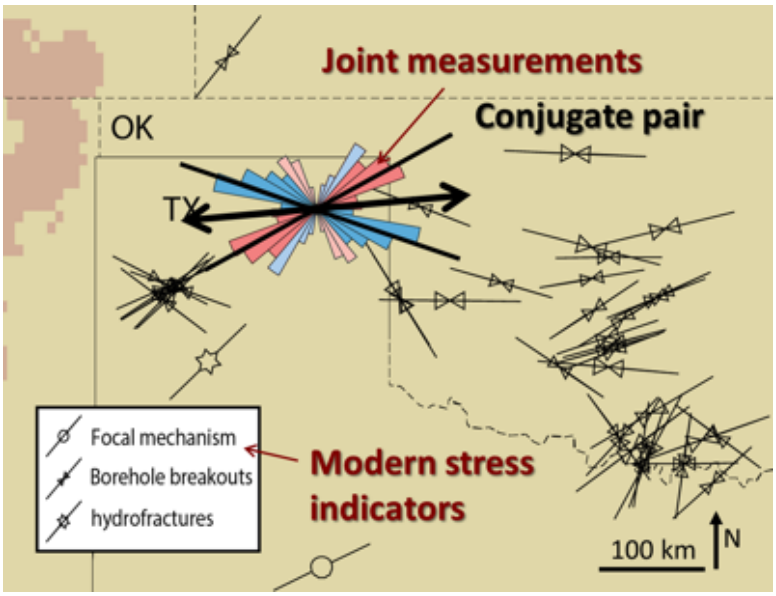


Figure 1.20: Stress map showing the present-day stress in the Texas Panhandle area. The joint systems (black, solid line) represent the product of shear stresses about the maximum horizontal stress showing by arrows (modified from Heidbach et al., 2008).

1.6.2 Fluid Movement through Fractures

The High Plains Aquifer in the Farnsworth Oil Unit is composed of sand and sandstone with various fluid transmissivity. The indurated sandstone lithofacies contains more cement than the less resistant massive and argillaceous sandstone lithofacies, and so limits fluid flux in the aquifer. Above the aquifer is a thick section of caliche and chert that would further buffer any seepage of CO_2 and CH_4. Numerous joints, including well-developed systematic joints and cross joints, were observed in the indurated sandstone lithofacies of the Ogallala Formation and in the siliceous caprock. Vegetation along the surface joint traces demonstrates that the fractures are important conduits for water and may therefore also control the exchange of gas between the aquifer and the atmosphere. Joint aperture highly impacts fluid transmissivity. According to numerous studies (e.g., Ortega, R.A. Marrett, and S. Laubach, 2006. Guerriero et al., 2010,Hooker, S. E. Laubach, and R. Marrett, 2013), joint aperture typically has a power-law population distribution. This distribution suggests that gases like CO_2 and CH_4 will not leak equally through all joints. In fact, it is typical for the vast majority of flow to occur along a small percentage of fractures in jointed formations (e.g., Nelson, 2001, Pashin et al., 2004). Where gas flows through a fracture of uniform aperture, it will ideally form a line-source plume (Fig. 1.21).

Alternatively, where joint aperture is nonuniform, point-source plumes may emanate from segments of joints where aperture is increased. Therefore, even though numerous fractures occur in near-surface formations in the Farnsworth area, few fractures may form significant

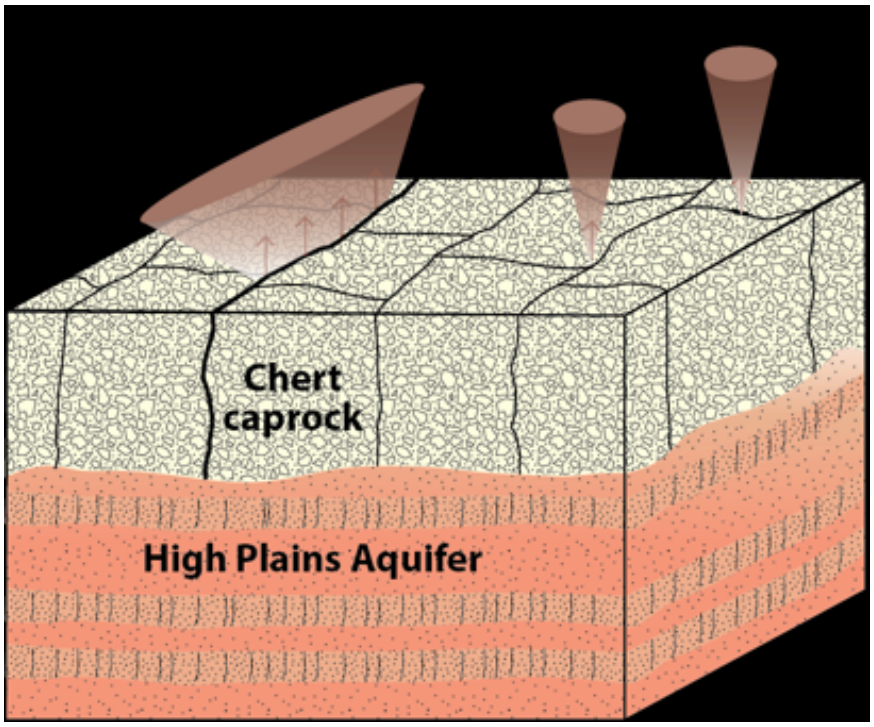


Figure 1.21: Conceptual model of CO_2 movement through fractures.

CO_2 and CH_4 migration pathways. Thus, if gas is moving through joints in the Ogallala Formation and the chert caprock, only few of fractures will support flow, and of those only parts of fractures may have significant aperture, perhaps either along localized joint segments or at joint intersections. In addition, the effects of fractures may be obscured by diffusion as gas migrates upward through the caliche section and soil profile. These hypotheses will be tested when surface monitoring equipment and UAVs are deployed at the Farnsworth Oil Unit.

1.6.3 Sources of CO_2 and CH_4

In the Farnsworth Oil Unit, geologic risk associated with CO_2 injection in the upper Morrow appears minimal because numerous sealing strata exist in the subsurface. Moreover, the Permian evaporites form a widespread seal that protects the High Plains Aquifer from any potential leakage associated with CO_2-enhanced oil recovery. Accordingly, the greatest risks are associated with infrastructure, and even this risk is low because the surface and production casing strings intervene between the wellbore and the USDW of the High Plains Aquifer (i.e., the Ogallala Formation).

Because the risk of reservoir leakage to the surface is very low in the Farnsworth Oil Unit, the CO_2 and CH_4 that the surface and airborne monitors detect may largely be the product of near-surface processes, such as microbial processes and reactions between carbonate and water

(e.g., Klusman, 2005; McIntyre et al., 2008) (Fig. 1.22). CO_2 and CH_4 can be generated in soil and aquifers via microbial acetate fermentation. CH_4 also can be generated via microbial CO_2 reduction, which is a process that can consume fugitive CO_2. Another possibility is that CO_2 may be generated by calcite dissolution driven by meteoric recharge and groundwater movement, which may obscure the effects of natural fracturing in the shallow subsurface. Dissolved CO_2 in water also may generate carbonic acid, which will in turn be buffered by reaction with carbonate in the Ogallala Formation and the caliche section. Dissolution processes appear to be operating today, as evidenced by development of residual loam atop the caliche section (Fig. 1.6 D).

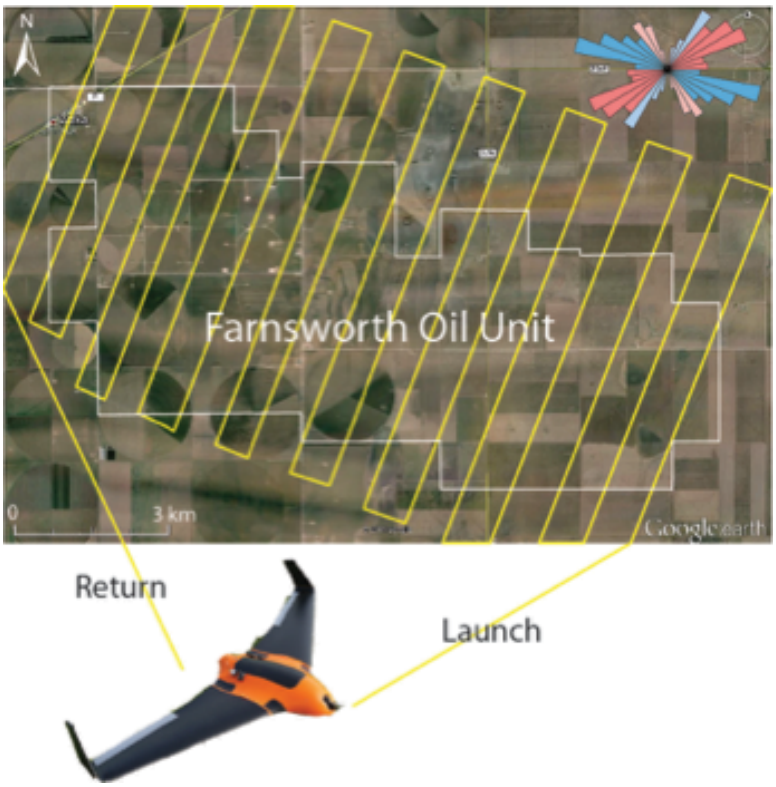


Figure 1.22: Idealized aerial sampling pattern for terrains where orthogonal joint networks may influence the flux of gas into the atmosphere.

1.6.4 Design of UAV Flight Pattern

Based on the results of this research, surface and airborne monitoring programs should take fracture architecture into consideration to maximize the probability of detection of major CO_2 and CH_4 fluxes emanating from the land surface. Flight paths should ideally be oblique to systematic and cross-fracture orientations to increase the likelihood of fracture-related plume detection (Fig. 1.23). Future research will focus on field deployment of surface and

UAV monitoring technology. During deployment the design UAV flight paths should not only consider the geologic architecture but also numerous other variables, such as, wind direction, wind velocity, flight altitude, and flight duration. After determining the actual CO_2 and CH_4 flux patterns in the Farnsworth Oil Unit, the results of surface and airborne monitoring will be used to adjust the geologic models, optimize surface sensor placement, and optimize UAV flight patterns.

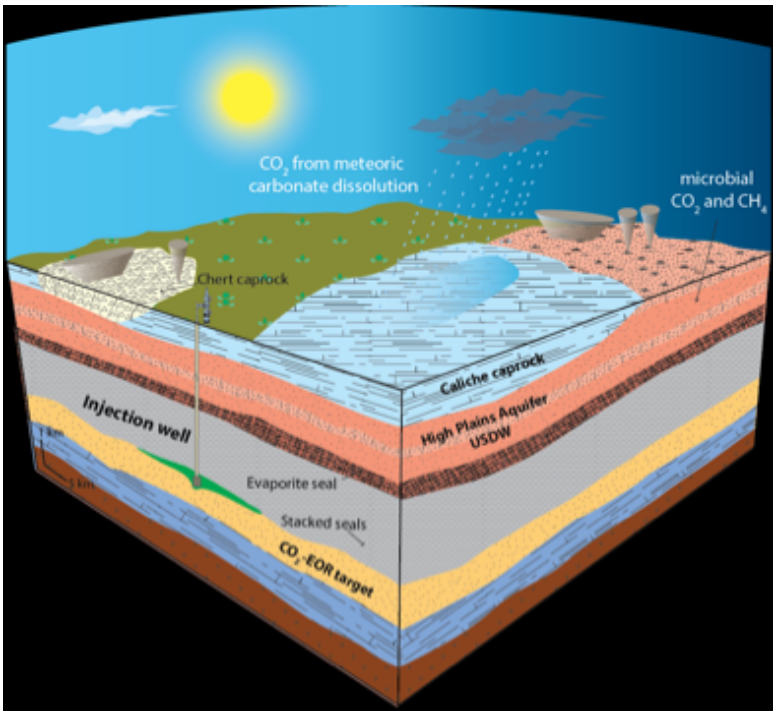


Figure 1.23: Conceptual model showing CO_2 and CH_4 exchange among the aquifer, caprock, and atmosphere.

1.7 Conclusions

Aquifer protection is a central imperative of underground injection control in the United States. Surface and airborne gas monitoring programs are becoming an important part of environmental protection in areas favorable for subsurface storage of carbon dioxide. Understanding structural architecture and its effects on the flux of fluids, specifically CO_2 and CH_4, in the shallow subsurface and atmosphere is helping design and implement next generation monitoring technologies, including unmanned aerial vehicles (UAVs). An important aspect of this research is using subsurface fracture data to inform the design of flight pathways for UAVs in the Farnsworth Oil Unit of the Anadarko Basin.

The High Plains Aquifer of the northeastern Texas Panhandle includes sandstone of the Ogallala Formation and Quaternary strata that include sand, clay, chert, and caliche. The High Plains Aquifer disconformably overlies Permian and Triassic redbeds and is dominated by weakly to moderately indurated sandstone in the Miocene-Pliocene Ogallala Formation. Three lithofacies were identified in the Ogallala Formation, including the massive sandstone lithofacies, the argillaceous sandstone lithofacies, and the indurated sandstone lithofacies. Ogallala strata are overlain by Quaternary strata that include a siliceous caprock west of the Farnsworth Oil unit and a thick (25 m) section of caliche east of the oil unit. The target zone for CO_2 storage and enhanced oil recovery in the Farnsworth Oil Unit is in upper Morrow sandstone of Lower Pennsylvanian age at subsurface depths greater than 2,000 m. Field study of near surface strata reveals that indurated sandstone and a chert caprock strata contain numerous joints that provide crucial insight into aquifer architecture and subsurface flow pathways. Length, orientation, spacing, and cross-cutting relationships of more than 1,700 joints were measured in the field and in high-resolution satellite imagery. The joint networks consist of well-developed systematic joints and cross-joints. Systematic joints are strongly aligned and are commonly longer than 60 m. Cross-joints tend to terminate at systematic joints and so joint length is typically equal to systematic joint spacing. The fractures are liner to curvilinear in plan view. In vertical section, the joints are typically curvilinear and strata-bound, cutting indurated strata and terminating within friable sandstone.

Analysis of joint indicates that that strike of the systematic joints varies among beds and regionally. Two distinctive joint systems were identified in the study area. The older system was observed in outcrops east of the oil unit and has vector mean azimuths of 115 and 31. The younger system was observed throughout the study area, with joints having vector mean azimuths of 67 and 146. Joint spacing follows a log-normal statistical scaling rule. In the west of the Farnsworth Oil unit, the general spacing of the older system joints are 2.0 m of SE mode and 1.9 m of NE mode. Joints of younger system contain the general spacing of 1.7 m at NE mode and 1.8 m at SE mode. In the west of the oil unit, the general spacing of the younger system joints is 3.3 m in the NE mode and 3.4 m in the SE mode. These fractures appear to be the product of regional tectonic stress and may have a significant effect on flow in the High Plains Aquifer system, as well as near-surface gas flux. Based on the results of this research, design of UAV flight paths should be oblique to fractures in a way that maximizes the likelihood of CO_2 and CH_4 flux from of systematic joints and cross joints. The risk of leakage from CO_2-enhanced oil recovery operations at Farnsworth is low, and multiple potential natural sources of CO_2 and CH_4 have been identified in near-surface formations. These near-surface sources are predicted to dominate shallow subsurface and atmospheric gas flux.

2 Land-Based Sensor System

2.1 Introduction

The primary goal of this project was to evaluate components, design the instrument platform, test and deploy a monitoring system based on low cost carbon dioxide (CO_2) and methane (CH_4) sensors. Sensor technology is developing rapidly driven by the demand for gas sensors for home and commercial use. Increasing demand has brought the costs down and the accuracy and reliability of these gas sensors up. At the time that the sensors were being evaluated, the technology for CO_2 sensors was somewhat better than that for CH_4. With this in mind, the field units were designed to accept new sensors if better sensors became available. Data were collected from both fixed, ground-based units and a UAV equipped with both CO_2 and CH_4 sensors. In this section, sensor selection, sensor platform design, testing, and construction, field testing, deployment at the Farnsworth site, and data analysis will be discussed.

2.1.1 Sensor Selection

A technological niche exists for a relatively inexpensive device which can monitor a large number of sites easily with little human interaction. Evaluating and selecting a sensor for each gas was the first step in the process. The second step in the process involved adapting available technology to develop a distributed sensor system that was capable of collecting, storing, and transmitting data. In the interest of cost minimization, most if not all of the components needed to be commercially available.

2.1.2 Design and Construction

To aid in the development of the sensor platform an Arduino Development Prototype and Deployment system was used to design, develop, and deploy prototypes. The Arduino acted as an interim control board to test the reliability of the setup in the field. This provided an inexpensive way to iterate hardware and software. The final design that resulted from this process is described in this report.

2.1.3 Field Evaluation

Testing of the sensor nodes was performed at a field site on the Oklahoma State University (OSU) campus. The goal was to collect baseline data, identify possible issues that might be encountered during field deployment, and develop the network technology. All sensor nodes were first field tested at OSU field site before being deployed. One set of fifteen sensors was kept at the site from the time that the sensor nodes became available in October of 2015 through the end of the project to test the longterm performance of the system. The results are presented later in this chapter.

2.1.4 Deployment

There were two field sites in which the sensor arrays were deployed. One site was at the Unmanned Airport Flight Station (UAFS) at OSU and the other at the Farnsworth, TX field site. At the UAFS, six communication nodes for the array and 28 sensor nodes were installed around the perimeter. Flight controllers at the airfield were concerned that UAVs might collide with the sensor nodes during takeoff and landing operations if the nodes were placed in the area immediately in front of the runway. The Farnsworth, TX sites was more challenging in terms of logistics and placement. The monitoring area that was selected was heavily farmed. In this environment, the Southwest Partnership regularly loses approximately thirty percent of the collars that they use in conjunction with their Li-Cor CO_2 instrument. It turns out that almost anything that will not damage farming equipment gets run over. With this in mind, we obtained permission to mount the sensor nodes on power poles that line the roads in the area. At the Southwest Regional Partnership, special hardware was designed to allow mounting without damaging the utility poles. In total, three communication nodes and 34 sensor nodes were installed.

2.1.5 Farnsworth site summary

The Farnsworth, TX site was maintained for more than seven months. An ice storm in February damaged a number of the utility poles that the sensors were mounted on, and the sensors had to be removed earlier than planned.

2.2 Sensor Selection

2.2.1 Considerations

When designing monitoring devices, the sensor for gas concentration analysis must be matched to specific operational requirements, such as precision, reliability, and power consumption. Example sensor elements include those used in HVAC air handlers (Yang et al., 2014; Chung

and S.-C. Lee, 2008), chemical processing units (Won and K. S. Lee, 2012), oil well monitoring devices (Yi, Xiao, and Zhang, 2010; Somov et al., 2013), and environmental monitoring (Pering et al., 2014; Black et al., 2012; Guohua et al., 2012; Karunanithi et al., 2009; Shendell et al., 2012). Potential sensors ranged from small, inexpensive chemiresistive sensors to complex and more costly optical systems. The low-cost chemiresistive based methane sensors are typically used in gas warning systems (Chiu and Tang, 2013). On the other end of the sensor spectrum, Light Imaging Detection and Ranging (LIDAR) is an accurate and effective method for remote monitoring of industrial sites, for example oil wells (Ho et al., 2001). However, these devices are limited by cost and operational complexity, and are not suitable for portable low-power monitoring devices. Sensors need good sensitivity and precision around the baseline atmospheric concentration for each analyte, which is around 400 ppm for carbon dioxide (Blasing, 2016; Dlugokencky and Tans, 2016). For methane, the baseline atmospheric concentration is under 2 ppm (Turner et al., 2016; Bamberger et al., 2014; Dlugokencky, 2016). And thus, four methane, sensors were selected and tested near their baseline. In addition to concentration, power consumption, reliability, and ease of integration are also important factors for remote operation. Sensors were selected based on their sensitivity at atmospheric levels, ease of use, power consumption, price, and market availability. Previous researchers have cited concerns with electrochemical sensors for these gases, as they have a short lifetimes and lack robustness (Neri, 2015). Therefore, the selection process was focused on sensors that were commercially available in large volumes (at least 1000 units) at low-cost (defined here as less than $100 per unit in bulk).

Commercially available carbon dioxide and methane sensors were selected and evaluated based on the needs for a portable low-power monitoring device. Table 2.1 lists the selected carbon dioxide sensors with the important properties obtained from the manufacturer. Table 2.2 lists the methane or hydrocarbon sensors and their respective properties. The K-30, COZIR, Dynament, and Telaire sensors are all NDIR sensors. These sensors were chosen as low-cost, lightweight sensors with satisfactory detection parameters for carbon dioxide. Dynament also provides a dual gas NDIR sensor (MSH-DP/HC/CO2/) designed to measure both carbon dioxide and methane concentrations. This ability was attractive given low-cost and portability requirements. The carbon dioxide and methane Gascard sensors sold by GHG Analytical were an order of magnitude more expensive than the other NDIR sensors, which have a cost between that of the lowest cost sensors on our list and that of the bench-top analyzers. Their specifications combined with the included pressure and temperatures compensation make them attractive enough to make up for the expense. In addition to the Gascard sensor, the Dynament hydrocarbon sensors (MSH-P/HC and MSH-DP/HC/CO2/) were chosen as inexpensive candidates for methane detection. Chemoresistive sensors include the MQ-4 from Hanwei Electronics and TGS-2600, TGS-2610, and TGS-2611 manufactured by Figaro Engineering Inc. sensors. The TGS sensors are used in commercial methane detectors, and the TGS-2600 sensor has been previously evaluated for atmospheric applications (Eugster and Kling, 2012). There are several different MQ versions optimized for hydrocarbon sensing.

Table 2.1: Manufacturer listed properties of evaluated carbon dioxide sensors.

Sensor	Supplier	Type	Sampling Method	Cal. Range	Op. Range
K-30 SE-0018	CO_2Meter	NDIR	flow or diffusion	0-5000 ppm	0-10000 ppm
COZIR AMB GC-020	CO_2Meter	NDIR	flow or diffusion	0-5000 ppm	0-10000 ppm
Gascard CO_2	GHG Analytical	NDIR	flow	0-50000 ppm	0-50000 ppm
MSH-P/CO2/NC/5/V/P/F	Dynament	NDIR	diffusion	0-2491 ppm	0-5000 ppm
MSH-DP/HC/CO2/NC/P/F	Dynament	NDIR	diffusion	100-2500 ppm	0-5000 ppm
Telaire T6615	General Electric	NDIR	flow or diffusion	0-2000 ppm	0-2000 ppm

Sensor	Warm Up	T	Humidity	Auto-cal	V Input	Avg. I
K-30 SE-0018	<1 min	0-50°C	0-95%	Yes	4.5-14 VDC	40 mA
COZIR AMB GC-020	<3 s	0-50°C	0-95%	Yes	3.25-5.5 VDC	1.5 mA
Gascard CO_2	30 s	0-45°C	0-95%	Yes	7-30 VDC	250 mA
MSH-P/CO2/NC/5/V/P/F	45 s	-20-50°C	0-95%	No	3.0-5.0 VDC	75-85 mA
MSH-DP/HC/CO2/NC/P/F	45 s	-20-50°C	0-95%	No	3.0-5.0 VDC	75-85 mA
Telaire T6615	10 min	0-50°C	0-95%	Yes	0-5 VDC	33 mA

Table 2.2: Manufacturer listed properties of evaluated methane sensors.

Sensor	Supplier	Type	Sampling Method	Cal. Range	Op. Range
MQ-4	Futurelec	Chemoresistive	diffusion		200-10000 ppm
Gascard CH_4	GHG Analytical	NDIR	flow	0-50000 ppm	0-50000 ppm
MSH-P/HC/NC/5/V/P/F	Dynament	NDIR	diffusion	0-5000 ppm	0-10000 ppm
MSH-DP/HC/CO2/NC/P/F	Dynament	NDIR	diffusion	5000-11000 ppm	0-10000 ppm
TGS-2600	Figaro Engineering	Chemoresistive	diffusion		1-30 ppm
TGS-2610	Figaro Engineering	Chemoresistive	diffusion		1000-25000 ppm
TGS-2611	Figaro Engineering	Chemoresistive	diffusion		500-10000 ppm

Sensor	Warm Up	T	Humidity	Auto-cal	V Input	Avg. I
MQ-4				No	5 VDC	<150 mA
Gascard CH_4	30 s	0-45°C	0-95%	Yes	7-30 VDC	250 mA
MSH-P/HC/NC/5/V/P/F	30 s	-20-50°C	0-95%	No	3.0-5.0 VDC	75-85 mA
MSH-DP/HC/CO2/NC/P/F	30 s	-20-50°C	0-95%	No	3.0-5.0 VDC	75-85 mA
TGS-2600				No	5.0±0.2 VDC	4.2±4 mA
TGS-2610				No	5.0±0.2 VDC	5.6±5 mA
TGS-2611				No	5.0±0.2 VDC	5.6±5 mA

Sensors with no listed warm up time required 7 day burn in time

The MQ-4 sensor was chosen as this variant was specifically tuned for methane.

2.2.2 Sensor performance and data analysis

Testing methods

To test the sensors under controlled conditions, a gas mixing apparatus (Fig. 2.1) was constructed. This apparatus allows gas flows of a known concentrations to be prepared from a gas cylinders containing known gas concentration. A high-quality bench-top analyzer (California Analytical Instruments, Inc. ZRE Non-Dispersive Infrared Analyzer) sensitive to both carbon dioxide and methane was used to provide baseline analyses. The analyzer was periodically calibrated using the calibrated gas mixture to ensure accuracy. Gas cylinders containing specific mixtures of gases depending on the experiment being performed were used. These gas mixtures were provided by and certified to these concentrations within ±2% of the declared value by Airgas Inc. Calibrated gas mixture were diluted using either air or nitrogen gas by a set of flow controls to produce specific partial pressures of the analyte gases. In the case of sensors configured for gas flow, the sensor being tested was connected directly to the apparatus by flexible hose connection. For diffusion based sensors, the sensor was placed in an exposure chamber. All experiments were referenced to the California Analytical Instruments ZRE Analyzer.

Data Collection and Sensor Response Time

Data were analyzed by first performing a baseline correction, if required, at each carbon dioxide and methane concentration. Both the mean (μ) and standard deviation (σ) were calculated from each individual baseline and average at each calibration point. The $\mu_{ZRE,peak}$ was the average response of the California Analytical Instruments Inc. ZRE. A plot of the calibration is shown in Fig. 2.2

$$x = \mu_{ZRE,peak}$$
$$y = \mu_{sensor,peak} - \mu_{sensor,baseline}$$

Figure 2.2 shows the response of two selected carbon dioxide sensors, the K30 and Gascard, to 600 and 1400 concentration steps of carbon dioxide. Both optical sensors respond quickly with no overshoot. These steps are just above and below the OSHA standard of 1000 ppm. The K30 reports the concentration directly, while the Gascard reports a value between zero and one, with one being the maximum concentration (3% carbon dioxide for the sensor). Both optical sensors respond very quickly to an increase in carbon dioxide inside the environmental chamber. The data demonstrate relative similarity in behavior of the selected sensors in the presence of concentration changes well above the limit of detection.

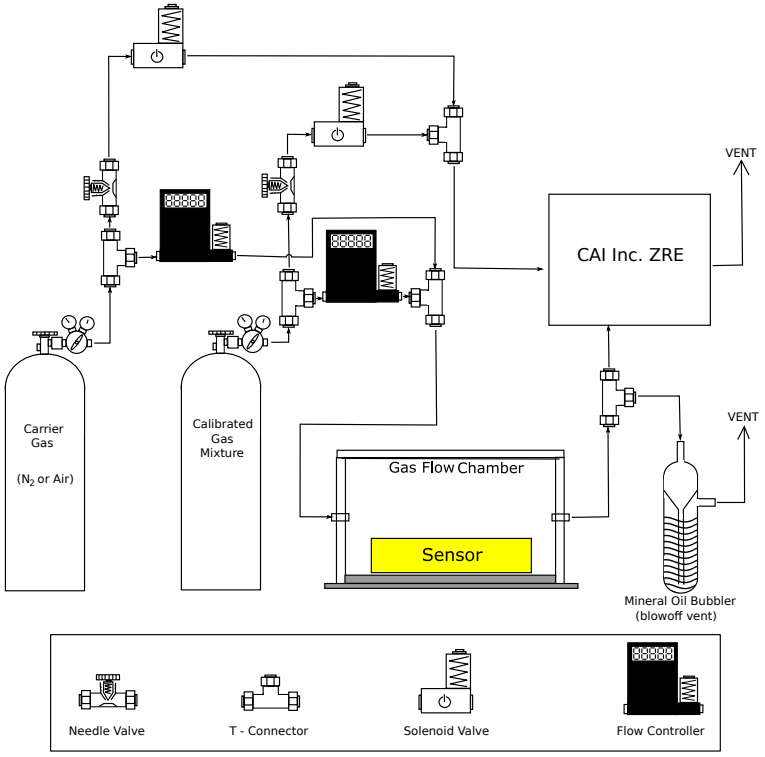


Figure 2.1: Component diagram of controlled gas exposure apparatus with chamber for diffusion-type sensors.

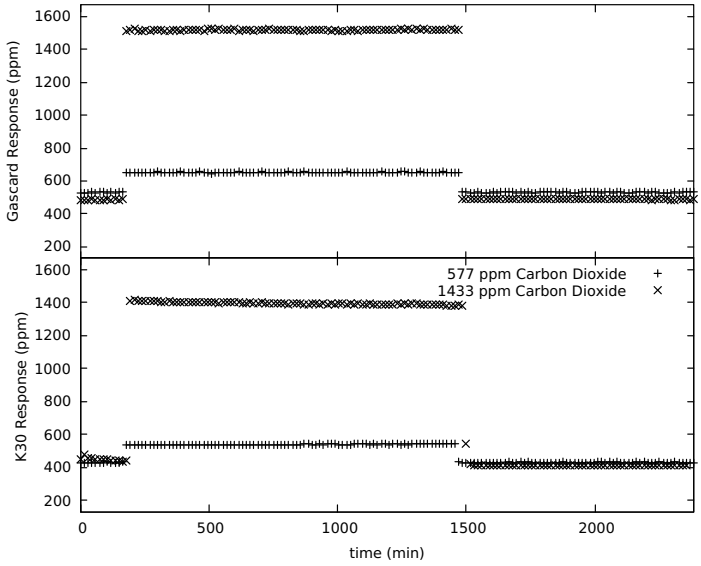


Figure 2.2: The Gascard (top) and K30 (bottom) sensor response over time at high (1433 ppm, plotted with "×" on the plot) and low (577 ppm, plotted with "+" on the plot) concentrations of carbon dioxide. Note the sharpness of the response by these sensors.

Figure 2.3 shows the response of two chemiresistive sensors optimized for methane detection, the MQ-4 and TGS-2611, to 100 and 1000 ppm concentration steps of methane. These experiments were performed with a small sensor enclosure (internal volume 2.54 cm^3). The higher limit is the maximum National Institute for Occupational Safety and Health's (NIOSH) recommended safe methane concentration. Conversely, the lower limit is more representative of a leak, which could be a significant distance from the source. The TGS-2611 sensor had a much faster and more step-like response to changes in methane concentration than the MQ-4 sensor. Similar behavior was also observed for a decrease in concentration. The magnitude of response is larger and noise is smaller for the TGS-2611 sensor. These issues are also seen when comparing the baseline measurements of both sensors, depicted in Fig. 2.4. Additionally, these plots show that the TGS-2611 sensor has less baseline drift than the MQ-4. In field applications, baseline changes due to humidity and temperature may play a significant role and must be taken into account.

Figure 2.5 depicts a plot of the temporal response of an MQ-4 sensor to five different concentrations of methane between 500 and 2200 ppm. The sensor produces significant overshoot, often exceeding 100% of the final response. Normally, sensors are characterized by rise time, the time for the response to change from 10% of the mean response to 90% of the mean response. These overshoots prevent this metric from being used. And thus, the settling time, the time required to reach steady state after a concentration change, is more useful. The settling time for the MQ-4 sensor did not appear to be concentration dependent. For all concentrations, a stable value within 2.5% of the mean was produced after 78±10 s, when averaged over all response time experiments. The overshoot may be a result of the complex set of chemical reactions that occur on the surface of the sensor taking a finite amount of time to reach steady state after a concentration change. Another cause may be a small discontinuity in the gas flow during a concentration change. Similar or faster response times, given Fig. 2.3, are expected for the TGS sensors.

Precision and Baseline Noise Tests

Measurement of the precision of each sensor was performed using the exposure apparatus discussed in Fig. 2.1. Baseline carbon dioxide tests were performed at a measured concentration of approximately 400 ppm in nitrogen gas, while the baseline methane tests were performed using medical grade air. The 0 ppm methane concentration was chosen as the baseline because the expected atmospheric or environmental level will be less than 1 ppm (Turner et al., 2016; Dlugokencky, 2016). Medical grade air was utilized since the chemiresistive sensors required atmospheric concentrations of oxygen to correctly measure the methane concentration. Gas concentrations were verified with the bench-top analyzer (California Analytical Instruments, Inc. ZRE Non-Dispersive Infrared Analyzer). The precision of the sensors was determined by a 20 to 30-hour data collection run at a known concentration and uniform flow rate.

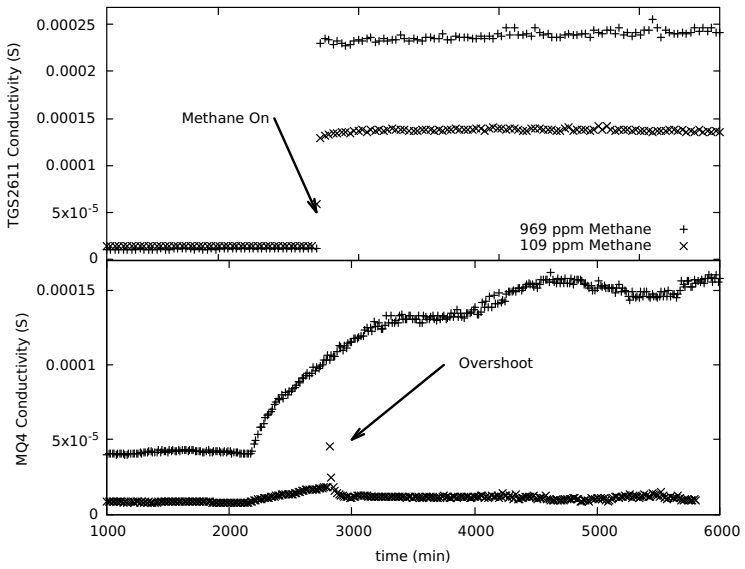


Figure 2.3: Plots of the TGS-2611 (top) and MQ-4 (bottom) sensor response over time at high (969 ppm, plotted with "×" on the plot) and low (100 ppm, plotted with "+" on the plot) concentrations of methane. Note the slower response of the MQ-4.

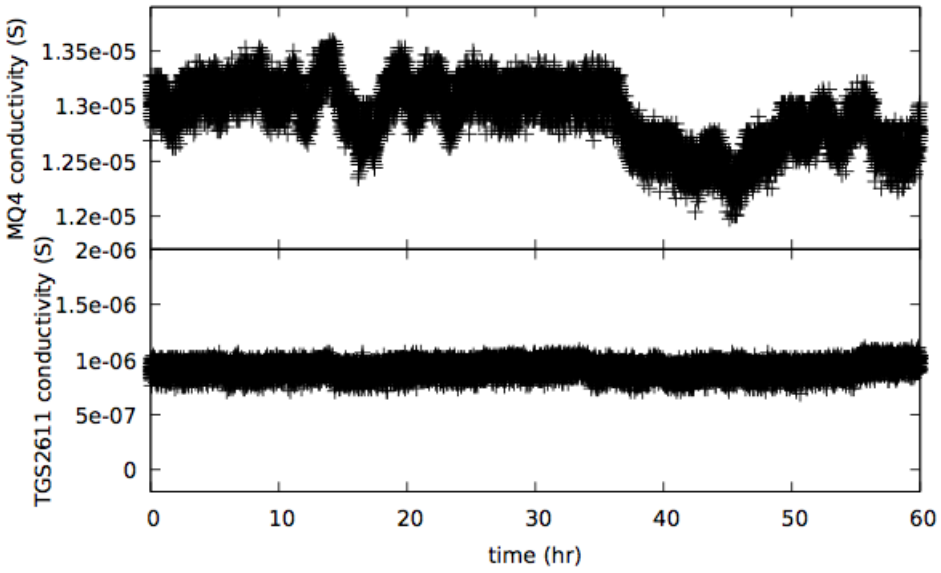


Figure 2.4: Long term baseline data were collected at the atmospheric baseline conditions (using a bottle of compressed medical grade breathing air (approximately 0 ppm methane), in the gas mixing chamber. The baseline fluctuations can be observed in these plots.

An initial Fourier analysis showed no significant periodic variations in the output during the tests, so a distribution of the digitized sensor output around the mean was utilized. Data streams were processed to provide the individual difference from the mean reported value, and

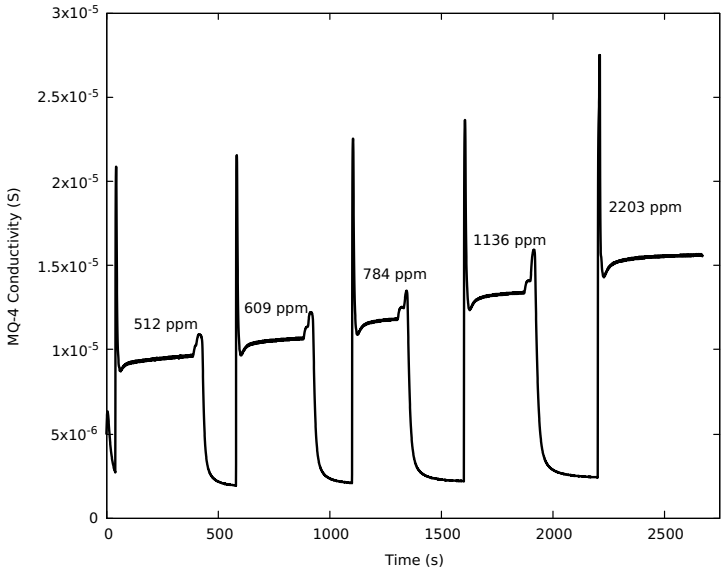


Figure 2.5: The MQ-4 sensor showed noticeable delay between introduction of gas and production of a stable response. At the introduction of gas for each concentration change, a significant overshoot is observed.

a histogram of these differences was plotted. Digitized sensor outputs have a finite number of possible output values, so no further bins were created while producing the analysis. Although the data utilized for this analysis were obtained at typical environmental concentrations, the measured deviation around the mean for each sensor was independent of the concentration of analyte gas. There was insignificant correlation between the concentration of the analyte gas and the σ obtained at each concentration, as quantified by the Pearson's Correlation Coefficient ($\rho = -0.173$). The results from this analysis can be seen in Table 2.3.

For the carbon dioxide sensors, the Gascard and K-30 sensors both showed low noise or smallest standard deviation around the mean, σ_{GAUSS}. Noise performance of the Telaire sensor was slightly worst by a factor of approximately 2 to 3. Variations for COZIR and the Dynament (MSH-P/CO2/) sensors were significantly greater that the Gascard, Telaire and K-30 sensor. Finally, the dual gas Dynament (MSH-P/HC/CO2/) sensor produced the largest deviation from the mean, possibly due to the dual sensing elements. Fluctuations around the mean is slightly lower than the quoted resolutions in the manufacturer's documentation of 50 and 100 ppm. The probability distribution of responses around the mean by the GE Telaire sensor is a signal peak with a σ_{GAUSS} that is approximately 2 to 3 times as large as that of the K-30 and Gascard sensors. The geometry of the GE Telaire similar to the K-30 sensor. They share comparable sensing mechanisms and path lengths. Results from the larger σ_{GAUSS} unit was unexpected. A direct comparison of the response for both the Telaire and K-30 sensors was performed simultaneously, both with and without ambient light. The data showed that the Telaire sensor is sensitive to ambient light, while the change in response from the K-30 sensor was negligible. This makes the Telaire sensor less desirable as these sensors

will be used outside.

For the optical methane sensors (see Table 2.3), the Gascard for methane produced a σ_{GAUSS} with a low RMSE (root-mean-square-error). The single gas Dynament (MSH-P/HC/) hydrocarbon sensor produced a very low σ_{GAUSS} and performed well in terms of precision. For this sensor, the 18 ppm precision is lower than the 50 ppm resolution quoted in the manufacturer's documentation at the low-end of the 1% concentration range of the sensor. With the inclusion of the calibration and limit of detection results, discussed below, the overall performance of these sensors are constant with the 50 ppm resolution. The σ_{GAUSS} for the dual-gas Dynament (MSH-DP/HC/CO2/) sensor was not included in this table. As mentioned above, the dual-gas Dynament (MSH-DP/HC/CO2/) sensor only reported two values for methane around the mean rather than a distribution of several values, which is not an adequate number of data points to provide a sufficient fit. This result is also consistent with the 100 ppm resolution quoted in the manufacturer's documentation for concentration of less than 10% methane, and is not indicative of a design issue.

The result for the chemiresistive methane sensors in Table 2.3 with a σ_{GAUSS} and a RMSE results in Table 2.3. The distribution curve for the MQ-4 sensor was not included. Since the MQ-4 sensor displayed significant baseline drift when compared to the TGS-2611 sensor (Fig. 2.4), the standard deviation was instead calculated directly from a relatively flat region of the baseline. This different treatment is not inconstant with the use of the sensor in many applications where drift is a result of temperature and humidity. Previous studies by Solis et al. have shown the noise of chemiresistive TGS-26xx sensors to be temperature dependent (Solis et al., 2005). In general, dynamic background subtraction is critical for the performance of the chemiresistive sensors. The σ_{GAUSS} results of the TGS-2600 and TGS-2610 were similar. This is expected due to their similar sensing mechanisms. Baseline noise, as quantified by σ_{GAUSS}, for the methane optimized TGS-2611 sensor was lower than the other TGS sensors. Although the background fluctuations are low for these sensors, the non-linear response and baseline drifts have significant effects on the limits of detections discussed below.

2.2.3 Sensor Calibration and Limit of Detection

The calibration curve of each sensor was established by varying concentrations of gas through system and, when appropriate, subtracting the baseline reading to the average concentration after stabilization. The carbon dioxide sensors were calibrated using points from 34.5 to 1020 ppm. For the methane sensors, calibration cures were generated from points between 1.85 and 995 ppm. In the development of a typical calibration curve, the carrier gas was first introduced. When a stable baseline was obtained, the calibrated concentration was introduced for 24 hours. After 24 hours, the calibration gas was removed and the carrier gas reintroduced for another 24 hours to allows the system to stabilize before the next measurement. During this procedure, the data from each sensor was continually collected. Using this method, the average baseline and response to each concentration was extracted. This process allows any

Table 2.3: Standard Deviation of Gaussian Fitted Probability Distributions and Root-Mean-Squared Error.

	Sensor	σ_{GAUSS} (ppm)	RMSE
Carbon Dioxide	K-30 SE-0018	1.91	0.219
	COZIR AMB GC-020	14.1	0.304
	Gascard CO_2	2.12	0.223
	MSH-DP/HC/CO2/	86.4	0.197
	MSH-P/CO2/	17.6	0.217
	Telaire T6615	4.42	0.185
†Methane/Hydrocarbon	MQ-4	0.48†	
	Gascard CH_4	35.7	0.222
	MSH-P/HC/	3.54	0.152
	TGS-2600	1.56	0.225
	TGS-2610	9.69	0.237
	TGS-2611	0.25	0.208

† This value was determined directly from the experimental response.

delay caused by the gas exchange when the carrier gas was introduced to be removed before analysis. Any initial overshoot and ringing, as observed for the chemiresistive sensor were also edited from the data. The California Analytical Instruments Inc. ZRE Non-Dispersive Infrared Analyzer was used to verify the concentrations at each point on the calibration curves. The instrument was calibrated using a gas mixture composed of carbon dioxide and methane gas in a nitrogen carrier gas.

Consistent with the Beer-Lambert law, the optical absorption based sensors all showed a linear response. The chemiresistive sensors produce non-linear calibration curves. Data sheets from the MQ-4 and TGS sensors showed significant non-linearity, especially at low concentrations. The response curve was modeled using a Langmuir-like or Langmuirian form, which provided a response consistent with that specified by the manufacturer. Attempts to use a power function yielded poor results. Additional corrections with kinetic analysis have been suggested (Barsan, Schweizer-Berberich, and Göpel†, 1999; Ahlers, Müller, and Doll, 2005), however a Langmurian form provides a simple fit. The following equation was utilized:

$$f(x) = \frac{a \times b \times x}{1 + b \times x}$$

Using this equation, a and b, along with the asymptotic standard errors, were determined by fitting the experimental data using the Levenberg-Marquardt algorithm in *gnuplot* (T. Williams and Kelley, 2016).

An example calibration plot with linear (for optical sensors) and non-linear (for chemiresistive sensors) trendlines is shown in Fig. 2.6. Since the calibration curves were produced with respect to the California Analytical Instruments Inc. ZRE, these curves provide an estimate of the accuracy as opposed to the precision discussed above. In general, it was found that all of the sensors must be calibrated before deployment. With the exception of the carbon dioxide and methane Gascard sensors, there was poor accuracy without an initial calibration. Fortunately, all the optical sensors were easily recalibrated, and the recalibration saved internally on the sensor. The chemiresistive required separate calibrations for each sensor, which must be applied by the monitoring device.

The limit of detection is the minimum concentration that can be detected as significantly different from the background (Long and Winefordner, 1983; Currie, 1997; Mocak J. et al., 2009). The International Union of Applied Chemistry (IUPAC) defines the limit of detection as three times the standard deviation (σ) from the background. For the sensors discussed here, the raw output from the sensors must be transformed into concentration value. Error in the calibration will affect the limit of detection. In the case of measurements requiring a calibration curve, Long and Winefordner provide a review of the various definitions as well as several examples (Long and Winefordner, 1983). If the calibration curve is linear, the procedure to calculate the limit of detection is straightforward (Long and Winefordner, 1983). For optical sensors, the limit of detection can be calculated from the errors in the slope and intercept as obtained by the calibration curve produced using the California Analytical Instruments Inc. ZRE as a reference. For the non-linear chemiresistive sensors, error propagation can also be used to determine the limit of detection. Given the non-linearity and steepness of response of the chemiresistive sensors at the detection limit, the limit of detection determined by 3σ can be significantly different than the limit of detection determined after correcting for errors in the calibration curve. It should be also noted that these tests were carried out in a controlled environment which eliminates many sources of error. In an application without environmental controls and unknown gas composition, these limits may be subject to additional errors. As an example, chemiresistive sensors are very sensitive to changes in temperature and humidity.

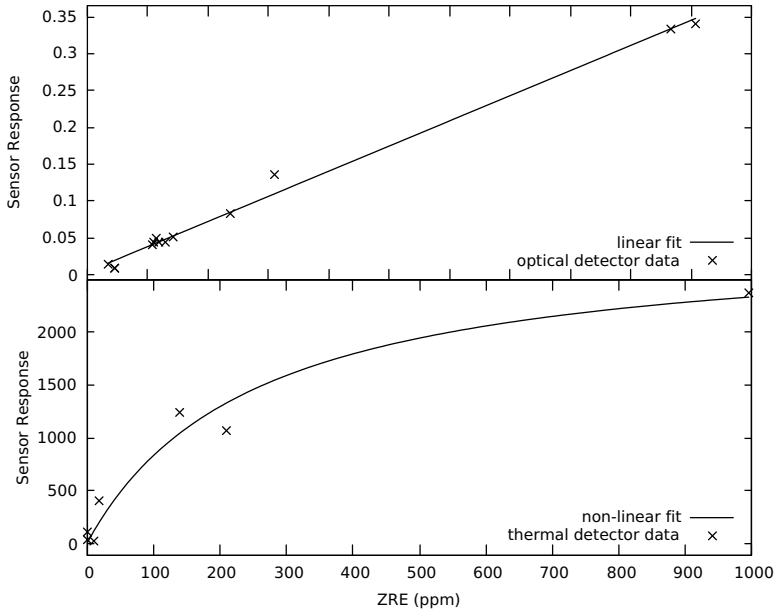


Figure 2.6: The top part of the figure shows the response of an optical carbon dioxide sensor (GasCard, response is a fraction of the 50000 ppm span) with the expected linear behavior. The bottom part of the figure shows the response of a chemiresistive methane sensor (MQ-4, response is value obtained from the 12-bit A/D converted) with a fit to the non-linear response.

Table 2.4: IUPAC and Calibration Corrected Limits of Detection (ppm).

	Sensor	IUPAC	Corrected†
Carbon Dioxide	K-30 SE-0018	5.7	25
	COZIR AMB GC-020	42	74
	Gascard CO_2	6.5	54
	MSH-DP/HC/CO2/	260	280
	MSH-P/CO2/	53	76
	Telaire T6615	13	27
Methane	MQ-4	53	82
	Gascard CH_4	110	151
	MSH-P/HC/	11	170
	TGS-2600	74	120
	TGS-2610	74	110
	TGS-2611	11	16

† A linear calibration was utilized for the optical sensor, while the chemiresistive utilized the non-linear Langmuirian fit.

2.3 Sensor Array Design, Costs and Construction

2.3.1 Introduction

To facilitate data collection, a tiered hierarchy was used for the sensor network. An overview is shown in Table 2.5. Tier 0 is the low-cost, low-power, and transportable sensor nodes. The nodes collect the data, so they require gas sensors, hardware for solar power, and radio for simple wireless communication to other nodes. Tier 1 is the commination nodes. These nodes are larger, more expensive, and more difficult to transport and setup. In addition to having the same set of sensors as the Tier 0 devices, the larger size and extra power allowed for the inclusion of a more accurate methane sensor. These communication nodes receive data from the Tier 0 sensors, store the data locally, and transmit the data via a cell modem to the Tier 2 node. A Tier 2 node is the highest tier of the network. This is a computational workstation housed at the base of operations. This computer is connected to the municipal power supply and the internet. It acts as a receiver of the field data, and has the computational ability to process and analyze the data.

Table 2.5: Hierarchy and respective requirements of devices in the network array designed for this project.

Tier and Name	Device Requirements
Tier 2 Data Repository Server	Large Reliable Storage Space Connected to the Municipal Power supply Directly Connected to the Internet
Tier 1 Communication Node	Moderate Power Tier 0 Set of Sensors Optical methane sensors Wireless Mesh Network Node Cellular Modem
Tier 0 Sensor Node	Low Power Diffusion based gas sensors Wireless Mesh Network Node

The nodes can be generalized as a set of functional units, such as depicted in Fig. 2.7. The functional units are grouped as “Power”, “Processing Unit”, and “Sensor Elements and Electronics.” The communication node has an extra unit for the optical methane sensor. Although the power supply capacity of each node on the tier differs, the power supervisor and supplies required for each unit are essentially the same. The “Processing Unit” is between tiers as well.

With low-power environmental sensors, the specific needs of the collected data must be weighed against the complexity and cost of the integrated system. The performance of the

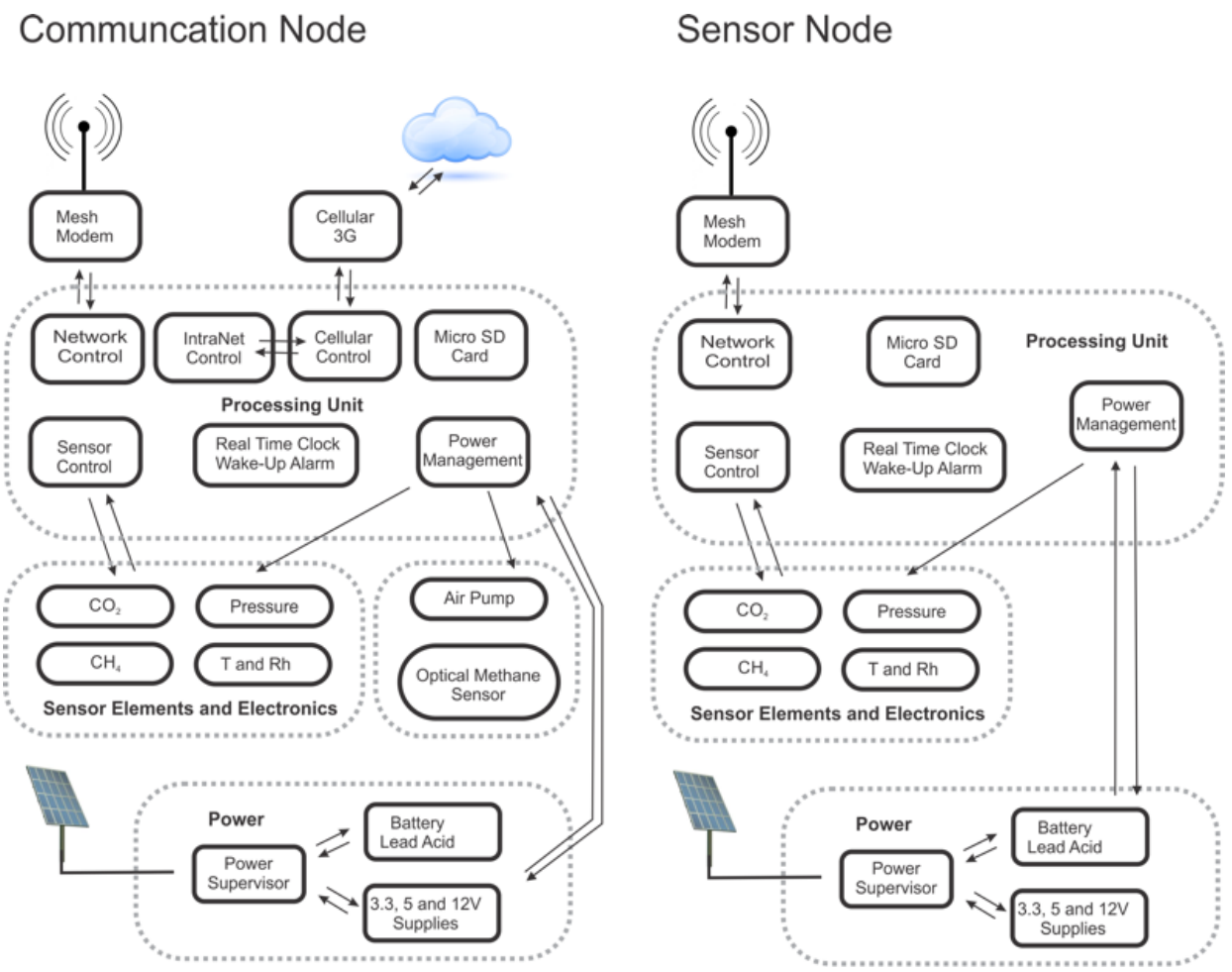


Figure 2.7: Block diagram of the communication and sensor nodes.

selected sensor for testing are shown in Tables 2.3 and 2.4. A K-30 sensor was picked for our system due to performance, low-power and low-cost. In the case of methane, there is no commercially available sensor that provides ppm sensitivity at atmospheric levels, with low power consumption, low price, and availability. The ready availability and low cost of the MQ-4 sensor lead to the selection of this sensor for our units. For the larger communication nodes that have a larger solar array, the more expensive Gascard methane sensor was chosen. This sensor require a mechanical pump for sampling, which results in additional power consumption, a mechanical failure point, and sampling issues. On the other hand, this sensor also corrects for temperature and atmospheric pressure, eliminating additional sensors for accurate work.

Table 2.6 shows the estimated cost for construction for the communication and sensor nodes. The cost is estimated based on the construction of approximately 15 communication and 120 sensor nodes with final design. Shipping cost were not included. The PCB costs includes board and the assembly of all components on the board. With an increase in the

number of units, there was significant cost savings in the manufacturing and assembly of the printed circuit board. The cost of the enclosures included the batters and solar arrays required for operation. In this work, an additional $20 was spent on a new battery for the sensor nodes. During the design process, a number of different chemoresistive sensors include the MQ-4 from Hanwei Electronics and TGS-2611 manufactured by Figaro Engineering Inc were tested. The MQ-4 was chosen due do its reasonable performance, cost and availability within the time scale of the project. There was no attempt made to quantify the labor cost for the assembly.

Table 2.6: Component Cost of the Sensor and Communication Node.

		Sensor Node	Communication Node
Control Board	PCB	$59.02	$117.91
	Parts	$130.72	$168.51
Sensor Board	PCB	$19.28	$19.28
	Parts	$65.49	$65.49
Sensors	MQ-4	$4.90	$4.90
	K-30	$85.00	$85.00
	Gascard		$1,448.00
Air Sampling	$3D Part	$18.18	
	Pumps		$255.21
	Lexan Box		unpriced
Cell Board	PCB		$33.00
	Parts		$204.06
Enclosure (includes battery and solar array)		$165.56	$1,099.96
	Total	**$548.15**	**$3,501.32**

2.3.2 Design Overview and Cost

The nature of the sensor array allows it to cover a large area but imposes limitations on the power of the system. Since the area to be covered may include elevation changes, anthropogenic developments, and other uncontrollable factors, it would not be advisable to constrain the array to wired power connection as the array needs to be adaptable to the site. Therefore, each unit was designed with internal and self-sustaining power supplies and methods of communication between the nodes. Two different types of commercially available solar power generation and storage units with weatherproof enclosures were selected from Tycon Power Systems. The majority of the sensors were housed in the Remote Pro 2.5 W Continuous Remote Power System die cast enclosure (Fig. 2.8A) and a few sensor units with features requiring more power were housed in the Remote Pro 15 W Continuous Remote Power System steel enclosure (Fig. 2.8B). The 2.5 W enclosure includes a 12 V battery rated

for 9 Ah of use, a charging and distribution circuit, and a 10 W solar panel. The 15 W enclosure includes two 12 V batteries rated for 98 Ah, a charging and distribution circuit, and a 60 W solar panel (Tycon Power Systems, 2014).

The power consumption of the Arduino development boards was found to quickly drain the solar cells and batteries. The Arduino prototype was found to use 2.7 W of power, which was higher than continuous use rating of 2.5 W for the small enclosure power system. To address this issue, a dedicated control system, power supply and power management circuits were designed and a voltage monitoring circuit was added. The final microcontroller can detect user specified threshold voltages, and selectively shut sub-sections of the monitoring device off if the voltages are too low.

Control and Sensor Boards:

The control board is designed in such a way as to minimize wasted power. A selection of parts considered for use in the board were measured for current and voltage to determine their power use. The control board designed for this project uses 1.7 W during full use. To prepare for times when the solar panel cannot provide continuous charging, a problem that potentially can last for long periods in the deployment area. Weather data statistics for the field site show that 32% of the year is at least half-cloudy and 8% of the year is heavily overcast (KSWO, 2015). The control board can select a power mode which reduces the frequency of sampling by some of the more energy expensive sensors. This is automatically triggered by the control program if the battery charge is determined to be at 70% charge. A hibernation mode was also programmed. Below a certain battery voltage threshold, the unit can hibernate to prevent complete power drain, and it will restart again once the battery registers above the threshold voltage. Using only 0.1 mW during hibernation, this mode prevents the units from requiring manned intervention during long periods of overcast skies.

Figure 2.9 shows the final control board with all components included. This board is included in all of the high level communication nodes in the network. Sensor nodes use the same printed circuit board, but do not have electrical components included for unit parts not incorporated into the sensor nodes, such as the Gascard sensor and cellular modem.

The board commands are coordinated by an Atmel ATmega 2560 lower power 8-bit microcontroller with 64 kB flash operating on a RISC architecture (referred to as ATmega 2560 in this report). The ATmega 2560 is under clocked to save power using a MA-506 8MHz ±30 ppm crystal oscillator with equivalent length traces. Logging is managed using timestamps from a DS3231 real-time clock. The real-time clock uses a CR-2016 coincell battery to maintain correct count of the date. Data can be logged using two memory storage methods. The first is 24LC1026 1024 kb serial EEPROM. The second is a micro Secure Digital stable storage, an Alps SCHD3A0100 micro SD card holder. For node to node communication, the XBee-Pro 900HP model was chosen. Both the communication and the sensor nodes types

Figure 2.8: The (A) Remote Pro 2.5 W Continuous Remote Power System and (B) Remote Pro 15 W Continuous Remote Power System–image from Tycon Power Systems' website shows typical units.

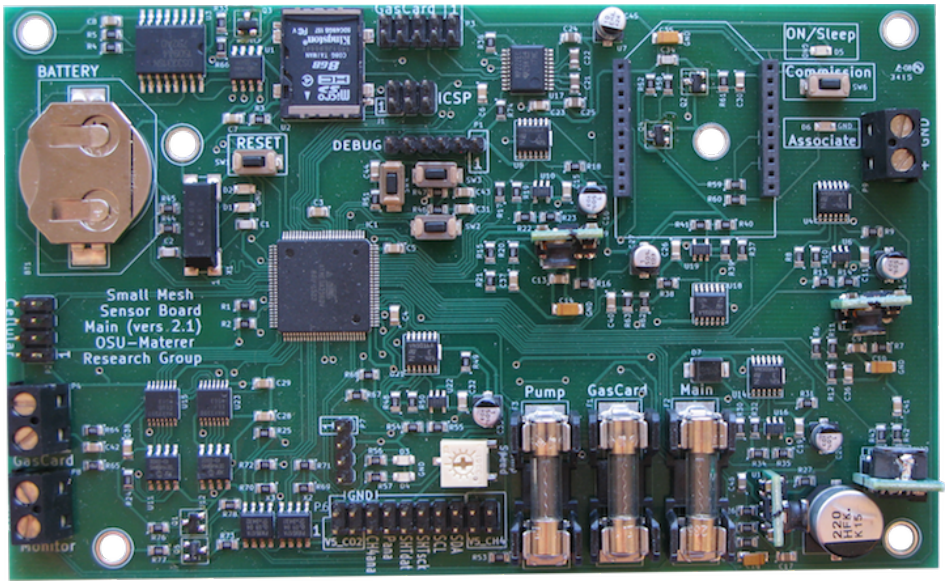


Figure 2.9: The fully built communication node control board.

share a set of the inexpensive sensors. These parts are collected on a breakout board referred as the "sensor board." Figure 2.10 shows the K series carbon dioxide, temperature and humidity, pressure, and MQ-4 methane sensors attached to the board. The communication node has a NimbeLink Skywire EVDO cellular modem on a separate breakout board, and also have an additional sensor, the Edinburgh Gascard for methane. Both the cell modem

and the Gasscard connect to the control board using a ribbon cable.

Air Sampling for the Sensor Node:

Samples of the ambient air at the field site are collected by both the small sensor nodes and the large communication nodes. To minimize the power expenditure, the small nodes are set to passively sample the environmental conditions. The passive sampling method requires the sensor components to be directly exposed to the environment. Placing an exposed circuit board into the relatively harsh conditions of the field site invites potential issues from moisture causing corrosion, animal life interfering with the fragile components, and damage from collisions. To mediate this problem, it was determined that a plastic housing to isolate the sensor components from the circuit boards and power supply electronics should be constructed. With the solar panel enclosures already selected, the sensors would be placed in such a manner that they would have contact with the environment through holes pre-drilled on the enclosure. These holes are oriented towards the ground to prevent collection of rainwater and moisture. The plastic housing was designed in a program called OpenSCAD, a free 3-D computer aided design (CAD) program which renders a 3D object from a script file (Kintel and Clifford Wolf, 2011). The interface objects were printed by i.materialise on an EOSINT P 700 3-D printer by selective laser sintering of polyamide granules. The 3-D printed part is attached using 3/4" standoffs which provide clearance from the board and sensors (see Fig. 2.11). An isolating enclosure for the temperature and pressure sensors is mounted flush against the PCB circuit board. Any void is filled with an epoxy potting compound to seal the areas which are exposed to the environment from interior of the enclosure. No potting compound is necessary for the carbon dioxide sensor, as part is made to press against the outer edge of the membrane on that sensor and forms a seal.

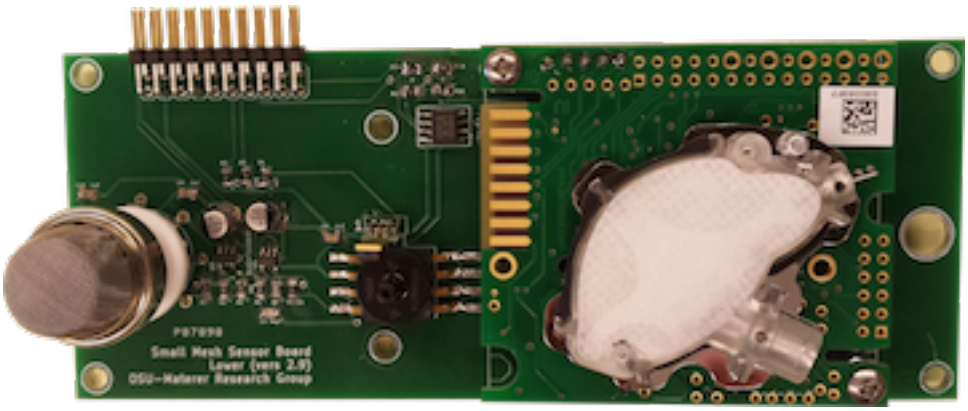


Figure 2.10: A picture of the sensor board with encapsulating materials removed.

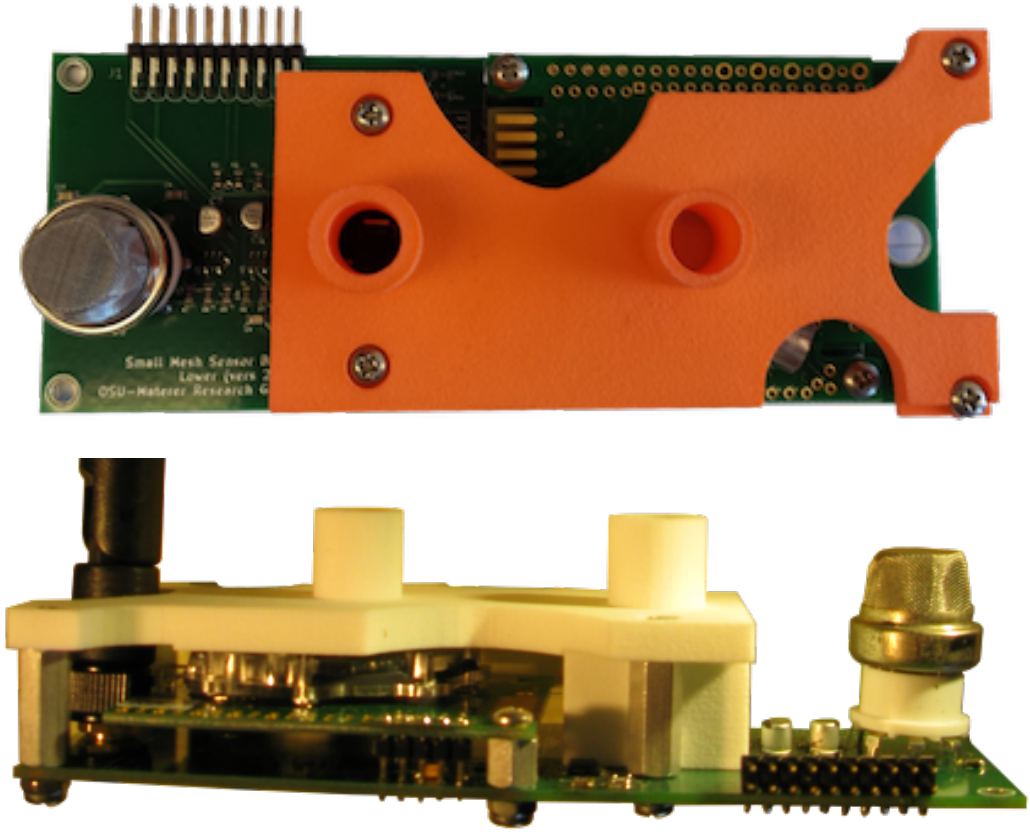


Figure 2.11: Top and side view of the 3-D printed part attached to the standoffs on the sensor board, showing the flush interface with the board and the carbon dioxide sensor.

Air Sampling for the Communication Node

The larger power reserves in the communication nodes allow active sampling. For these larger nodes, a Thomas 1410D/2.2/E/BLDC diaphragm pump pulls air from outside of the controller through a hole on the underside of the enclosure, through a 0.45 μm particle filter, to the pump diaphragm. Air is pushed out from the pump to a small plastic housing with the sensors that are used on the node sensors, then through the Gascard, and finally vented outside the enclosure. Since it was essential to use the same sensor board in the communication node as is used in the sensor nodes to enable direct comparison of results, an internal housing was designed to hold the sensor board.

2.3.3 Communication Scheme

In the final version of the software, the unprocessed sensor data and diagnostic information is collected every 15 minutes and stored in the ring buffer (see Block Diagram in Fig. 2.7). About one month of data can be stored before the data must be written to the SD card. Depending on the battery charge, the new data are written to the ring buffer on the SD card once an hour. If the system charge is low, the system will wait until the next specified time to transmit the data. The SD card writing and transmission can be performed separately, depending on the current battery charge. The network routing is optimized every hour. Packets can transfer from one node to another node as they travel to the communication node. This increases both the range and reliability of the system. The power-managed SD card provides a local backup.

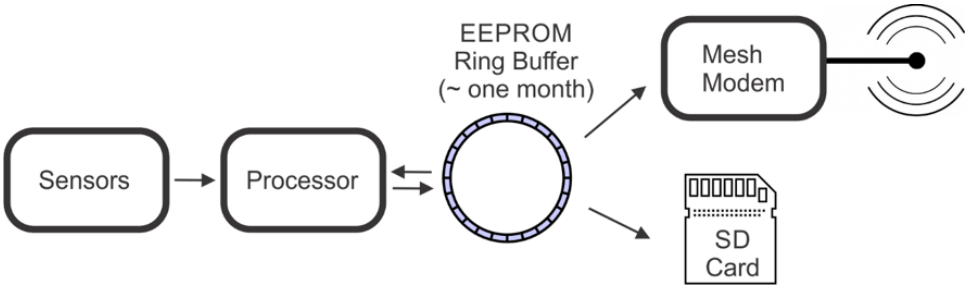


Figure 2.12: Block diagram of the data collection process on a sensor node.

File rotation is performed every week to ensure that the data file is kept at a reasonable size.

The communication nodes direct the self-assembly for the mesh network. Data comes at specific times from each mesh sensor and is identified by the modemś 64-bit identification number. The processor writes all the data to the ring buffer and periodically transfers the data to the SD card (see Block Diagram in Fig. 2.13). Again, file rotation is performed,

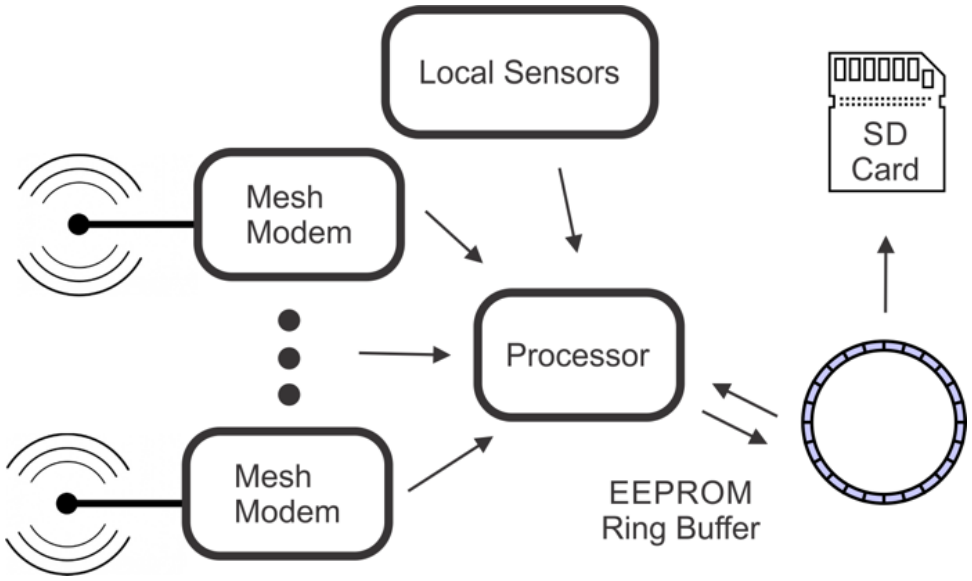


Figure 2.13: Block diagram of the data collection process on a communication node.

keeping the files at a manageable size. The communication nodes have been tested with a network containing 20 sensors nodes. Currently, the maximum number of sensors per communication node should scale to 55 sensors nodes.

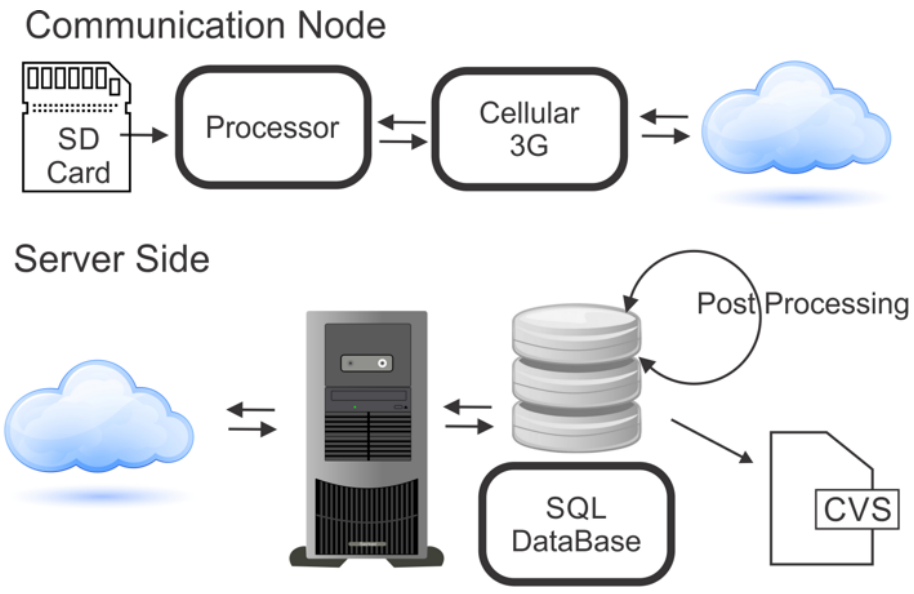


Figure 2.14: Communication to the server for each communication node.

Once an hour, the communication nodes connect to a server at OSU. The nodes transmit the directory of the SD card, containing the collected data from all of the sensor nodes, to the server (Fig. 2.14). At this time, it is possible for the server to request files from any sensor

node on the mesh network, or change performance parameters contained on each sensor node. In a typical operation, the server looks at the file sizes on the SD card and requests any new data as indicated by an increase in the file size. The final line count is recorded by the server along with the new file size to ensure that the next data transfer continues where the previous one left off. The raw data are stored in a SQL database. Periodically, the raw data are processed to obtain the measured values from the sensors and other monitoring points. Each set of communication nodes has a unique database for the raw and processed data. This second step process allows all checksums to be validated, and a variety of post-processing to be performed without modifying the raw data.

Sensor networks data and node parameters are stored in SQL database which is queried during data upload. Basic information on the sensor history is also stored in this database and can be accessed at http://sensormesh.chem.okstate.edu/sensors/. The data collected are only accessible though the SQL server, but we plan to have the data available through a password protected web page at both the network and sensor level. The data in the CSV file can be grouped by location, communication nodes, or even a given sensor node. Currently, there are over one million data points on the SQL server.

2.3.4 Production of Sensor and Communication Nodes

Circuitry for the individual units was produced in stages. The sensor breakout board was printed and assembled early in the first quarter of 2015, the control board for both the communication and sensor nodes was printed and assembled in the 3rd quarter of 2015, and the cellular breakout board was printed early in the fourth quarter of 2015. All printed circuit boards and pick-and-place assembly was contracted through Advanced Circuits in Aurora, Colorado. Since the sensor boards were designed first, they were the first to be constructed. These completed sensor boards were used during the Arduino prototyping phase so device design would employ the actual sensors. This enabled development and honing of the serial communication with the sensors to and from the microcontroller. The control boards were produced after developing several prototypes and printed in a single batch.

Construction of sensor nodes began early the second quarter of 2015. When the sensor circuit board and plastic part were attached to the enclosure, a small amount of silicone was applied to the flat surface of the plastic part, and the piece slipped into the correct position on the enclosure. Due to problems with small invertebrates making homes in the sensor holes during the prototyping phase, a 1 mm mesh screen is glued over the holes before the unit is deployed. This final assembly phase was completed in waves of ten units at a time. After the engineers assembled the nodes, the devices were given a quality control inspection and programmed. The final unit is shown in Fig. 2.15. A total of 122 sensor nodes were constructed.

Construction of the communication nodes began near the end of the 1st quarter in 2015. Clear Lexan boxes were built in house to house the sensor breakout board for active sampling.

Figure 2.15: A photo depicted the internal arrangement of boards and wiring within a completed sensor node.

Circuit boards were mounted to an orange back plate secured within the enclosure (Fig. 2.16). Only the top half of the plate was practical for mounting circuit boards, as the large batteries inside took up half of the volume of the enclosure.

Vinyl tubing was used to plumb the unit. Input and output ports for gas were placed through the bottom of the enclosure to pull gas from a location with a similar orientation to the passive interfaces on the sensor nodes. The ports were placed on opposite ends of the bottom side of the enclosure to minimize recycling of gas which had already been analyzed. Each communication node was tested in the field, much like the sensor nodes, for quality assurance of the units. A total of 15 communication nodes were constructed.

Figure 2.16: (TOP) The layout of components within the large communication nodes. The batteries take up most of the space inside the enclosure, so components were arranged to be bolted to the top of the back plate. (BOTTOM) Behind the batteries, tubing to direct the air for the active sampling is laid.

2.4 Field Evaluation

2.4.1 Introduction

Testing of the sensor nodes was ongoing in a network that was deployed on the OSU campus (see Fig. 2.17). The goal was to collect baseline data, identify possible issues which might be

encountered during field deployment, and to develop the network technology. Each sensor nodes was attached to one T-post and the communication nodes were attached to two T-posts.

Figure 2.17: Approximately 50 sensors in the field at OSU for testing and long-term data collection.

2.4.2 Sensor Nodes

To date, a total of 122 sensor nodes have been tested for performance of sensor elements and power management, each for a minimum duration of four months. One set of 15 units has been collecting data since Oct 2015, providing a long term dataset for baseline determinations.

Initial testing was performed with an Arduino board acting as an interim control board to test the reliability of the setup in the field. In this test, four units were outfitted with Arduino control boards with data logging and deployed in a 4 yard x 4 yard region on the OSU campus. These units proved to be resilient through a week-long storm during late May 2015. This includes a storm with gusts exceeding 70 mph. In this storm, some some sizable pieces of debris struck the some of the units and caused them to fall from their poles. While the enclosures suffered heavy, but cosmetic damage, all units continued to function properly. This incident demonstrated the resilience of the units and indicated that the sensor mounts needed to be strengthened. During these tests, we found and corrected several unexpected issues with the software. In addition, we confirmed the importance of power management to the success of the field development of the CH_4 sensor.

As more sensors were added various issues, from bugs in the embedded software to construction issues, were identified and addressed. In this testing, we found failure rates of about 1% of our sensors, mostly due to unexpected construction issues. During assembly, we found at least two sensor assemblies and four control board that failed out of a total of

150 units. There was also one failure related to the solar battery charger. A checklist was developed and continually refined to eliminate problem in the final assembled unit. This reduced the failure rate to zero for the sensors installed in the field. The pressure sensor had three units fail, which was the highest failure rate of all of the sensors. Only one methane sensor and one carbon dioxide sensor developed abnormal readings.

2.4.3 Communication Nodes

The OSU field site was also critical for the testing of the communication nodes and the uplink through the cellular network to a server. Fifteen communication nodes were built, which contain the cell modem and a high quality optical gas CH sensor. Three have been tested at Farnsworth, three were tested on the OSU campus, and six more are installed at OSUś unmanned airport flight station. By analyzing the test data several improvements were made. Code to automate the sensor configuration over the XBee network and cell networks was successfully created to gather over 1,000,000 data points.

2.4.4 Evaluation data

As the mass production of units was completed, a set of ten sensors nodes was deployed in the same proving ground area as the previous four prototype sensors. This new network was dubbed the Long-Term Study Array. These sensors have been actively returning data starting on Tuesday, 20 October 2015. As an example of the data collected for the temperature and carbon dioxide the graph was generated (see Fig. 2.18) from the CO_2 and temperature sensors for one of the units at the OSU test site between Tuesday, 20 Oct 2015 and Friday, 29 Apr 2016. Some seasonal variations in CO_2 due to temperature can be seen. At higher magnifications, the day/night variation of the CO_2 levels is also visible.

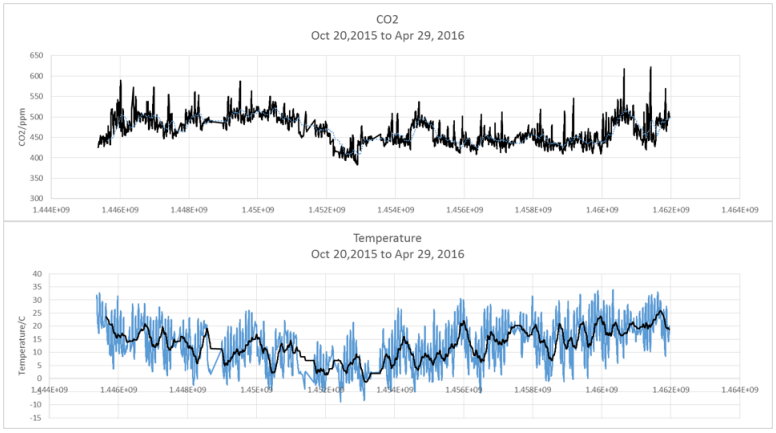


Figure 2.18: Carbon dioxide and temperature sensor data for a unit at the OSU test side between Tue, 20 Oct 2015 and Fri, 29 Apr 2016.

There is no local reporting agency for carbon dioxide and methane concentrations, and thus validation of the concentration values reported by the sensors is not simple. Instead, the data were considered good if those reported by each sensor tracked well with the other sensors in the network. However, the temperature, humidity, and pressure data can be compared against data reported from a local weather station (KSWO, 2016). Data collected from an archive of reports from KOKSTILL4 database provided values for the maximum, minimum, and average reported temperature, humidity, and pressure. These values were plotted as three lines of the individual values of all of the sensors. The temperature and humidity data, shown in Fig. 2.19 and Fig. 2.20 respectively, are inline with the weather station data. The sensors show a regular cycling of these values, corresponding to the day-night cycle. The peaks and troughs of this cycle match closely with the minimum and maximum weather station data. The plot of pressure data (Fig. 2.21) differs, in that the weather station data and the test data seem to be offset. The altitude of the KOKSTILL4 station (935 ft. above sea-level) does not differ enough from the test site (944 ft. above sea-level) to suggest that there should be a significant difference in pressure. It is possible that there is some minor variation between the pressure in the two locations, approximately 2 miles apart. The sensor used by the weather station may also be inaccurate, there is no published calibration data for this station. It is also possible that the pressure sensors used by the sensor nodes are incorrect. Due to the large number of sensors involved and the careful calibration of each one, this is less likely.

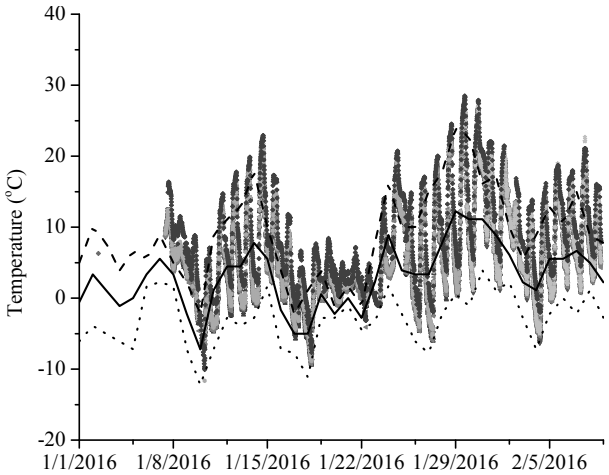


Figure 2.19: Temperature data collected from test sensors tracks with the reported weather data. Diurnal cycling is apparent. The weather data is from KOKSTILL4 weather station (KSWO, 2016). The dashed line and dotted line are the maximum and minimum value observed for each date, and the black line is the average value observed for that date.

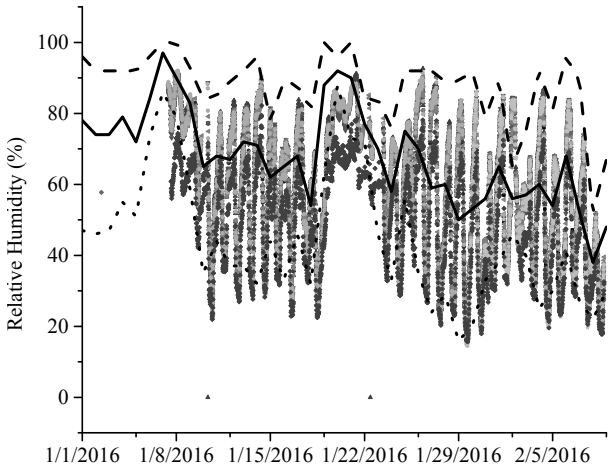


Figure 2.20: Humidity data collected from test sensors tracks with the reported weather data. Diurnal cycling is apparent. The weather data is from KOKSTILL4 weather station (KSWO, 2016). The dashed line and dotted line are the maximum and minimum value observed for each date, and the black line is the average value observed for that date.

2.4.5 Field Performance

The testing site on the OSU campus provided valuable long-term information about the capabilities of the sensor network. Sensors were shown to be capable of withstanding dangerous weather and seasonal extremes of temperature.. The long-term study provides a reasonable baseline for the local area, and the data are shown to track well as a group. All sensors deployed in this group were shown to produce data reasonably close to accepted values. Observation of the day/night variation (see Fig. 2.13) provides good evidence that the sensor and network is preforming as expected.

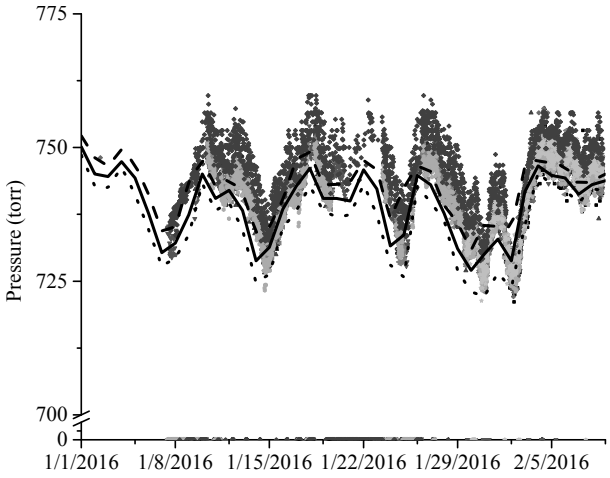


Figure 2.21: Pressure data collected from the test sensors is acceptably precise yet consistently lower than the reported weather data. This may indicate deviation of the sensors from the true value, local variation in pressure, or an inaccurate report from the weather station. The weather data is from KOKSTILL4 weather station (KSWO, 2016). The dashed line and dotted line are the maximum and minimum value observed for each date, and the black line is the average value observed for that date.

2.5 Array Deployment

2.5.1 Introduction

There are two field sites where the sensor array was deployed. One site is the unmanned airport flight station at OSU and the other is at the Farnsworth, TX field site. At the unmanned airport flight station, 6 communication nodes for the array and 28 sensor nodes were installed around the perimeter. Like the evaluation site, each sensor node was attached to one t-post and the communication nodes were attached to two posts. Farming practices in and around the Farnsworth Field made it necessary to mount the sensor nodes on utility poles. At the request of the site maintainers from the Southwest Regional Partnership, special hardware was designed to allow mounting without damaging the utility poles.

2.5.2 Unmanned Aircraft Systems Airfield site

The first external site was the Oklahoma State University Mechanical and Aerospace Engineering's Unmanned Aircraft Systems (UAS) Airfield. This site, located approximately 16 miles east of the Stillwater campus, is one of the few locations owned by a university which

permits the testing of unmanned aerial vehicles. The fifteen acre tract of land includes a garage, control facility, and polymer mat runway. A network of sensors developed for this project were deployed at the site in conjunction with flights of aerial sensors.

2.5.3 Installation

At the UAS field, a network consisting of 34 nodes (Figure 2.22) was installed. Metal T-posts were erected in regular intervals along the fence line. Unlike the network discussed in Section 2.4 which consisted of a single network of sensors, this deployment includes multiple subnets. In Figure 2.22, the individual subnets are color coded. This allowed for more communication nodes with the optical methane sensor and for testing the network support for multiple subnets.

Figure 2.22: Configuration of the sensor the unmanned airport flight station. The larger circles represent communication nodes and the smaller circles are the sensor nodes.

2.5.4 Initial data analysis

Data were successfully collected by the sensors in the network, and sent to the server on the OSU main campus. Data for the study were collected of a period of approximately nine months. Collection is still ongoing, but the data is not included in the study. During this period, there were variations in pressure, temperature, and humidity, as can be expected of an outdoor testing period of this time scale. Data from each sensor were collected and sorted by time of day. The range of values reported for these time points can be seen in Figure 2.23. The data from the carbon dioxide sensor shows that the concentration seems

to be highest in the early morning hours, and decreases during the day. Just the opposite was observed for methane, with peak concentration occurring in the middle of the day. This cycle matches with reports on diel concentration flux by various authors and in various environments (Tarnawski et al., 1994; Yun et al., 2013; Lai et al., 2012; McGinn et al., 2014). The generally accepted cause of these variations is the flux between the ground and flora with the atmosphere. Carbon dioxide levels increase during the night as plants cease to photosynthesize. Methane levels increase during the day as the sun drives processes by the microorganism in the soil that produces methane. Changes in relative humidity also play a role in this cycle by reducing the partial pressure of carbon dioxide and methane components with respect to the partial pressure of the water vapor.

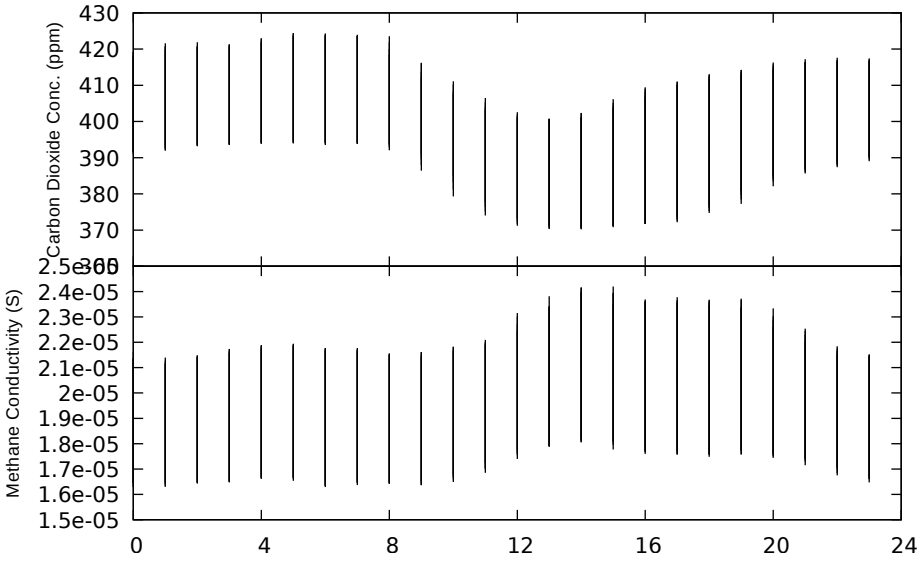


Figure 2.23: The range of values recorded during each daily cycle at set time points. The plot shows the concentration changes in the typical day/night cycle at the site.

2.5.5 Farnsworth Texas site

The ultimate goal of the networked sensor array was to deploy it at an injection well site. A location was selected in the panhandle of Texas near the town of Farnsworth. The Farnsworth Field has been prolific producer of hydrocarbons for many years, and is now being produced using a CO_2 enhanced recovery process. A detailed study of the enhanced recovery operation is managed by the Southwest Regional Partnership (Ball, Henry, and Frezon, 1991; White et al., 2014). The test site at Farnsworth is a pilot carbon dioxide injection well in the Morrow geological formation. Most of the site consisted of farmland, which required special care in placing the sensor nodes. At the site where the sensors were deployed, both cotton and corn were grown during the course of this experiment. Upon investigation of the site, it

became apparent that sensor arrangement would have to be along the existing access roads. As the topsoil is actively used in center pivot irrigation farming. Deploying sensors within these circles would inhibit agricultural use, and create problems. Logistics coordination with Southwest Regional Partnership and the transporting of the sensors to a site several hours from the Stillwater campus was challenging. To prevent damage to the sensor nodes, the sensors were mounted on existing wooden utility poles at the site. This limited the number and location of the possible sites. A total of 33 units were deployed to the site in three subnets. The location of these sensors and networks is shown in Figure 2.24. The poles are on access roads surrounding the primary injection well.

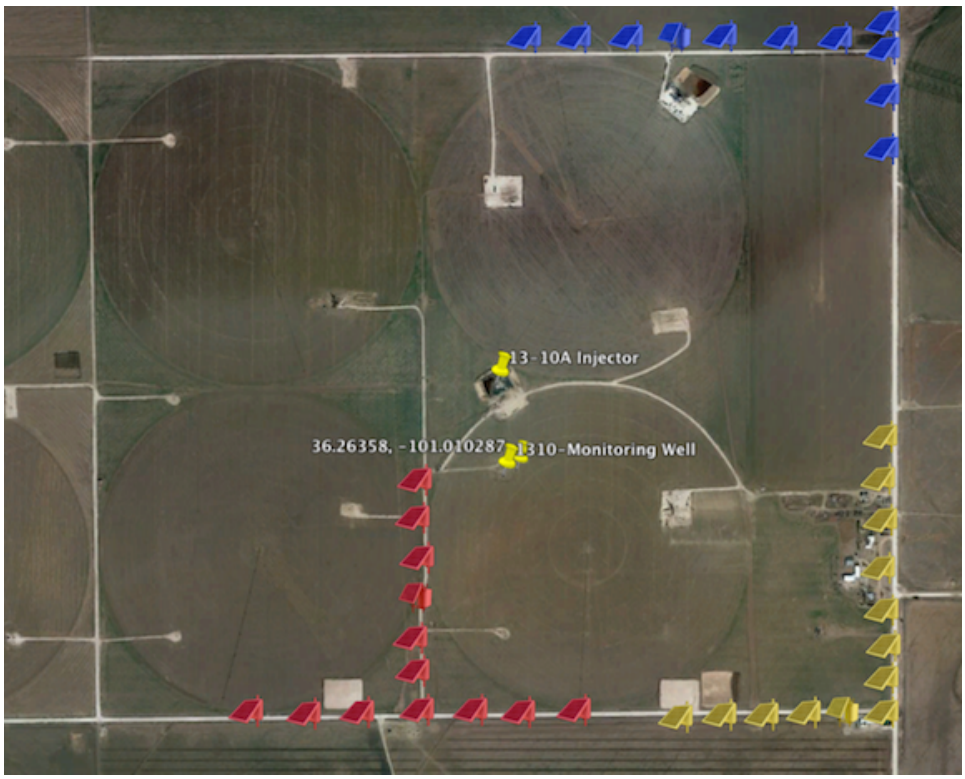


Figure 2.24: Google Maps satellite image of the Farnsworth, TX site depicts the approximate locations of sensors at the site. The colored groups depict the individual subnets in the sensor network. Sensor locations were determined in part by the placement of existing power poles.

To facilitate installation on utility poles, a different mounting method from the existing T-post configuration was designed. A piece of U-channel material (both aluminum and high density polyethylene) was fitted with holes to mount the sensor with bolts on the broad part of the channel and two slots on each arm of the channel. These slots allow pieces of stainless steel strapping to be fit around the pole and secure the sensor. Workers on ladders could quickly mount these units with the strapping, as shown in Figure 2.25. After practice, each sensor took about 10 minutes to install.

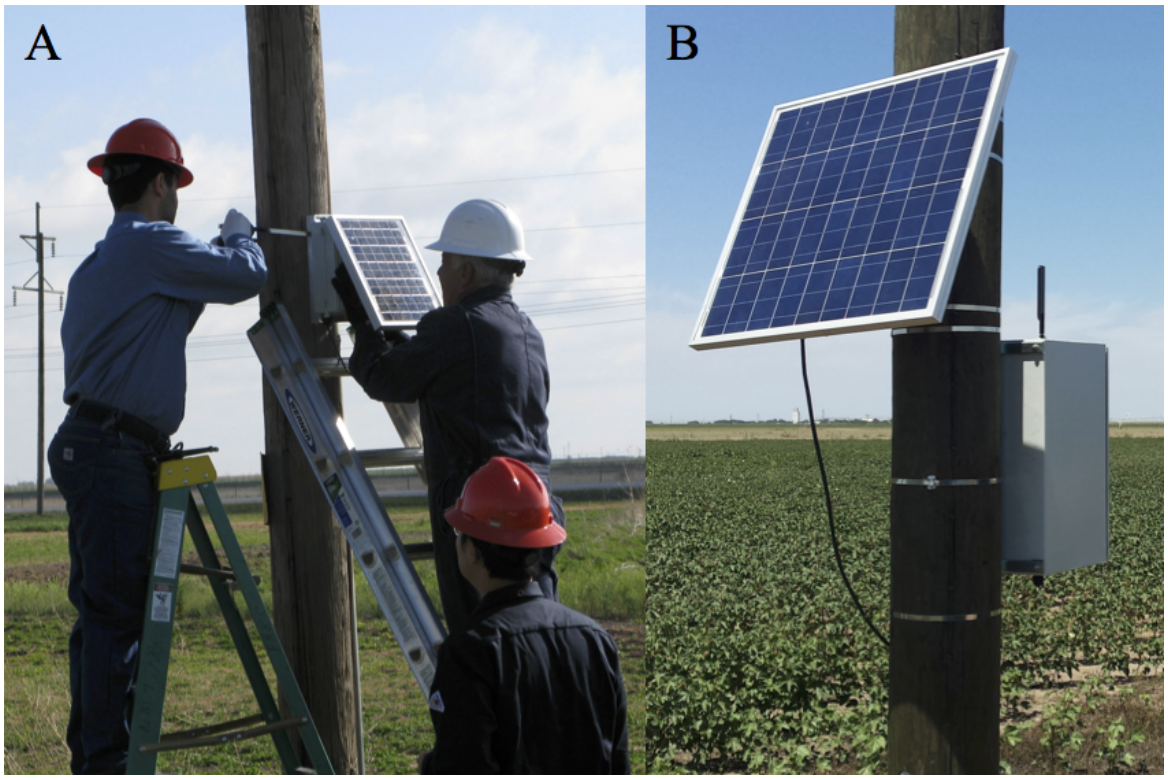


Figure 2.25: (A) Sensor and (B) communication nodes installed in Farnsworth, TX. The nodes, equipped with special mounting brackets, were strapped to the poles with steel strapping. At the request of the site maintainers, great care was taken to not damage the poles.

The first network (blue in Figure 2.24) was deployed in April 2016, and the other two subnets were deployed in August of 2016. The units collected data until February 2017. Around this time, the Farnsworth, TX region had a severe ice storm. This storm destroyed many of the power poles in the region and caused some damaged to a few of the sensors. The sensors were then retrieved from the Farnsworth site as the project was near completion.

2.6 Data Analysis and Results for Farnsworth, TX

2.6.1 Introduction

At the Farnsworth field site thirty-four nodes were deployed. All thirty-four of these nodes track the time-evolving concentration change of both carbon dioxide and methane. In addition to collecting concentration data, three of the nodes served as communication nodes. All the sensors recorded the temperature, pressure, humidity, and gas concentration data every 15 minutes. Periodically the sensor nodes transfer their data to the communication node and then communication nodes upload the data to a server on the OSU campus. There were a

few time periods where there was no data recorded because of low sunlight or unsuccessful transfer of the data. Since each sensor node had an SD card, the information collected during these down periods was not lost. It was, however, necessary to retrieve the data manually..

2.6.2 Level value correction

After organizing the database, results were plotted versus time. Figure 2.26 shows the time-evolving carbon dioxide concentration over a five month period from two representative sensors located at the Farnsworth field site. These two sensors have similar carbon dioxide measurement trends with about a 100 ppm offset. When discussing the calibration (see Section 2.2.3), it was noted that every sensor requires calibration before use. This offset between sensors is caused by different device geometry and the physical properties of the gas. The geometry of the intake and exit to the measurement region is different for the diffusion and flow through sensors. Each sensor was initially "zeroed" to 400 ppm and the K-30 units should autocalibrate. However, in practice there are still differences in the offset between the various sensors. Given that the focus of this work is to detect a significant "event" or time period of high gas concentration then the specific measured concentration is not as important as the ability to detect changes in gas concentration. This will be discussed in more detail later.

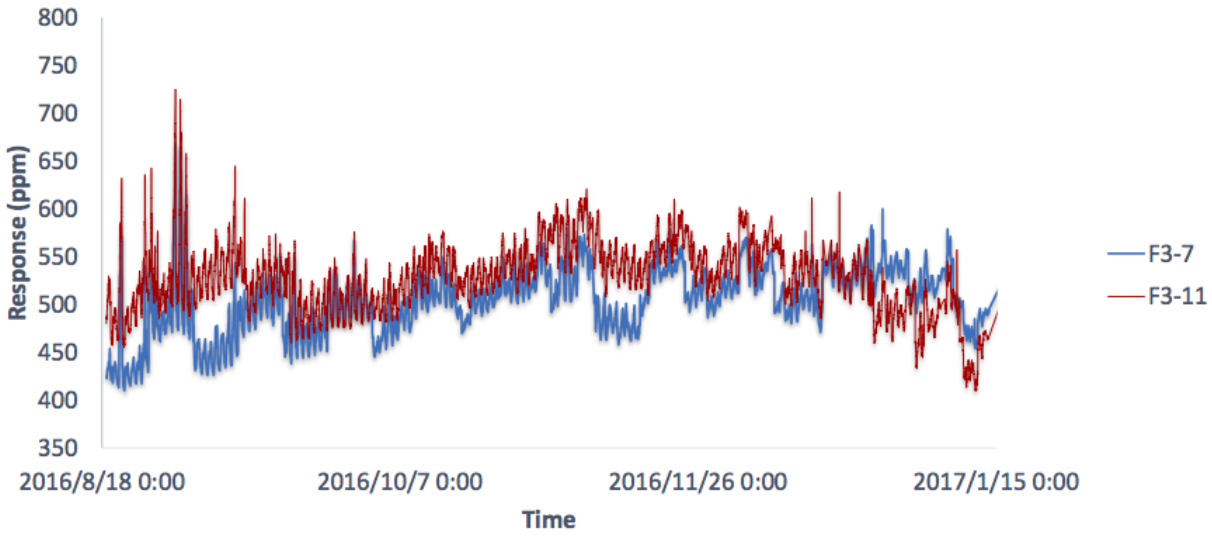


Figure 2.26: Time-evolving carbon dioxide concentration from two sensors (F3-7 and F3-11) in Farnsworth, TX field site from Aug. 18th, 2016 to Jan. 15th, 2017.

To investigate this further, a similar plot was made for all the sensors for both carbon dioxide and methane. The results can be found in Fig. 2.27 and 2.28. These plots show that

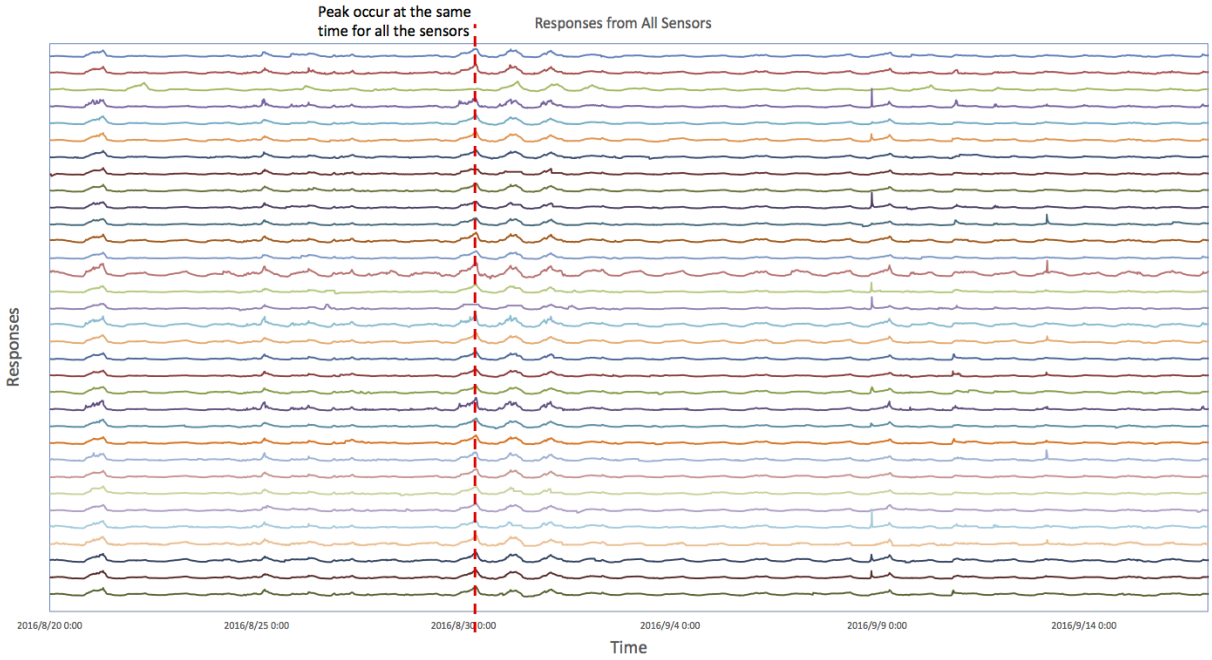


Figure 2.27: A comparison of all carbon dioxide sensors at the Farnsworth field site for 7 months. The peaks show higher detected levels of gas concentration. These peaks seem to occur at similar time periods for all of the sensors. See Appendix 2 for a plot of CO_2 data that covers approximately 320 days.

there is a similar trend between all of the sensors when they are compared over a long time period.

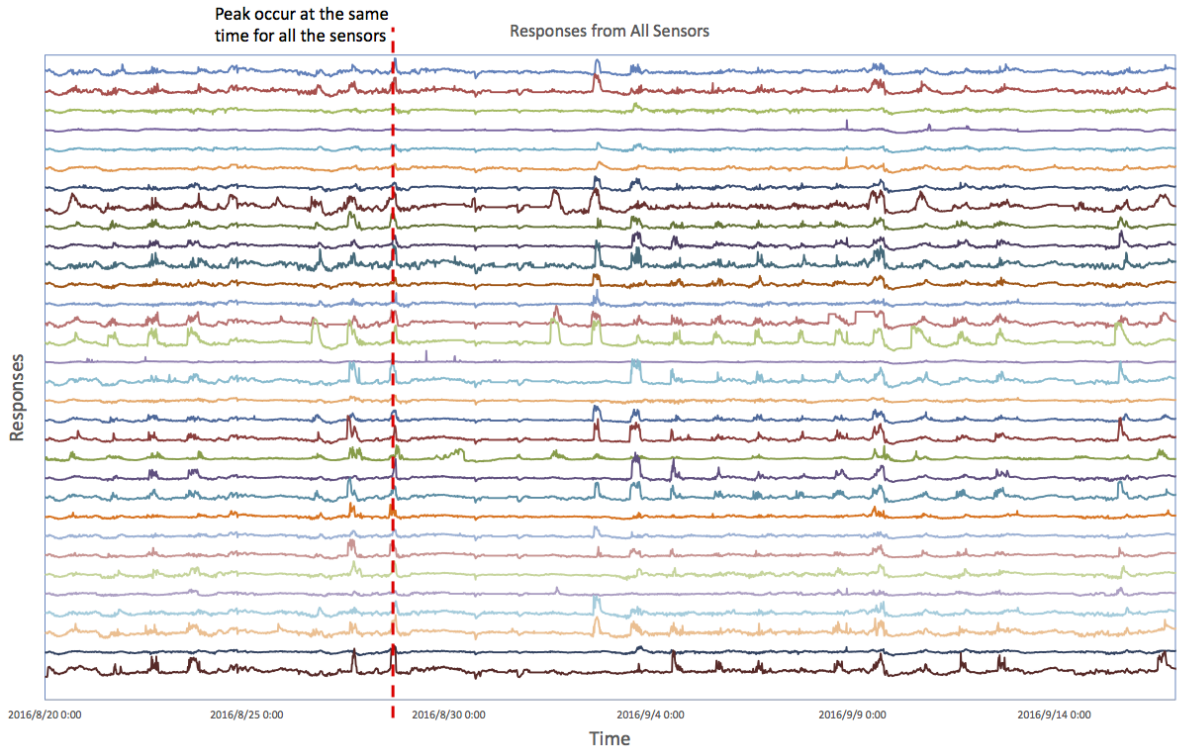


Figure 2.28: A comparison of all methane sensors at the Farnsworth field site for 7 months. The peaks show higher detected levels of gas concentration. These peaks seem to occur at similar time periods for all of the sensors.

2.6.3 Data Filtering

Since the results between sensors were so similar then individual sensor performance did not need to be investigated. The results were averaged and plotted as a histogram in Fig. 2.29. Plotting the values as a histogram shows how the data is distributed and highlights measurements that are much higher than the mean. These regions of higher concentration are called events.

The plot in Fig. 2.29 shows that the results can be treated as a normal distribution. An iterative data processing algorithm was used to find the mean and standard deviation for both the carbon dioxide and methane measurements. Measurements that were continuously higher than two standard deviations from the mean for at least two hours were identified. Two continuous hours were chosen to ensure that there was not a single measurement that triggered an event. This meant that at least eight measurements were taken at these elevated values. Two standard deviations were chosen to ensure the values were significantly different than the mean with a high confidence interval. The two standard deviations metric provides a confidence interval of 95%. An example of the affect of filtering is shown in Appendix 3.

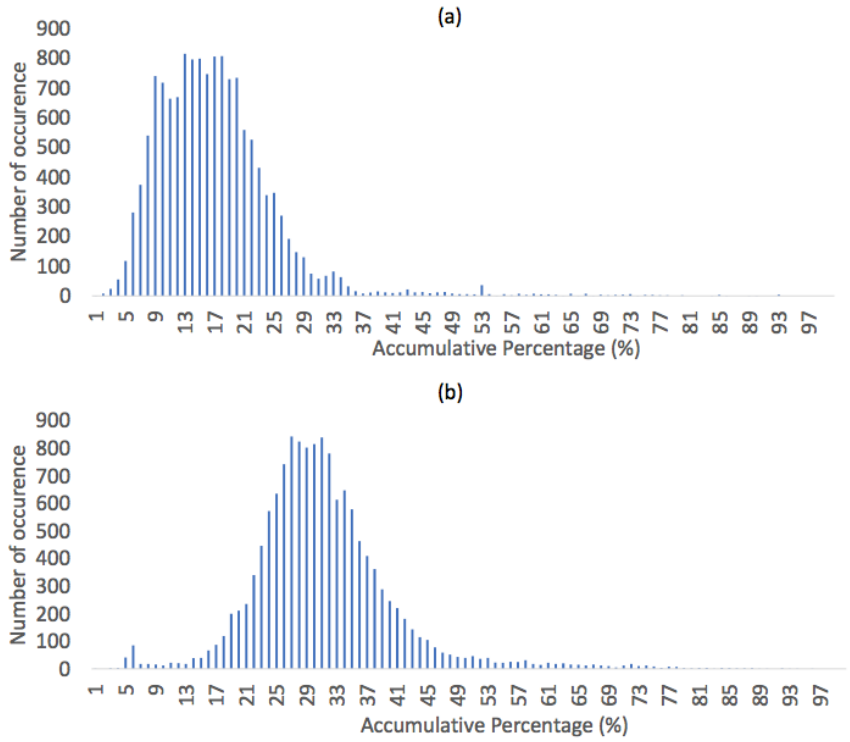


Figure 2.29: A concentration distribution histogram of (a) carbon dioxide and (b) methane from the Farnsworth, TX field site.

Other values for the time period and variance from the mean could have been used, but these values were chosen as they were reasonable, could be defended, and still, show the validity of the sensor network. Results from other time periods and confidence intervals are shown in Table 2.7. The analysis shows that thirty-three carbon dioxide events and thirty-five methane events were identified with the chosen ranges.

Table 2.7: Possible number of events that occur under different choices of gas concentration of duration between Aug. 20th, 2016 and Jan. 13th, 2017.

	Number of Carbon Dioxide Events			Number of Methane Events		
$x\,(\sigma\,(daily))$	y =1 hrs	y =2 hrs	y = 3 hrs	y =1 hrs	y =2 hrs	y = 3 hrs
1	117	78	59	164	105	75
2	46	33	26	60	35	25

2.6.4 Event details

Figure 2.30 shows the magnitude and duration of the carbon dioxide and methane events that met the event criteria set out above. When no events are detected, it shows a horizontal line at zero. Also, these plots have been placed above one another so that the values could be compared. Although a similar number of events were detected for carbon dioxide and methane, Fig. 2.30 shows that these events do not occur simultaneously. Several events for both carbon dioxide and methane are shown in Fig. 2.31. Additional examples of CO_2 and CH_4 data are shown in Appendix 4.

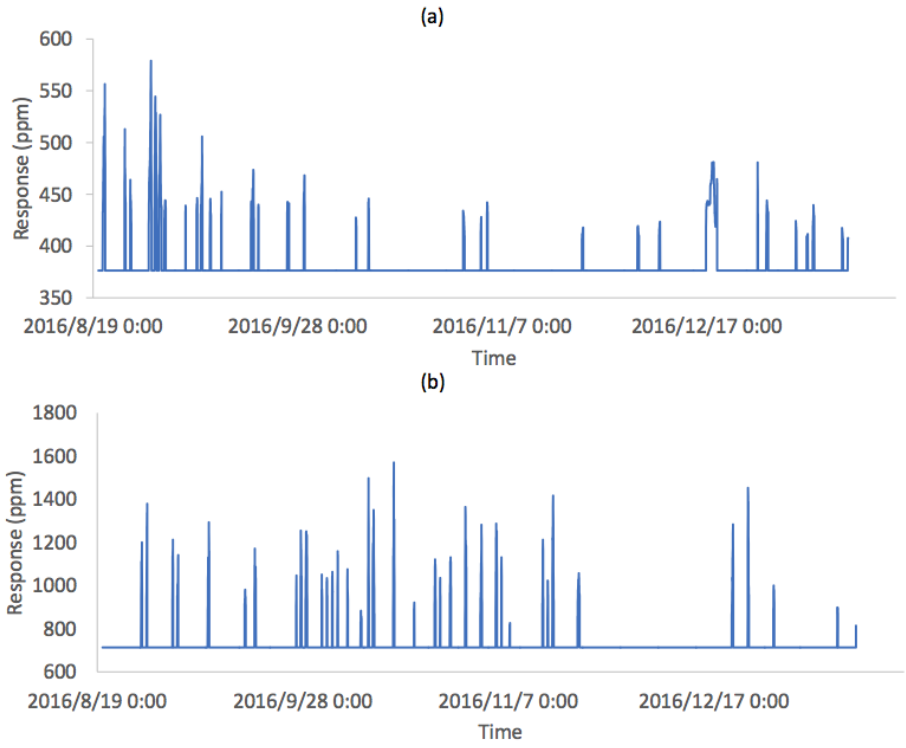


Figure 2.30: Identified carbon dioxide and methane events shown over time. The events are not shown to occur simultaneously.

2.6.5 Significance of the Findings

These results show that the sensor network is capable of reliably making a measurement of carbon dioxide and methane in the field. Furthermore, these measurements can be used with reasonable criteria to identify statistically significant time periods of elevated measurements or events. Given a temporal resolution of fifteen minutes this study showed that there was

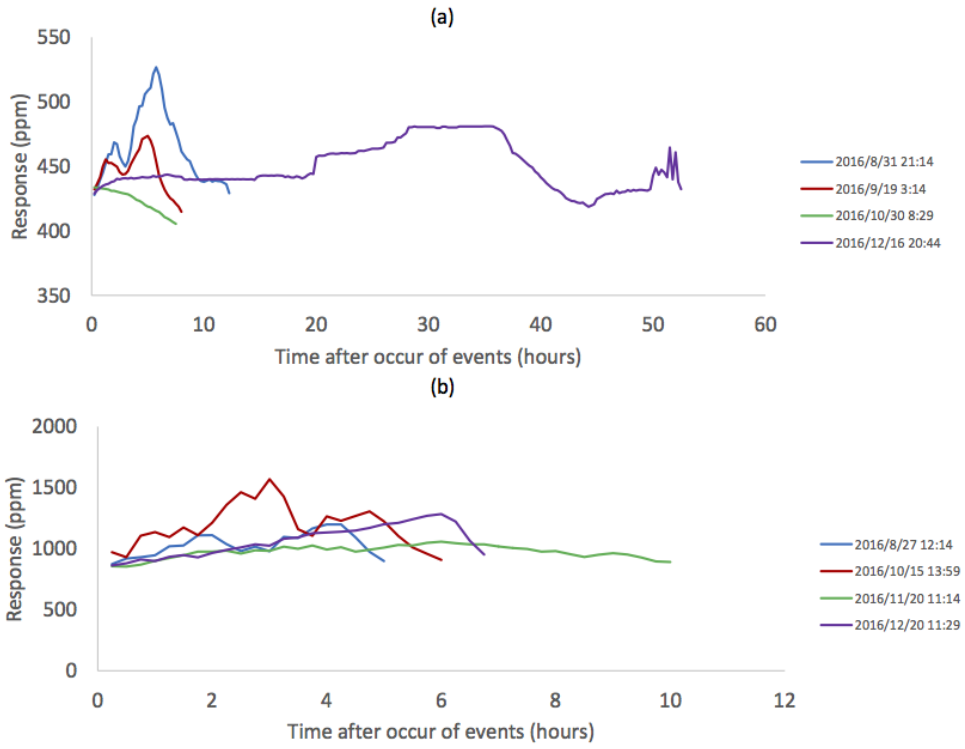


Figure 2.31: The concentration change versus time for several typical (a) carbon dioxide and (b) methane events.

not a significant difference in the spatial measurement of carbon dioxide or methane at the Farnsworth site. Also, the carbon dioxide and methane events did not correspond to one another.

Studying the events in more detail it can be seen that the event durations were typically less than six hours. This suggests that whatever is causing the release of gas is not constant. It indicates that the detected events are related more to a one-time occurrence such as a malfunctioning piece of equipment or a leak at the injection site. By changing the frequency at which data is collected, it may be possible to get a better fix on the source of the gas discharges.

3 UAV Design, Evaluation, and Deployment

Overview

The team evaluated concepts for possible use in the airborne monitoring portion of the project. This included both COTS and custom platforms, as well as custom and COTS autopilots. While a final selection was not made in either category due to limited details on the sensor requirements, tests were performed using readily available systems to determine requirements and limitations of possible flight plans. Fixed wing and rotary wing platforms were both evaluated for use and are discussed in turn below. In short, while a fixed wing vehicle was determined to be the best option for large surveys, a rotary wing system was developed to perform quick profiles and evaluate with greater ease and safety variations in CO_2 as a function of altitude, since they are much easier to fly near the ground.

Oklahoma State University's Unmanned Aircraft Flight Station (UAFS) is used for flight-testing of vehicles developed in the program. The flight station is located 11 miles east of Stillwater Oklahoma and offers 1 square mile of flight test area over Oklahoma State owned pasture with a 600 ft north south runway, and a smaller NE to SW runway. The facility has two hangers, one for storage and readying of aircraft and systems, the other is used to house ground control stations and test hardware in an air conditioned environment.

All flights were conducted in accordance with FAA requirements and Certificate of Authorization Applications (COAs) were obtained for flights as necessary. This included receipt of a blanket COA by the FAA in April 2016 to allow flights under 400 ft. in Class E airspace and flights under Part 107 rules starting in October 2016. Prior to that date, permission for flights were only approved at OSU sites in and around Stillwater, OK. Guidelines and best practices for FAA COAs for small-unmanned aircraft were followed. [Elston et al.] These include the OSU Unmanned Aircraft Flight Station (UAFS) for basic research and flight testing, the Marena site, and the DOE ARM SGP site located in Lamont Oklahoma. The latter two sites include dedicated diagnostics that log detailed meteorological and atmospheric information useful for system performance assessment and data analysis. A COA was obtained for the Farnsworth site in 2016 enabling flights during the day under visual meteorological conditions (VMC) and within visual line of site (LOS). In 2017 this COA was amended to allow night flights as well.

Technical Description

System Development and Testing To determine the efficacy and response time of airborne chemical traces, a simple sensing system has been developed to help guide the UAV tasks. The data collection system for the CO_2 detection is centered around a Sensair K30FR CO_2 sensor; a low-cost non-dispersive infrared sensor capable of a measurement range of 0-5000 ppm CO_2 with an accuracy of 3% and a response time of 2 seconds. The data is recorded to an onboard SD card by an Arduino microcontroller. The system measures the CO_2 levels as well as the corresponding GPS coordinates, altitude, temperature, humidity, pressure, and flow rate. It is a fully self-contained system powered by a LiPo battery. The system is shown in Figure 3.1. A test of the system measuring CO_2 levels over a controlled burn are shown in Figure 3.2. Note the elevated CO_2 levels over the burn area.

The data collection system was installed in a COTS aircraft for initial testing to keep the system simple and cost low. The aircraft is a Skywalker X-8 as shown in Figure 3.3; the hand launched UAV which was equipped with the data system and enough power for up to an hour of flight. The aircraft was controlled by a Pixhawk autopilot system that was programmed to search for leaks in a 'lawnmower' search pattern over the desired area. A full sweep was done at several altitude intervals to locate and characterize the leak. Final instrument integration is shown in Figure 3.4.

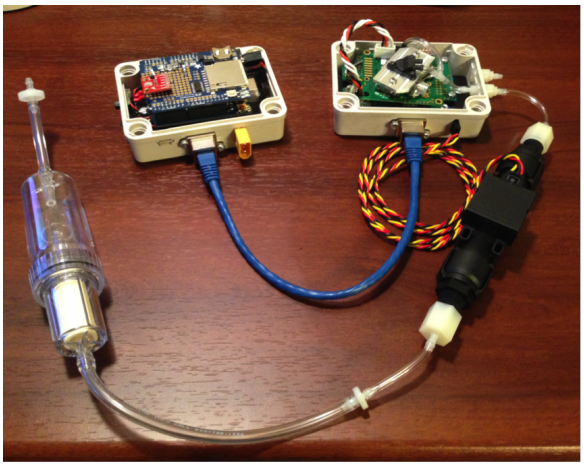


Figure 3.1: Airborne CO_2 sensor test bed.

The CO_2 detection system is show in 3.1 . The air sample flows in to an intake tube at the front of the aircraft at a maximum rate of 7.4 L/Min. This high flow rate allows a quick response time (¡2 sec) by the CO_2 sensor. The intake the air travels through particle and water filters and then a flow meter before entering the sensor box containing the CO_2 sensor. The sensor box contains sensors to provide the temperature, humidity, and pressure of the air

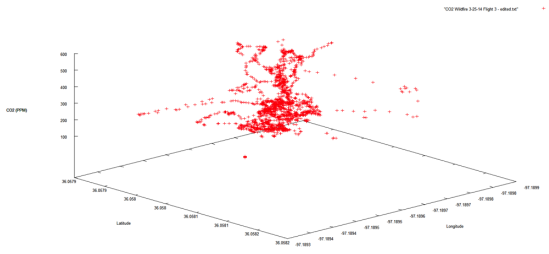


Figure 3.2: Plot of CO_2 levels and their location as measured over a controlled burn.

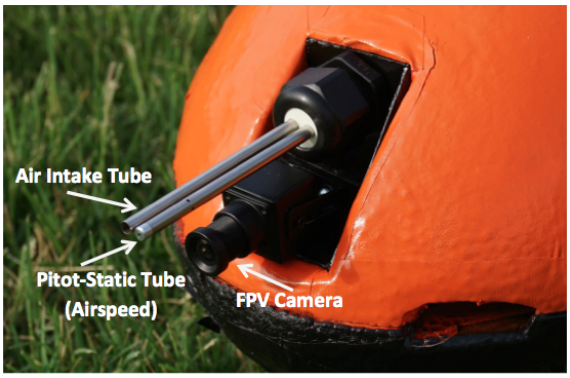


Figure 3.3: Sensor intake.

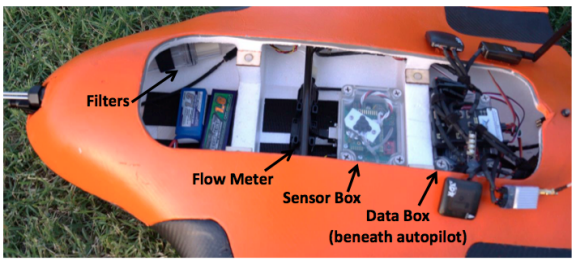
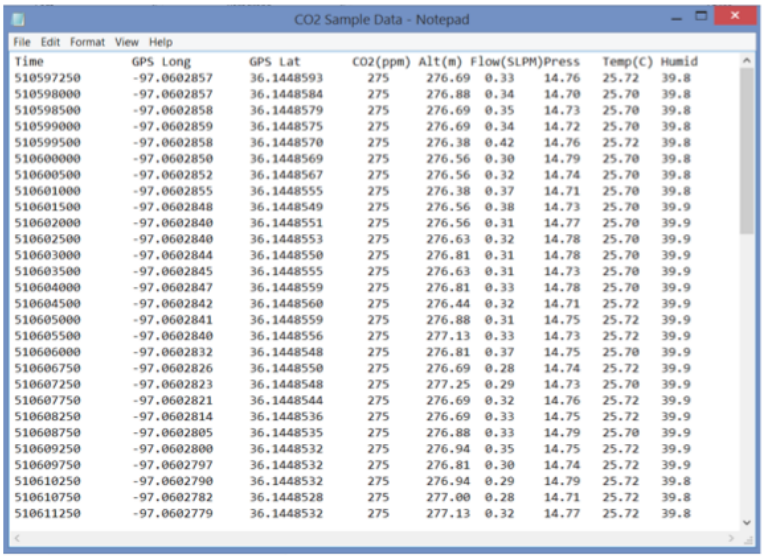


Figure 3.4: Instrument integration.

flow. The sensor box connects through ethernet cable to the data box containing the Arduino microcontroller, SD card logger, altimeter, and connections to the GPS and battery. This allows the system to independently operate inside or outside of the aircraft and log the time, GPS location, CO_2 level, altitude, flow rate, pressure, temperature, and humidity; the system logs at half second intervals (2 Hz) but this is user selectable.

The autopilot used on the UAS is a Pixhawk running the ArduPlane control software. It is wired to a GPS, an airspeed sensor with a pitot-static tube, a wireless first-person video (FPV) system for the pilot to watch while flying, and a wireless telemetry uplink so the flight plans and settings can be changed while in the air. The autopilot system was first tested in a smaller aircraft to test its reliability and make sure the programming was accurate. This aircraft was programmed to fly a similar flight plan as would be used in testing and performed well enough to be integrated into the final aircraft.


CO2 Sample Data - Notepad

File Edit Format View Help

Time	GPS Long	GPS Lat	CO2(ppm)	Alt(m)	Flow(SLPM)	Press	Temp(C)	Humid
510597250	-97.0602857	36.1448593	275	276.69	0.33	14.76	25.72	39.8
510598000	-97.0602857	36.1448584	275	276.88	0.34	14.70	25.70	39.8
510598500	-97.0602858	36.1448579	275	276.69	0.35	14.73	25.70	39.8
510599000	-97.0602859	36.1448575	275	276.69	0.34	14.72	25.70	39.8
510599500	-97.0602858	36.1448570	275	276.38	0.42	14.76	25.72	39.8
510600000	-97.0602850	36.1448569	275	276.56	0.30	14.79	25.70	39.8
510600500	-97.0602852	36.1448567	275	276.56	0.32	14.74	25.70	39.8
510601000	-97.0602855	36.1448555	275	276.38	0.37	14.71	25.70	39.8
510601500	-97.0602848	36.1448549	275	276.56	0.38	14.73	25.70	39.9
510602000	-97.0602840	36.1448551	275	276.56	0.31	14.77	25.70	39.9
510602500	-97.0602840	36.1448553	275	276.63	0.32	14.78	25.70	39.9
510603000	-97.0602844	36.1448550	275	276.81	0.31	14.78	25.70	39.9
510603500	-97.0602845	36.1448555	275	276.63	0.31	14.73	25.70	39.9
510604000	-97.0602847	36.1448559	275	276.81	0.33	14.78	25.70	39.9
510604500	-97.0602842	36.1448560	275	276.44	0.32	14.71	25.72	39.9
510605000	-97.0602841	36.1448559	275	276.88	0.31	14.75	25.72	39.9
510605500	-97.0602840	36.1448556	275	277.13	0.33	14.73	25.72	39.9
510606000	-97.0602832	36.1448548	275	276.81	0.37	14.75	25.70	39.9
510606750	-97.0602826	36.1448550	275	276.69	0.28	14.74	25.72	39.9
510607250	-97.0602823	36.1448548	275	277.25	0.29	14.73	25.70	39.9
510607750	-97.0602821	36.1448544	275	276.69	0.32	14.76	25.72	39.9
510608250	-97.0602814	36.1448536	275	276.69	0.33	14.75	25.72	39.9
510608750	-97.0602805	36.1448535	275	276.88	0.33	14.79	25.70	39.9
510609250	-97.0602800	36.1448532	275	276.94	0.35	14.75	25.72	39.9
510609750	-97.0602797	36.1448532	275	276.81	0.30	14.74	25.72	39.9
510610250	-97.0602790	36.1448532	275	276.94	0.29	14.79	25.72	39.8
510610750	-97.0602782	36.1448528	275	277.00	0.28	14.71	25.72	39.8
510611250	-97.0602779	36.1448532	275	277.13	0.32	14.77	25.72	39.8

Figure 3.5: Sample data.

The selected CO_2 sensor (Senseair K-30) matches the manufacturer's specifications and is one of the best sensors available considering its high accuracy, low cost, and small size. Further ground testing was performed to quantify the performance characteristics of the sensor system that could affect its in-flight results. First, both the CO_2 sensor and flow meter were calibrated. Tests were then run to determine the sensor's response time with the intake system attached (the response time of the sensor itself has already been verified). The fuselage of the aircraft was also be placed in the wind tunnel at OSU to determine the in-flight flow rate, perform more response time tests, and calibrate the airspeed sensor.

The first objective of the initial test flights were to simply evaluate the aircraft's handling qualities in its final configuration as well as the autopilot's performance characteristics in

Figure 3.6: Final vehicle 1.

this airframe. The second objective for the test was to perform the first of several aerial measurements of the CO_2 levels at the OSU flight field to establish a baseline to compare against during future CO_2 emissions testing. The aircraft flys in a simple 50m grid pattern over the test area to collect CO_2 data points; this pattern is flown at increasing altitudes in 50m increments. Tighter and coarser grids were also evaluated to see which pattern the autopilot can adhere to best while still providing the most detailed measurements. Following this test, additional test flights were flown to further test the reliability of the system and provide background CO_2 measurements at various times during the day.

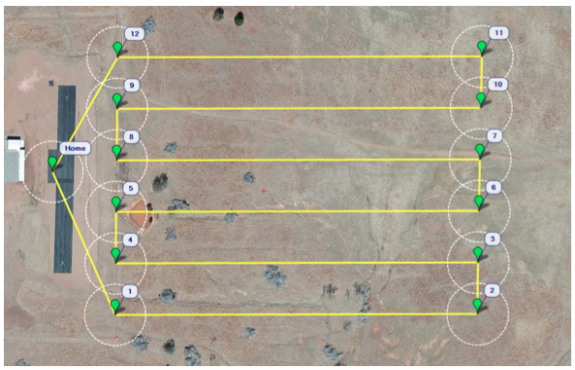


Figure 3.7: Proposed flight path at test site.

To improve system capability, the aircraft was retrofitted with landing gear. This increases payload weight and keeps the sensors protected during takeoff and landing. The system is shown in Figure 3.9. Flight pattern testing has continued, which has included testing waypoint navigation and PID tuning of the flight controller. Sample 'lawnmower' search pattern flight plans are shown in Figures 3.10a and 3.10b; the latter also shows the corresponding vehicle

Figure 3.8: UAV flight tests.

trajectories. Flight paths are overlapped to increase the coverage over the search area while still allowing large turns for the controller. Figure 3.11a shows a sample orbit pattern with increasing diameter to determine the viability of using a spiral pattern with increasing altitude for detection. Results are shown in Figure 3.11b. The smallest diameter orbit is clearly not achievable with the current system, but the larger two orbits are. This provides a lower limit for the orbit diameter for the current system.

Figure 3.9: Airborne CO_2 sensor test bed with landing gear.

A test of the system measuring CO_2 levels over a controlled release are shown in Figure 3.12. Note the elevated CO_2 levels over the runway where the release was made. The air sample flows in to an intake tube at the front of the aircraft at a maximum rate of 7.4 L/Min. This high flow rate allows a quick response time (¡2 sec) by the CO_2 sensor. The release was made using a 16 g CO_2 completely vented in under 5 s. The Senseair K-30 sensor clearly picks up the plume the point of release, but the maximum CO_2 level of approximately 3,500

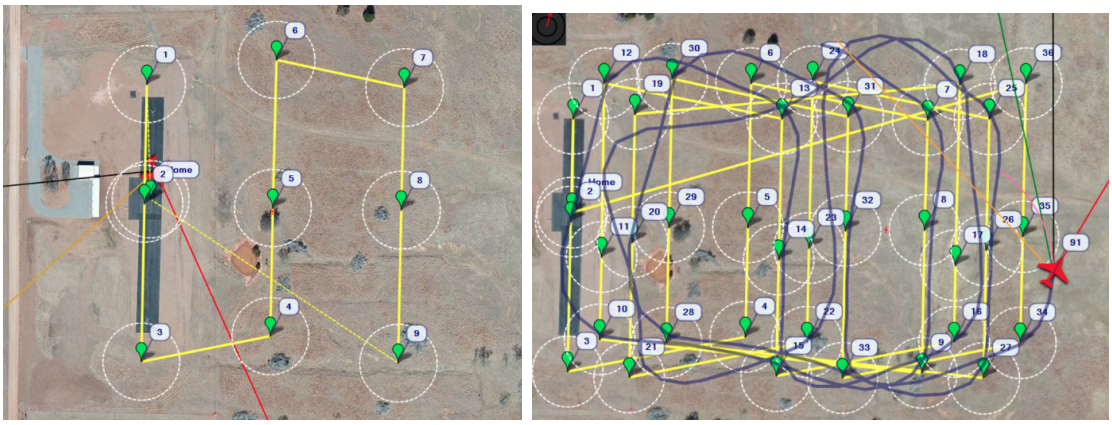


Figure 3.10: Search patterns and tests.

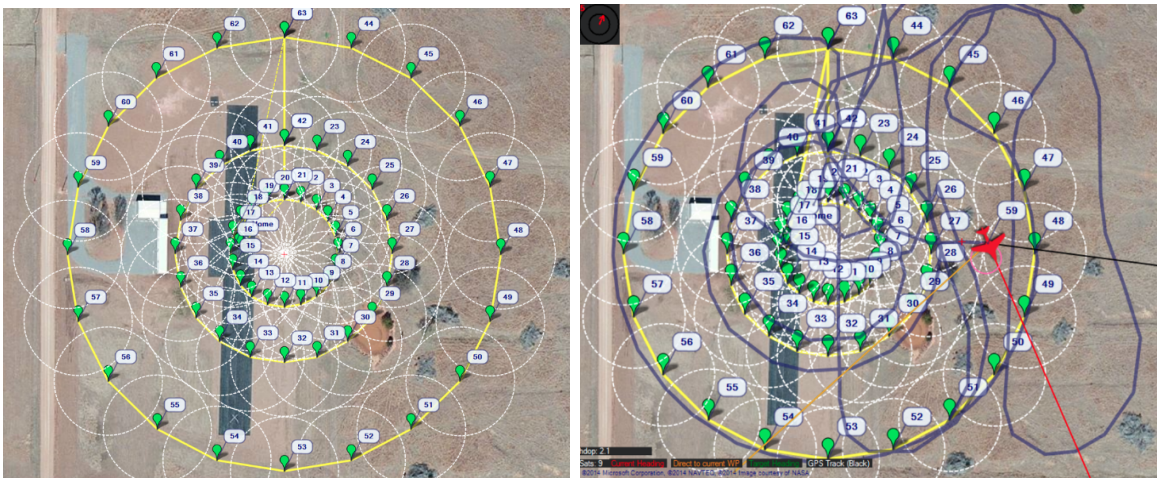


Figure 3.11: Orbit diameter waypoints and corresponding test.

ppm is not reached until approximately 10 s later well downstream of the release. Thus, the sensor delay needs to be accounted for as tests continue. Calibration of this system in the wind tunnel is ongoing to address this issue.

Calibration and Validation The first test performed in the lab was to calibrate the K-30 CO_2 sensor. To do this a CO_2 - N_2 mix was pumped through the K-30 as well as a California Analytical Instruments ZRE NDIR Gas Analyzer. The ZRE has a repeatability of 0.1% of full scale. The CO_2 mixture was pumped through the sensors at seven levels increasing from 0 to 1308 PPM CO_2 . The mixture flowed through the sensors at a rate of roughly 1.1 liter/min and was held at each point for ten minutes while recording at 2 Hz. From this a relationship was developed between the K-30's levels and that of the calibrated ZRE as shown in Figure 3.13; one standard deviation for each point is 1 PPM so the error bars are not visible. As can be seen, the K-30 output is linear, but requires a correction factor to adjust to the correct

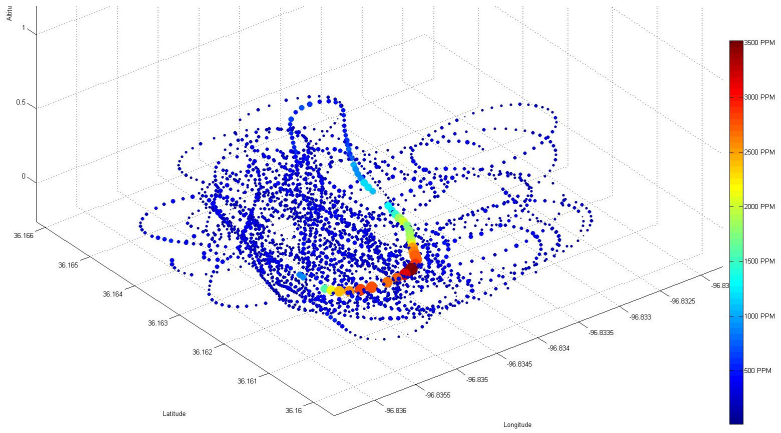


Figure 3.12: Sensor test with CO_2 source.

CO_2 value. The DAQ has been programmed to output all of the data at a rate of 4 Hz to the SD card as a text file.

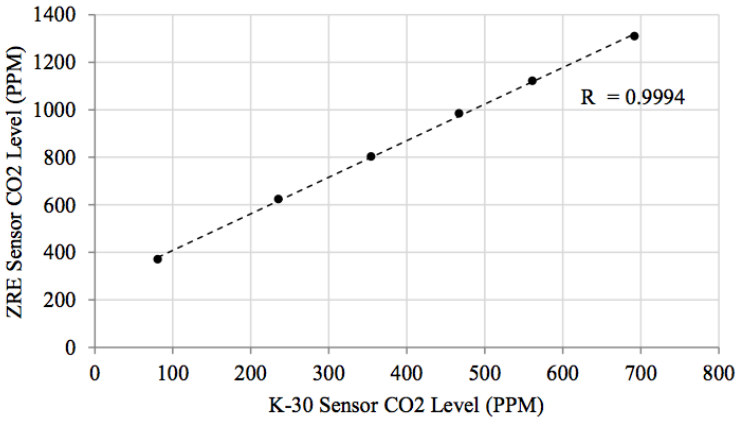


Figure 3.13: Sensor calibration.

Figure 3.14 shows a schematic of the payload layout in its final form. The air intake tube is integrated next to the Pitot-static tube and the exhaust port is located by the motor for reduced static pressure to promote additional air flow into the meter. Wind tunnel tests were next performed to determine how the flow rate through the sensor system varies with airspeed as well as what configuration of the sensor system provides the best flow rate. The aircraft was placed in a 3x3 test section of OSU's low speed wind tunnel and the onboard flow rate sensor was attached to a NI 6009 DAQ so the flowmeter and tunnel speed voltages could be recorded in LabView. Data was recorded at 1 kHz in two five- second runs per data point. The aircraft flies at a cruise speed of around 20 m/s so data points were collected over a range from 10 m/s to 28 m/s to cover typical flight speeds that may be encountered by the aircraft.

To determine the configuration of components in the flow system that provide the optimal flow rate, components were reconfigured between tests to see what impact each major component had on the flow. The intake tube, the 150 micron particle filter, and Honeywell flowmeter were present in all of the tests. First, as opposed to the air exiting into the fuselage of the aircraft, tests were run to see what effect running the outflow air through an expanding funnel located at a blunt area aft of the fuselage (a low-pressure region) would have on increasing the flow rate. The motor on the aircraft was also run during a series of tests to see its impact on the low-pressure region. Next, tests were run to observe how the flow rate changes when the water trap/hydrophobic filter was removed which is preferred for flights in high humidity but not always necessary. Finally, for comparison, a test was run with no filter, funnel, or even CO_2 sensor system to determine what the unimpeded (except the flowmeter and non-removable particle filter) flow rate would be. Figure 3.15 gives the results of these tests grouped by configuration.

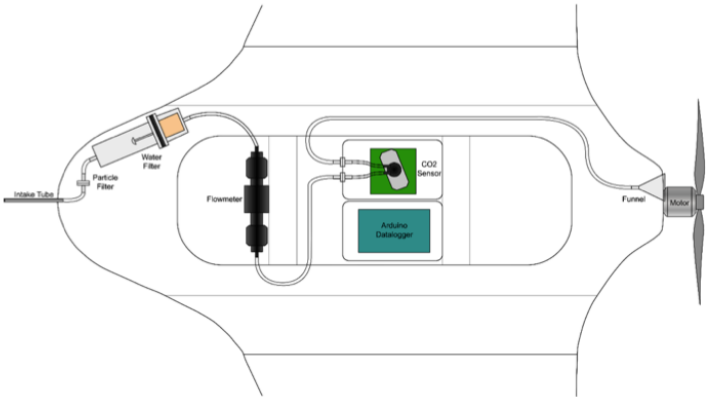


Figure 3.14: System configuration showing intake and exhaust port.

The wind tunnel test results show how strongly the sensor and system reduce the flow coming in to the aircraft. Seeing as the CO_2 sensor is a necessary component of the system, however, removing the water filter is the second best option for improving the flow rate of the system; additionally it avoids flow mixing that could occur in the water trap chamber and potentially reduce the quality of the readings. The exit funnel also has a noticeable increase on the flow rate, so it is incorporated in to the final aircraft configuration as well. Having the aircraft motor on reduces the effectiveness of the funnel somewhat, likely due to adding energy to the flow and reducing the separation that creates the low-pressure suction on the funnel.

The next tests performed in the wind tunnel focused on determining how the response time of the sensor was affected by the flow rate through the system. The plane was fixed in the wind tunnel and CO_2 was pumped in through a pitot tube mounted 17 inches in front of the plane intake. The gas was released for one second at a rate of 30 SCFH from a reservoir

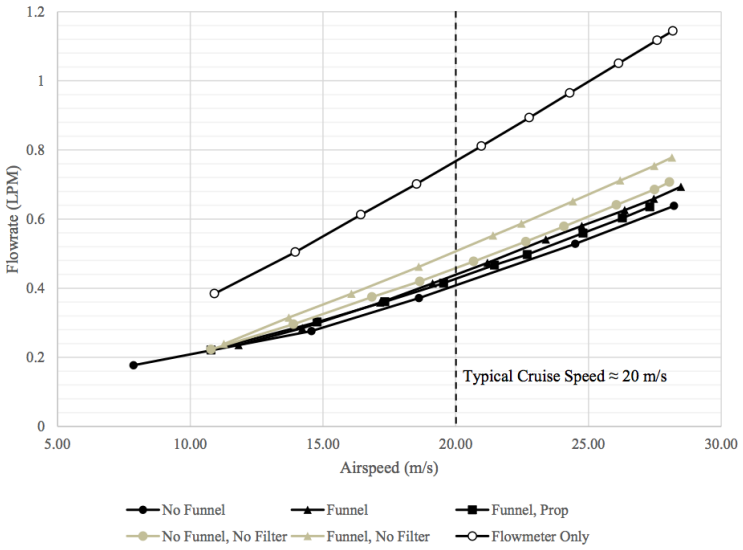


Figure 3.15: Flow rate as a result of airspeed and configuration.

of pure CO_2 . The valve was controlled by a LabView program which would record the CO_2 sensor values, flow rate, and airspeed at a rate of 6 Hz; the gas would be released every one minute and fifteen seconds to allow enough time for the sensor to detect the release and return to its baseline levels. Figure 3.16 shows a sample of two runs of the response test.

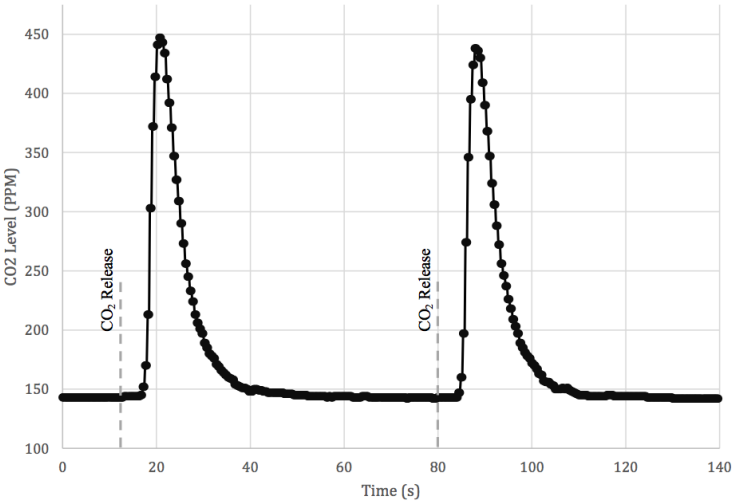


Figure 3.16: Sample CO_2 sensor response test at 0.435 LPM flow rate; 14.25 m/s airspeed.

These tests were run twenty times (twenty gas releases) per speed over the same airspeed range as the previous airspeed tests. All of the tests were run with what was decided to be the final flight configuration of the CO_2 sensor: the intake tube, particle filter, flow meter,

CO_2 sensor, and exit funnel; the water trap/hydrophobic filter was not present in the system. The time between the initial CO_2 release and when the sensor CO_2 level first began to rise was recorded as well as the time between the CO_2 release and the CO_2 peak level. This data was averaged and graphed in Figure 3.17 to observe the relationship between the flow rate through the CO_2 system and the response time of CO_2 detection by the system. The data shows that while the response time is not ideal for instantly detecting the plume, it is small enough to be manageable with some analysis of the flight data.

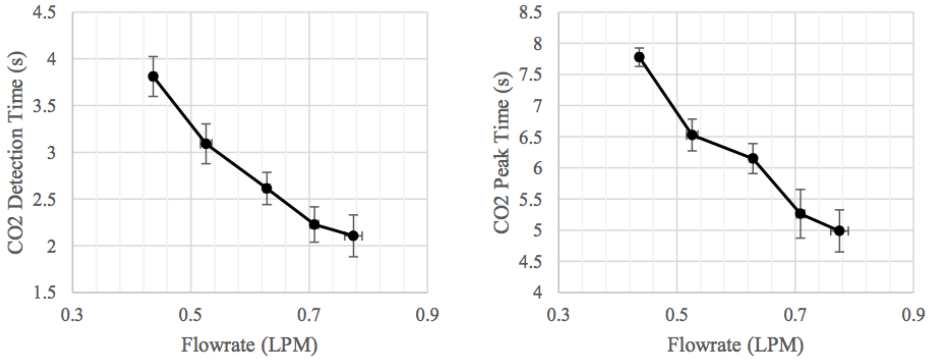


Figure 3.17: CO_2 detection and peak times based on flow rate.

Initially the model was controlled manually to ensure that control surfaces were correctly programmed and working as intended. Center of gravity and engine specifications were also verified. Once the aircraft was capable of manual trimmed flight the autopilot was tuned for autonomous flight in the simulator. Due to the increasing popularity of UAVs with hobbyists, initial PIDs were able to be obtained from DIY forums where others using X-8 with DIY autopilots have discussed the airframe's required tuning. From there, additional tuning was needed due to the more expansive payload required for this study. At this point the model handled well in manual flight and autonomous flight simulation was attempted.

For measuring wind speeds during the study, a Modern Device wind sensor (rev p) was acquired; it is shown in Figure 3.18. The sensor was obtained with the intention that it could act as a low cost variation of a hot wire probe. The wind sensor costs only tens of dollars compared to hundreds of dollars for the average hot wire probe. Additionally, the wind sensor is more robust than its counterpart allowing it to be airframe mounted which would likely break the hot wire in the event of a hard or grass landing.

The sensor performs similarly to a hot wire probe having an element that is heated to a programmed amount requiring a voltage to keep it at that temperature. As airflow passes over the heated element higher voltage is used to keep the element at the original temperature. This voltage change was calibrated with known velocities in a wind tunnel. The wind sensor calibrated data was compared with a hot wire probe. Figure 3.19 shows the results of each calibration curve. Both output voltages in the range of 1.25 volts to 2.5 volts for velocities up 30 mph. The sensor runs off of a regulated 5 volt power supply and outputs voltages up

to 3.3 volts. Based on these power requirements, robustness, and quality of data, the wind sensor was determined to be a suitable piece of hardware for future flights of the system.

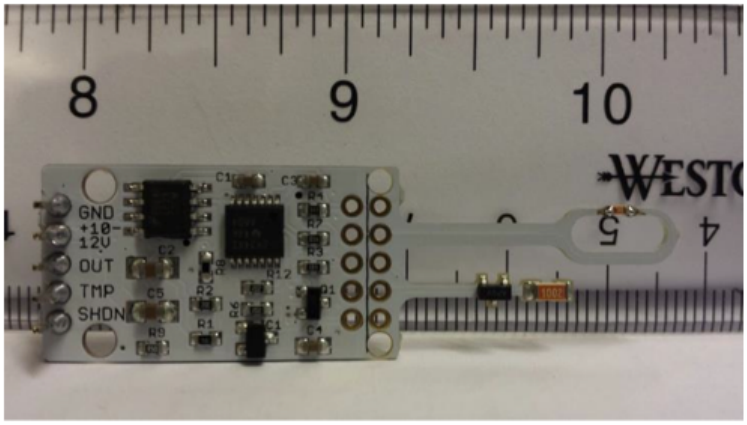


Figure 3.18: Modern Device Wind Sensor RevP..

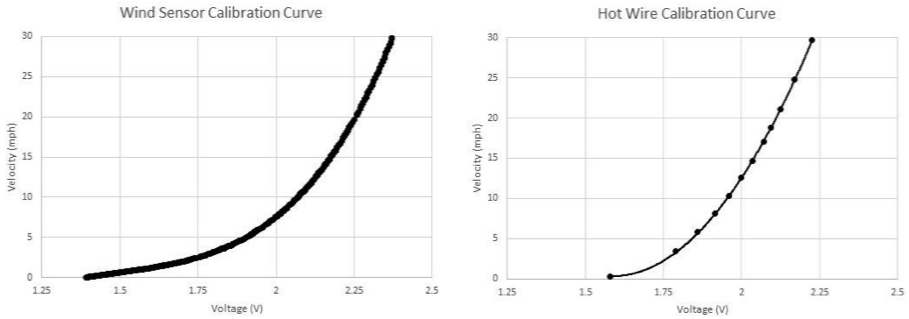


Figure 3.19: Wind Sensor and hot wire calibration curves.

CFD Simulation Initially the model was controlled manually to ensure that control surfaces were correctly programmed and working as intended. Center of gravity and engine specifications were also verified. Once the aircraft was capable of manual trimmed flight the autopilot was tuned for autonomous flight in the simulator. Due to the increasing popularity of UAVs with hobbyists, initial PIDs were able to be obtained from DIY forums where others using X-8 with DIY autopilots have discussed the airframe's required tuning. From there, additional tuning was needed due to the more expansive payload required for this study. At this point the model handled well in manual flight and autonomous flight simulation was attempted.

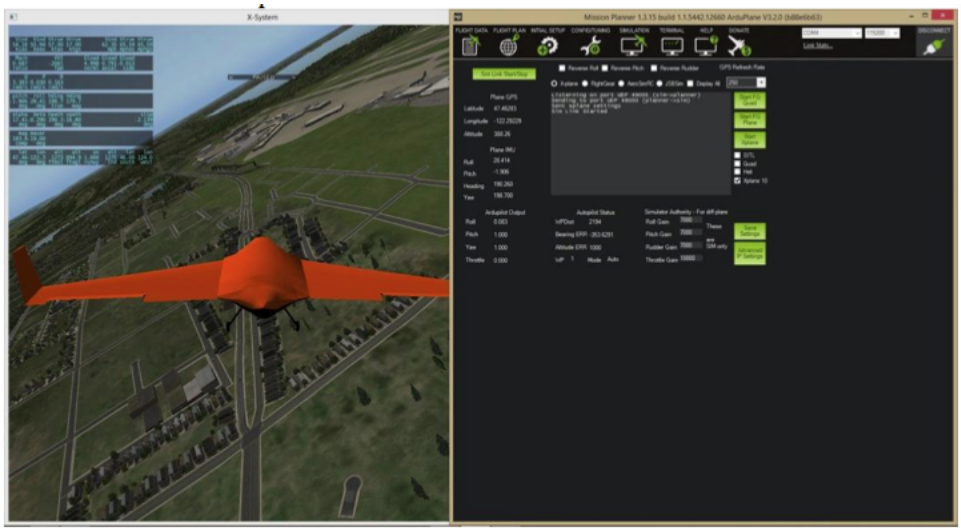


Figure 3.20: X-8 Autonomous flight simulation.

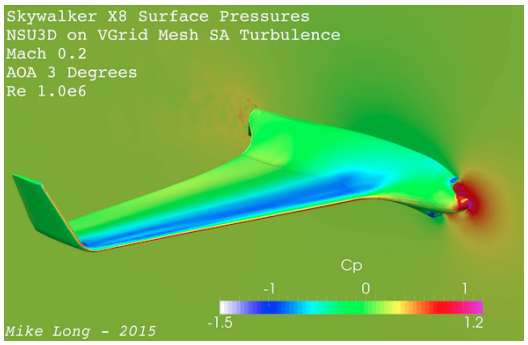


Figure 3.21: CFD simulation of flow over X-8 airframe for optimal sensor placemtn.

Figure 3.20 shows the X-8 simulation during an auto mode flight attempt. After multiple tunings and configurations of the model the automatic control through the simulation still struggled. It was determined that because the aircraft was a flying wing and only had elevons for control, difficulties were encountered in the Plane Maker modeling that only became apparent during the simulated automatic flight. The problem arose due to a lack of working elevon settings in Plane Maker which led to the use of ailerons and elevator being defined as the same control surface in Plane Maker. However, when programming the transceiver and receiver with the autopilot, attempts at both flying wing configuration and standard aileron plus elevator configuration were used to try to get the model to correctly perform as controlled. The only method of setup that worked in the controller was to setup the aircraft as a flying wing with two ailerons. For manual flight in the simulator this worked fine other than the fact that aileron inputs, while correctly banking the aircraft, were displayed as elevator inputs. So the aircraft would perform as controlled to do but appeared to be trying to pitch up or down depending on the commanded bank. Additionally the X-Plane software has increasing difficulty with smaller and smaller aircraft and weights. Due to the aircraft

being considerably lighter than the intended aircraft used in X-Plane the plane would very rarely flip and spin out of control before reestablishing correct flight. This could possibly be from simulated wakes of other aircraft in the program but its infrequent and short occurrence did not require deeper understanding of the issue. At this point the airframe would fly mostly stable with slight oscillations and poor altitude holding. Because it was known that X-8 aircraft were being flown with very similar tunings for DIY autopilots and the issues at hand being easily monitored in and corrected for real flight, this concluded the simulation testing.

Additional, CFD was used to evaluate optimal sensor placement over various airframe configurations. This allowed a relatively easy process to determine the impact of sensor location and airflow over the airframe on flow rate into the sensor. An example for the flying wing configuration is shown in Fig. 3.21, which shows pressure contours that provides input on best locations for sensor intake and exhaust utilizing natural pressure gradients.

Updated and Expanded System Due to weight limitations of the X-8 design, the system was upgraded to a larger platform. The UAS is based around the hand-launched Skyhunter aircraft. The autopilot has been upgraded to a Stabilis with a Pixhawk as a backup. The aircraft was extensively outfitted with video systems for additional safety and source observation. This included an infrared camera manufactured by DRS technologies mounted to the side of the aircraft, a FPV camera mounted to the nose, and a GoPro HD camera mounted to the nose hatch; all of these cameras were mounted at a downward angle for ground observation while in level flight. In addition, wireless video transmitters for the infrared and FPV were equipped as well as onboard digital video recorders for these two cameras.

To perform CO_2 tests during the fire flights the sensor system, including sensor box, data box, flowmeter, GPS, battery, strobe, and tubing had to be fitted in the narrow fuselage of the Skyhunter aircraft (with the designated call sign Firebird). The boxes were squeezed in tandem in the fuselage connected with a 0.5 m Ethernet cable. The flowmeter was placed atop the Ethernet cable with the battery nestled below and the strobe mounted beneath; leaving just enough room for the digital video recorders in the nose. The highest flow rate tubing configuration which could be fitted was equipped. This included the 0.135 inch ID intake tube with the 1 inch intake funnel attached, the particle filter (very important with the smoke and ash), no water filter, and a 1 inch exit funnel mounted on the trailing edge of the wing just between the aileron and edge of the prop arc. The entire system configuration is shown in Figure 3.23 and the final flight vehicle in Figure 3.24.

A custom ground control station (GCS) was developed to facilitate field tests. The unit is shown in Figure 3.25 and has the following characteristics:

- Field Transportable Communications Link

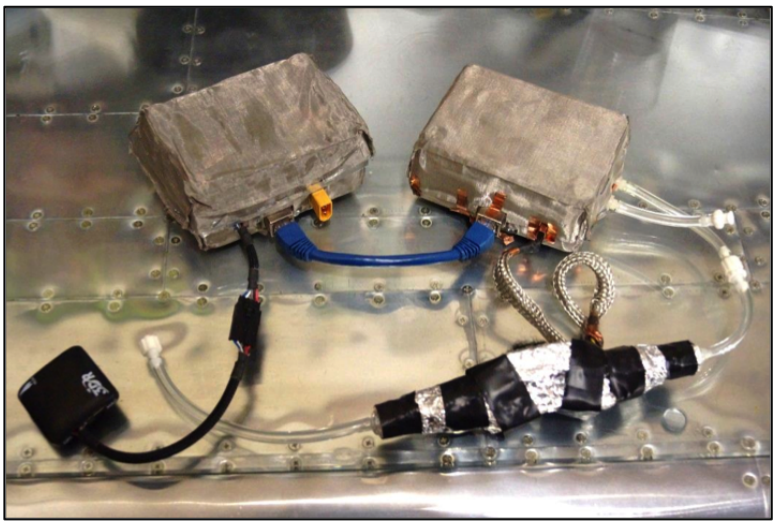


Figure 3.22: Final sensor setup.

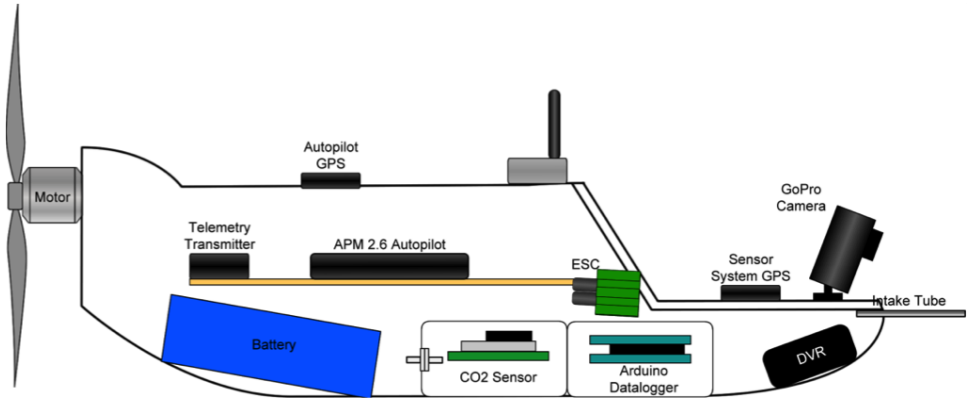


Figure 3.23: UAS payload and systems configuration.

- Pelican Case iM2590
- Custom Front Panel
- DVR Capture of Displays
- 120 VAC with distributed power (DC 12V)
- USB communication protocol
- Primary Display
- Semi-Rugged Panasonic Toughbook (CF53)
- Waypoint Navigation and Control
- 2 x 11 inch LCD displays

Figure 3.24: Skyhunter aircraft (with call sign Firebird).

- Attitude and Telemetry
- FPV streaming from aircraft

Figure 3.25: GCS.

Flight testing hours required for development have also been significantly reduced with the ability to throughly test a complete system in a HITL (hardware-in-the-loop) environment.

The autopilot is integrated into a flight ready aircraft complete with control surface actuators, on board power, data links, etc. Entire missions can be realistically simulated using custom designed X-Plane® flight models of the test aircraft. The HITL is capable of emulating several systems fault-scenarios and adverse environmental scenarios, including sensor and actuator failures, structural failure, probabilistic models of component wear, aerodynamic instabilities, and gusts/turbulence. A sample is shown in Figure 3.26.

Figure 3.26: GCS in HWIL simulation.

Plume Modeling for System Evaluation Standard plume models were used to evaluate results of the field tests. While it is known that natural plumes are not well represented by time-averaged models, standard plume models can be used to compare with the measured CO_2 concentrations at instantaneous points in time.Woods, 2010 There are multiple ways to model plumes, but a straightforward though not necessarily accurate method involves modeling the plume as a constant 3D source much in the same way as pollutants from a smokestack would be modeled. This type of modeling consists of two main components, diffusion and advection, as shown in Fig. 3.28. Diffusion is the tendency of concentrations of a gas or liquid to disperse over time. This is governed by Equation 1 as shown below which consists of the gradient of the product of the diffusivities (D) and spatial change of the concentration c as well as background velocity (v) and decay (λ).Modeling, 2012

$$\frac{\partial c}{\partial t} = \frac{\partial D_x}{\partial x}\frac{\partial c}{\partial x} + \frac{\partial D_y}{\partial y}\frac{\partial c}{\partial y} + \frac{\partial D_z}{\partial z}\frac{\partial c}{\partial z} - v\frac{\partial c}{\partial x} - \lambda c \tag{3.1}$$

When combined with advection, the movement due to the wind, and some decay, this produces a form of Gaussian model as provided in Equation 2 which is based upon a Gaussian distribution of the gas (as opposed to uniform distribution), where $\sigma = \sqrt{2Kx/U}$. K is the turbulent diffusivity, which is very difficult to accurately determine.

$$c(x,y) = \frac{Q}{\pi\nu\sigma_y\sigma_z} e^{-\frac{1}{2}\left(\frac{y^2}{\sigma_y^2}+\frac{H^2}{\sigma_z^2}\right)} \tag{3.2}$$

Q is the source strength and H the plume source height, which in this case may be the forest canopy height but will typically be zero for grassland fires. The standard deviations, σ_y and σ_z, are major components in determining the shape of the plume.It should be noted that this model is for a point source and thus is only valid when the plume is far from the source or much larger than the source footprint. An example of this distributions is shown in Fig. 3.27.

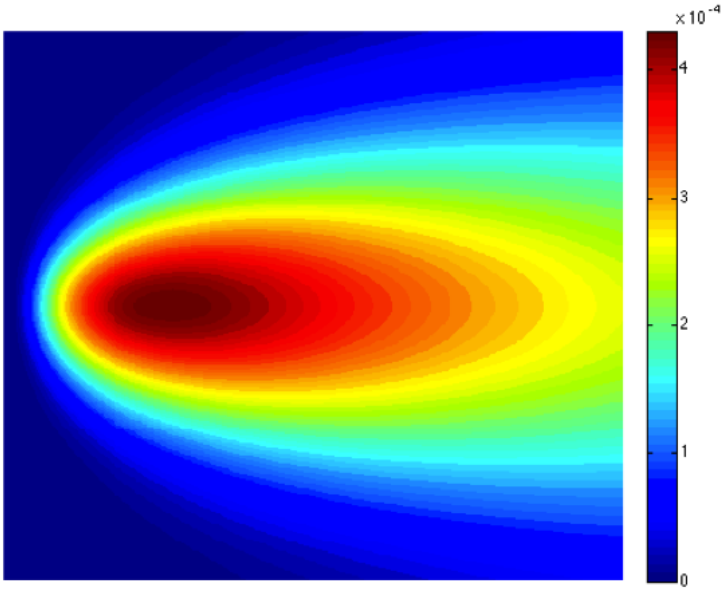


Figure 3.27: Sample scalar distribution from gaussian model.

As a system demonstration, the system was tested using a controlled burn on a small plot, shown in (Fig. 3.29). This type of controlled burn is limited to a small area on the ground, but the plume, laden with CO_2 , expands and advect with the prevailing winds. The measurement grid consisted of five north-south passes which were 2600 ft long to provide plenty of clean air between passes over the fire for the CO_2 levels detected to return to ambient and spaced 150 ft apart. Because the system could not turn tightly enough to make such close passes, the

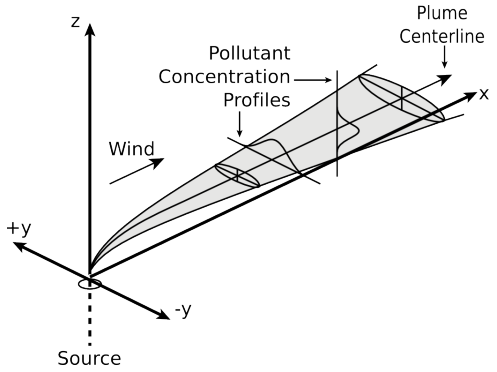


Figure 3.28: Gaussian plume.

measurement path includes interleaved passes. A plot on the south edge was burned before the test, but was only smoldering by the time of the measurement. The flow rate during this flight averaged around 0.5 LPM which gave a resulting response time of approximately 7 s, which is included in the data shown in Fig. 3.30a, with the burned plot highlighted in orange. The CO_2 levels very clearly increase when measured just south of the fire. The embedded filter removed small smoke particles and other particulates entering the sensor. This does not happen at all collected points, but this is likely due to the changing intensity of the fire, especially during the initial measurement points when the fire was just starting. For comparison, the model from HolzbecherModeling, 2012 was applied with the given wind speeds, a CO_2 diffusivity of 0.39 cm^2/s, the plume originating from the plot location, and the observation point at the same location as the sensor. This is shown in Fig. 3.3b. Since the points used to extra the Gaussian distribution is limited and is assumed to be a snapshot in time, the scaling values for the model do not necessarily match with the scale of the plot, so the plume model overlay should be simply taken as a reference visualization of the general plume path and not indicate a direct quantitative model of the CO_2 levels. Regardless, this can be extended to 3D and interpolated as a model of the full atmospheric plume, a portion of which is shown in Fig. 3.31.

Figure 3.29: Small plot burn used as surrogate CO_2 plume source.

For future analysis of this type, values of the CO_2 concentration can be compared with

other measured values, such as smoke intensity (using, for example, a pyrometer, chemical analysis, or in this case a direct visual measurement) and vertical velocity v as a measure of buoyancy and plume convection. Change in height from the barometric pressure sensor can be used to calculate the latter. This is shown in Fig. 3.32. Note that both smoke intensity and v appear to correlate well regions of increased CO_2 concentration. However, a thorough analysis has not been completed. In future tests for this use of the technology a 5-hole probe or variometer can be used for improved accuracy in determining v while dedicated light sensors can be used to provide a better assessment of smoke intensity.

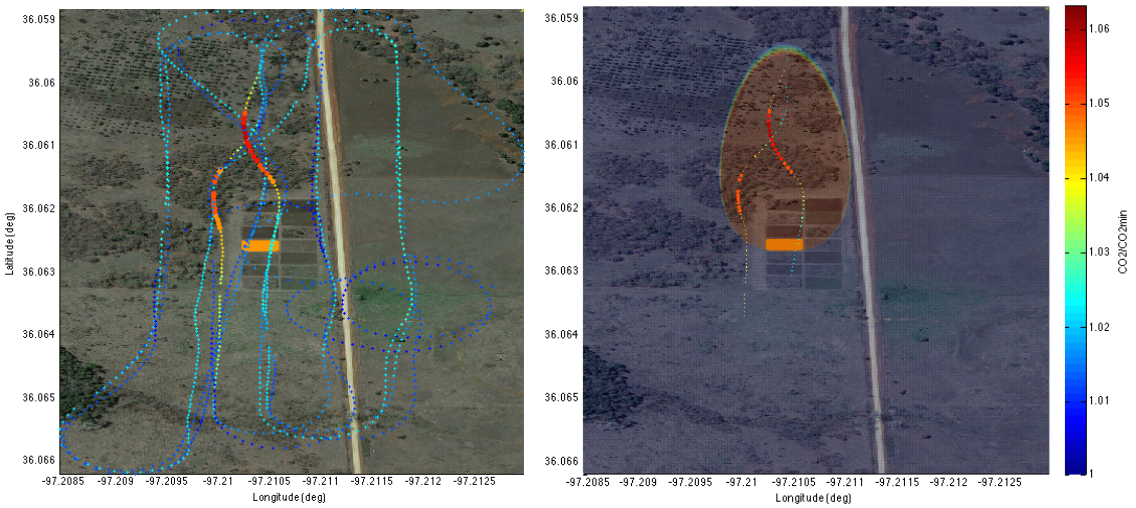


Figure 3.30: Sample CO_2 data with Gaussian plume model.

Attempts were made to increase the flow rate through passive methods such as varying the design of the inlet and exit configurations of the system. While these demonstrated a noticeable improvement, they still did not provide a flow rate with a high enough sensor response time. While a few seconds delay in peak detection time is short enough to shift the data to refine the location of the detected plume, the greater issue is the time it takes for the sensor to return to atmospheric background levels. This issue was highlighted best during controlled CO_2 release tests. In these tests remnants of the increased CO_2 air sample had not fully left the system, causing the CO_2 levels to simply increase with each subsequent pass through the plume. Further increasing the flow rate through the system would do much to alleviate this problem, but the required increase may be beyond the reasonable capabilities of passive improvements and it may be necessary to install a pumping system onboard the aircraft.

While the system is able to detect increased CO_2 levels, further testing and improvements are required to be able to accurately quantify the gas levels detected to better characterize the plume. This type of high accuracy measurement is necessary to detect small levels that may be but a few ppm above ambient but ensembled averages appear sufficient to assess variations

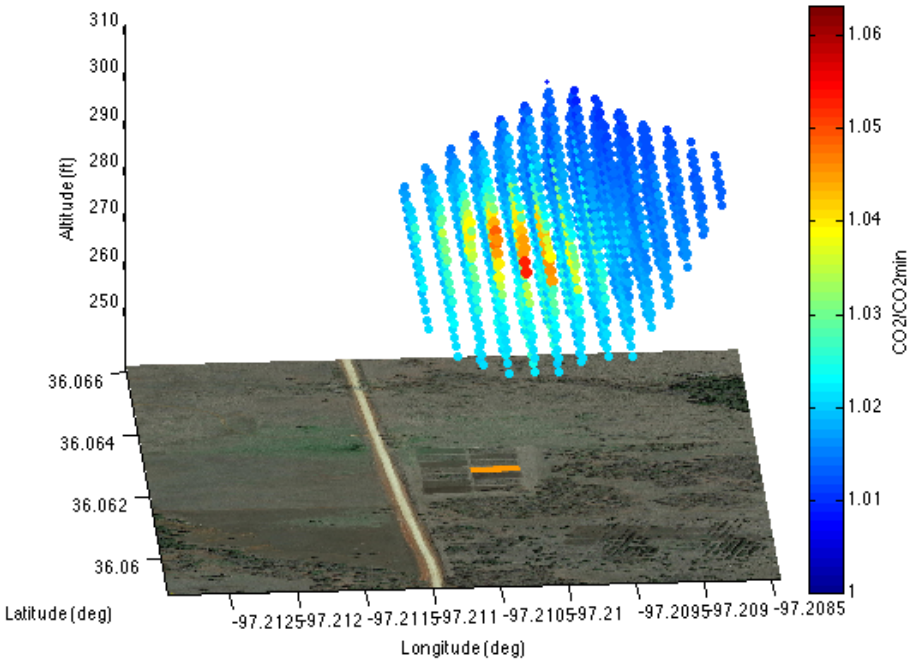


Figure 3.31: 3D plume model.

at this scale. Improved calibration techniques are a key part in attaining this level of accuracy, possibly including in-flight but at least pre- and post-flight calibration using reference gases as used in several previous works. In addition, further testing is needed to characterize the effects of flow rate on the values of carbon dioxide detected. This includes wind tunnel and laboratory testing to determine if the levels being detected are the actual values of CO_2 or if too high of a flow rate does not allow for the gas to be fully measured. Further validation of field tests would also be desirable, such as comparing the atmospheric levels measured by the aircraft to a network of sensors on both the ground and at various altitudes (such as mounted to towers or balloons). More detailed mathematical plume modeling could also be a key component of these tests.

The system used for these field tests was still a prototype configuration, which, while ideal for troubleshooting and the ease of making changes, was bulky and offers potential for a reduction of size and increase in capabilities. Because the Pixhawk already includes a GPS unit, altimeter, and SD logging, essentially the addition of the CO_2 sensor and flowmeter would allow for the elimination of nearly all of the sensor system boxes now included in the aircraft. This would remove nearly 2 lbs from the aircraft weight once the sensor system batteries and other components are included.

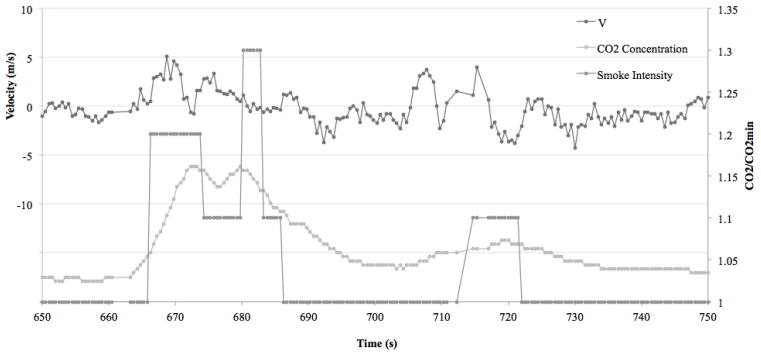
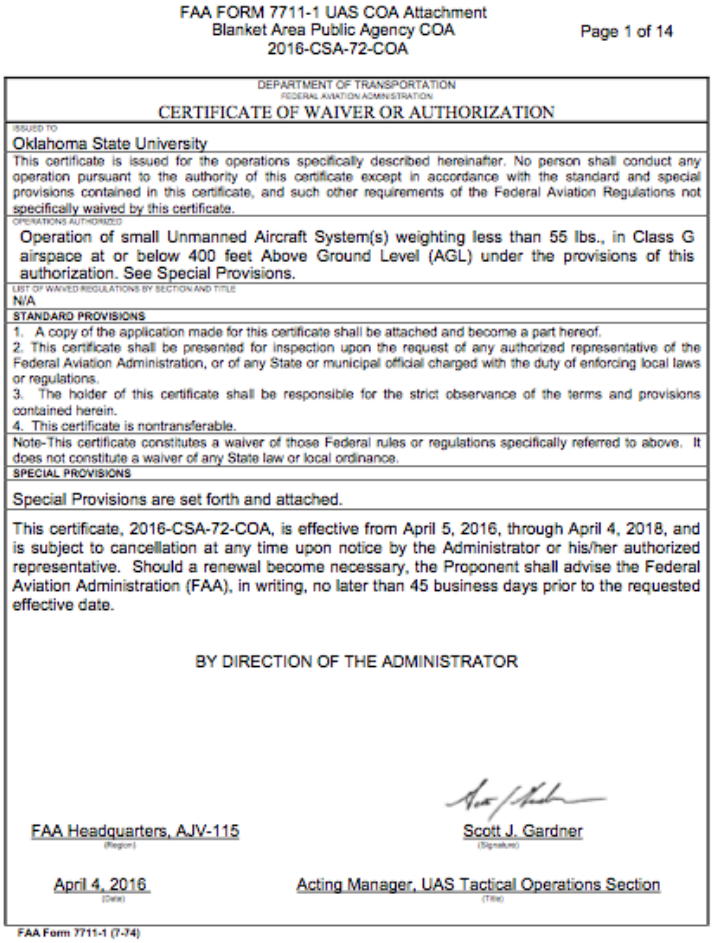


Figure 3.32: CO_2 levels compared with smoke intensity and vertical velocity.

Flight Testing at OSU and Farnsworth Sites OSU was granted a Certificate of Authorization (COA) for operations of SUAS within the NAS starting in April 2016 (2016-CSA-72-COA) as shown in Fig. 3.33. This allow operation of any SUAS $<$ 55 lbs below 400 feet anywhere in class G airspace with a 24 hr NOTAM. Additional COAs for operations above 400 feet can be obtained as needed. Approved and trained pilots-in-command (PIC) and observers are used for system operation and all PICs and observers meet current FAA requirements.**elston** Certified commercial and UAS pilots from the OSU School of Aviation are also utilized as needed.

Approximately 50 individual flights over the course of the project were flown to develop and test the platform and the sensor system. The majority of these flights occurred at the OSU Unmanned Aircraft Flight Station and were used to evaluate the handling qualities of the vehicle, the autopilot and ground control systems, and the data acquisition system. Once the operation of each component was verified, flights were conducted to evaluate the science objectives, namely the ability to detect a specific gas tracer and map the plume to trace it back to its source. As discussed in detail above, surrogate sources were utilized at the OSU sites to generate either controlled (tanks) or uncontrolled (burns) gases to evaluate the system performance. Since approval to fly at the Farnsworth site was not approved by the FAA until April 6, 2016, no flights occurred at Farnsworth until this time. A test flight was performed in May 2016 to evaluate field requirements and determine if any radio interference may be present at the Farnsworth location. There were no flight tests in June due to intermittent weather and conflicting flight operations, but tests were performed at the OSU site. While tests were conducted in July, high winds limited flight operations and likely prevented acquisitions of any useful data. Monthly flights were conducted thereafter. A summary of all flights with the fully operable DAQ system is shown in Table 3.1 and typical flight profile shown in Fig. 3.35

In between flight tests, the CO_2 system was improved with the addition of sensors to aid CO_2 tracking as well as miniaturization to improve aircraft performance. External temperature,



FAA FORM 7711-1 UAS COA Attachment
Blanket Area Public Agency COA
2016-CSA-72-COA

Page 1 of 14

DEPARTMENT OF TRANSPORTATION
FEDERAL AVIATION ADMINISTRATION

CERTIFICATE OF WAIVER OR AUTHORIZATION

ISSUED TO
Oklahoma State University

This certificate is issued for the operations specifically described hereinafter. No person shall conduct any operation pursuant to the authority of this certificate except in accordance with the standard and special provisions contained in this certificate, and such other requirements of the Federal Aviation Regulations not specifically waived by this certificate.

OPERATIONS AUTHORIZED
Operation of small Unmanned Aircraft System(s) weighting less than 55 lbs., in Class G airspace at or below 400 feet Above Ground Level (AGL) under the provisions of this authorization. See Special Provisions.

LIST OF WAIVED REGULATIONS BY SECTION AND TITLE
N/A

STANDARD PROVISIONS
1. A copy of the application made for this certificate shall be attached and become a part hereof.
2. This certificate shall be presented for inspection upon the request of any authorized representative of the Federal Aviation Administration, or of any State or municipal official charged with the duty of enforcing local laws or regulations.
3. The holder of this certificate shall be responsible for the strict observance of the terms and provisions contained herein.
4. This certificate is nontransferable.

Note-This certificate constitutes a waiver of those Federal rules or regulations specifically referred to above. It does not constitute a waiver of any State law or local ordinance.

SPECIAL PROVISIONS
Special Provisions are set forth and attached.

This certificate, 2016-CSA-72-COA, is effective from April 5, 2016, through April 4, 2018, and is subject to cancellation at any time upon notice by the Administrator or his/her authorized representative. Should a renewal become necessary, the Proponent shall advise the Federal Aviation Administration (FAA), in writing, no later than 45 business days prior to the requested effective date.

BY DIRECTION OF THE ADMINISTRATOR

FAA Headquarters, AJV-115
(Region)

Scott J. Gardner
(Signature)

April 4, 2016
(Date)

Acting Manager, UAS Tactical Operations Section
(Title)

FAA Form 7711-1 (7-74)

Figure 3.33: Blanket COA.

humidity, and wind sensors (as well as optional smoke sensor for flights over fires as a surrogate CO_2 source) were added to provide weather data. These sensors were mounted on a 3D printed mast protruding from the top of the aircraft to provide measurements from free-stream flow. Wireless ground sensors were developed to measure local atmospheric properties around the testing area. These sensors were created using over the counter electronic components and were assembled to measure temperature, relative humidity, barometric pressure, local wind speed, and direction. These values were measured at 1/2Hz and wirelessly transmitted to a local ground computer for display and storage. In addition, a low-cost five-hole pitot probe system is under development to measure wind speed and direction. New high-resolution (25 micron) 3D printing is being used to manufacture these pitot probes whereas previously it cost several hundred to thousand dollars to purchase probes machined from metal. The prototype of this pitot system as well as the atmospheric sensors were flown as part of the flight campaign as an initial validation.

Year	Month	Day	Location	Vehicle	Sensor System
2014	2	11	UAFS	X-8	CO2 V1
2014	3	13	MAR	Skyhunter	CO2 V0.5
2014	3	25	MAR	Skyhunter	CO2 V0.5
2014	7	1	UAFS	X-8	CO2 V1
2014	7	1	UAFS	X-8	CO2 V1
2014	7	14	UAFS	X-8	CO2 V1
2014	10	9	UAFS	X-8	CO2 V1
2014	10	15	UAFS	X-8	CO2 V1
2014	10	21	UAFS	X-8	CO2 V1
2015	2	1	MAR	Skyhunter	CO2 V1
2015	2	12	MAR	Skyhunter	CO2 V1
2015	3	5	UAFS	X-8	CO2 V1
2015	3	14	UAFS	X-8	CO2 V1
2015	3	20	MAR	Skyhunter	CO2 V1
2015	3	27	LCB	Skyhunter	CO2 V1
2015	3	30	LCB	Skyhunter	CO2 V1
2015	4	10	MAR	Skyhunter	CO2 V1
2016	5	25	FARN	Skyhunter	CO2 V2
2016	6	22	UAFS	Skyhunter	CO2 V2
2016	7	6	FARN	Skyhunter	CO2 V2
2016	8	17	FARN	Skyhunter	CO2 V2
2016	9	14	FARN	Skyhunter	CO2 V2
2016	10	12	FARN	Skyhunter	CO2 V2
2016	11	9	FARN	Skyhunter	CO2 V2

Table 3.1: Summary of data flights with UAS. Locations indicate the OSU Unmanned Aircraft Flight Station (UAFS), the Marena Mesonet (MAR), Lake Carl Blackwell (LCB), or the Farnsworth (FARN) sites.

The measured air properties were then displayed on a local computer using a custom graphical interface created using an open source visualization software called Processing. With processing, the recorded information would be shown at the approximate ground location for each sensor. These sensors have currently been used in support of on-going research at Oklahoma State University: including confirmation of wind changes during CO_2 data collection flights. With the sensors being developed in-house, the sensors have the capability to incorporate additional data collecting tools for additional data collection. Additional components, such as GPS, onboard data save abilities, and higher resolution sensors, were integrated and tested. Flight testing was conducted at both the DOE Atmospheric Radiation Measurement Southern Great Plains Site and Farnsworth with the system. Monthly test flights at the site were conducted starting in August 2016 through the end of the year. Sample data from a test flight at Farnsworth is shown in Figures 3.38a and 3.38b. The flight path is color coded based upon measured CO_2 levels. To date, no extreme CO_2 excursions have been detected using the airborne system.

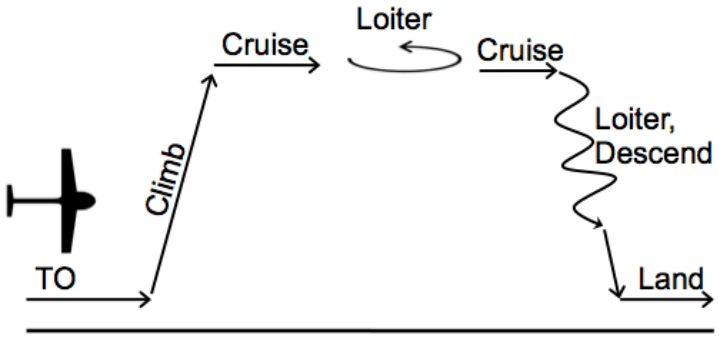


Figure 3.34: Mission profile.

Figure 3.35: Aircraft with improved sensor mast.

During the project, a larger more capable platform was also evaluated, as shown in Fig. 3.39, to carry more payload, fly for longer periods of time for longer surveys, and explore options for automated take off and landing for use by a more general end user. With this capability, no piloting skill is required since the vehicle can perform completely autonomously. This platform features a twin boom, inverted-V tail pusher design, with larger payload capability and endurance. Several demonstration flights have been flown to reduce risk and validate autonomous subsystems required for the plume encounters. The flight test methodology is modeled after civilian test flight procedures. The aircraft takes off and flies directly to a loiter way point to take initial measurements. The aircraft is commanded to enter the base leg of its search pattern and transition to autonomous search mode. The leader follows the traffic pattern as normal. The aircraft searches at varying altitudes and speeds. This evaluation was deemed successful though wasn't used in the Farnsworth tests due to need for a runway or improved road surface. However, it does demonstrate feasibility for further deployment to large areas.

Figure 3.36: Ground station at the Farnsworth site.

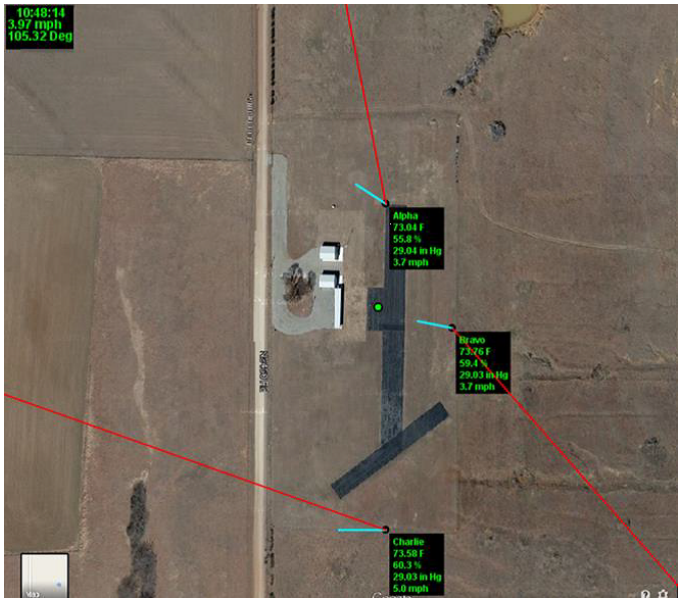


Figure 3.37: Graphical weather data interface showing local wind direction and conditions.

3.0.1 UAV Design, Evaluation, and Deployment Summary and Conclusions

Flight testing and analysis of the system is continuing, with combined tests of the ground and airborne detection system to evaluate performance and potential system improvements. One of the most important areas for improvement is the flow rate system. Attempts were made to increase the flow rate through passive methods such as varying the design of the inlet and exit configurations of the system. While these demonstrated a noticeable improvement, they still did not provide a flow rate with a high enough sensor response time. While a few seconds delay in peak detection time is short enough to shift the data to refine the location of the

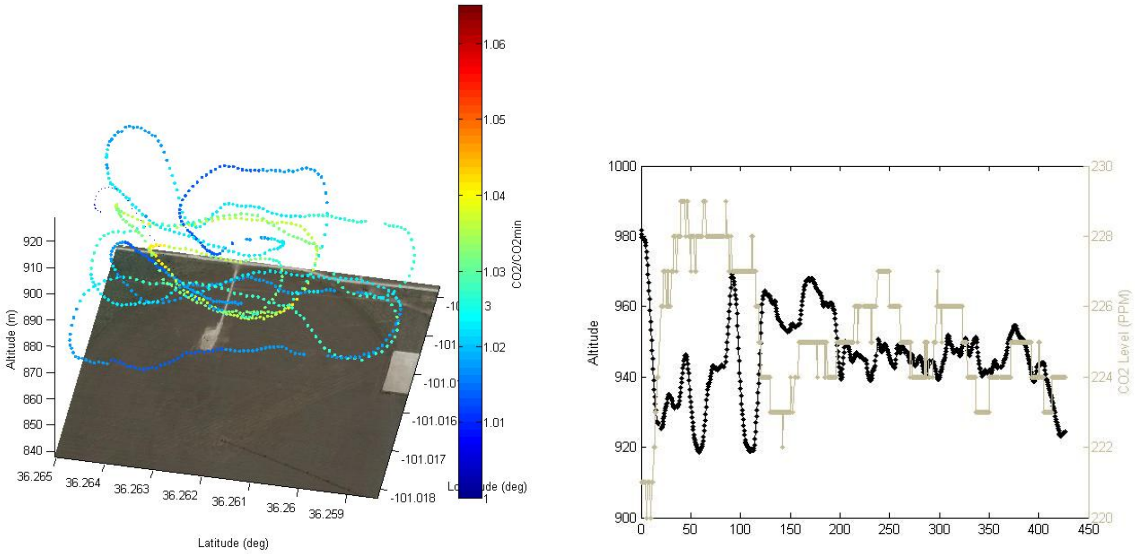


Figure 3.38: Flight track at the Farnsworth site and CO_2 levels.

Figure 3.39: Anaconda.

detected plume, the greater issue is the time it takes for the sensor to return to atmospheric background levels. This issue was highlighted best during controlled CO_2 release tests. In these tests remnants of the increased CO_2 air sample had not fully left the system, causing the CO_2 levels to simply increase with each subsequent pass through the plume. Further increasing the flow rate through the system would do much to alleviate this problem, but the required increase may be beyond the reasonable capabilities of passive improvements and it may be necessary to install a pumping system onboard the aircraft.

While the system is able to detect increased CO_2 levels, further testing and improvements

are required to be able to accurately quantify the gas levels detected to better characterize the plume. This type of high accuracy measurement is necessary to detect small levels that may be but a few ppm above ambient. Improved calibration techniques are a key part in attaining this level of accuracy, possibly including in-flight but at least pre- and post-flight calibration using reference gases as used in several previous works. In addition, further testing is needed to characterize the effects of flow rate on the values of carbon dioxide detected. This includes wind tunnel and laboratory testing to determine if the levels being detected are the actual values of CO_2 or if too high of a flow rate does not allow for the gas to be fully measured. Further validation of field tests would also be desired, such as comparing the atmospheric levels measured by the aircraft to a network of sensors on both the ground and at various altitudes (such as mounted to towers or balloons). More detailed mathematical plume modeling could also be a key component of these tests.

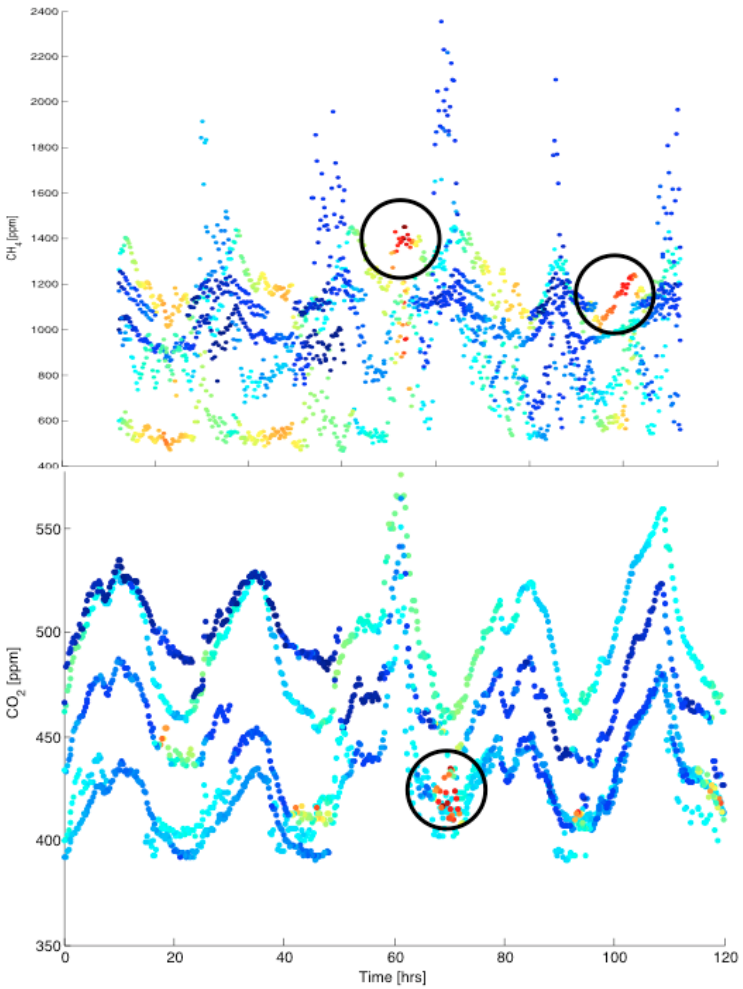


Figure 3.40: CH_4 and CO_2 detection.

Finally, there is room for improvement in the design of the sensor system itself. The system

used in this project was a prototype configuration, which, while ideal for troubleshooting and the ease of making changes, was bulky and offers potential for a reduction of size and increase in capabilities. Many autopilots (including the Pixhawk) offer the ability to modify the firmware to communicate with additional sensors through methods such as I2C. Because the Pixhawk already includes a GPS unit, altimeter, and SD logging, essentially the addition of the CO_2 sensor and flowmeter would allow for the elimination of nearly all of the sensor system boxes now included in the aircraft. This would remove nearly 2 lbs from the aircraft weight once the sensor system batteries and other components are included. Great care would need to be taken, however, to ensure that the additional sensor communication would not interfere with the processing capabilities of the autopilot and extensive testing would need to be performed to verify the reliability of the firmware changes. In addition to a reduction in weight and volume of the sensor system, integration with the autopilot would also provide increased capabilities to the system. A higher logging rate may be possible as well as integration with the aircraft wireless telemetry. This wireless datalink would allow for the CO_2 levels to be streamed live to the ground station instead of having to wait to download the data after the flight which would allow in-flight mission changes based on areas of interest; in addition the values could then be displayed on the on-screen display of the pilot's FPV video system for increased situational awareness and flight planning. Finally, integration with the autopilot system would allow for future improvements in automated path planning as are being worked on by other teams in the mechanical and aerospace engineering department. This would allow algorithms to be applied to the data as it is recorded to make the mission more efficient by focusing the flight path on the areas of greatest interest while only sparsely covering areas with no variation in gas levels. Many of these elements are being developed as part of other efforts at OSU not part of the project, but which may benefit future studies of this kind. A prototype implementation of the system is shown in Figs. 3.41 and 3.42.

Figure 3.41: Prototype mobile eddy covariance "tower" with integrated ultrasonic sensor with CO_2 detection.

Figure 3.42: Flight testing of mobile eddy covariance system.

4 Data Analysis and Sensor Optimization

4.1 Introduction

The work described in this chapter focuses on the development of optimal sensor allocation and UAS path planning algorithms to improve the speed and likelihood of detection of carbon dioxide and methane (CO_2 and CH_4) plumes. Techniques are described that extend recently developed methods (Allamraju et al., 2014) to data collection and path planning which help the sensor detect CO_2 and CH_4 leaks with high confidence. In order to achieve this, a framework of Markov Decision Processes (MDPs) is used that provides tools for path planning in presence of uncertainty. In particular, the problem of detection of gas plumes is modeled as a non-stationary Markov Decision Process. The non-stationarity in the decision process comes from the variation in the CO_2 and CH_4 gas concentrations due to biological or environmental activity. A mathematical simulation of a gas plume formation and concentration fluctuations as a function of time using a Gaussian Processes to model the gas density variations and their measurements. The simulation models recurring, but stochastic patterns, in the accumulation of CO_2 or CH_4 gas over an area. The model implements a non-stationary MDP planning algorithm developed Chowdhary, *et. al*, (Allamraju et al., 2014) to plan optimal paths in presence of the non-stationarity plume. Results indicate the feasibility of this approach, and have given valuable insights into improving the methods.

4.2 Background: Gaussian Processes and Gaussian Process Clustering

4.2.1 Gaussian processes

A Gaussian Process (GP) is a stochastic process, involving any subset of random variables that have a joint Gaussian distribution (Rasmussen and C. Williams, 2005). A GP is defined completely by its mean $m(x)$ and covariance function $k(x, x')$ and is written as

$$f(x) \sim GP(m(x), k(x, x')) \tag{4.1}$$

This states that a GP is a distribution over functions and draws from a GP that is a function which represents a generative model of the given data. Under GP regression the mean is assumed to lie in a class of functions

$$\mathcal{G} = \left\{ g(\cdot) \in \mathbb{R}^{\mathcal{X}} \;\middle|\; g(\cdot) = \sum_{i=1}^{\infty} \alpha_i k(x_i, \cdot) \right\}, \tag{4.2}$$

where $\mathcal{X} = \mathbb{R}^n$, $\alpha_i \in \mathbb{R}$, $x_i \in \mathcal{X}$.

The space $\mathcal{G}$ is a subspace of $\mathcal{H}$, a reproducing kernel in Hilbert space.

Given a set of data and observations $D = (x_i, y_i)_{i=0}^n$, where $x_i \in \Re^D$ and $y \in \Re$, we wish to make inferences about the relation between the input x_i and observations y_i. The data set defines a covariance matrix

$K_{ij} := k(x, x')$,

which generates a mapping ψ to an RKHS $\mathcal{H}$ such that $k(x, x') = \langle \psi(x), \psi(x') \rangle_{\mathcal{H}}$.

GP regression assumes that the uncertainty in the data and the model follow Gaussian distributions, while modeling the function estimate using a mean function $\hat{m}$ and a covariance function $\hat{\Sigma}$. Since the observations are Gaussian, the likelihood function $p(y|X, w)$ is also Gaussian. The initial prior is set to $p(w) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_w})$, and Bayes' rule is used to infer the posterior distribution $p(w|X, y)$ with each new observation y. Since the posterior is Gaussian, the update generates a revised mean $\hat{m}$ and covariance $\hat{\Sigma}$.

4.2.2 Gaussian process clustering

The inference method described above assumes that data D is generated from a stationary generative model. However most environments are non-stationary and generate data that cannot be classified into a single model. Therefore a method has been implemented to cluster data sets that are different as described in (Grande, Chowdhary, and J. How, 2013). The clustering technique maintains a history of all data sets it classifies as differentiable. For a GP, the log likelihood of a subset of points y can be evaluated as

$$\begin{aligned} \log P(y \mid x, M) = &-\frac{1}{2}(y - \mu(x))^T \Sigma_{xx} (y - \mu(x)) \\ &- \log |\Sigma_{xx}|^{1/2} + C, \end{aligned} \tag{4.3}$$

where $\mu(x) = K(X,x)^T(K(X,X)+\omega_n^2 I)^{-1}Y$ is the mean prediction of the model M and $\Sigma_{xx} = K(x,x)+\omega_n^2 I - K(X,x)^T(K(X,X)+\omega_n^2 I)^{-1}K(X,x)$ is the conditional variance plus the measurement noise.

The log-likelihood contains two terms that account for the deviation of points from the mean, $\frac{1}{2}(y-\mu(x))^T\Sigma_{xx}(y-\mu(x))$, as well as the relative certainty in the prediction of the mean at those points $\log|\Sigma_{xx}|^{1/2}$. Based on a set of points which are considered unlikely to have arisen from current model, a new GP model is constructed from these sets and is tested against all stored models using a non-Bayesian Hypothesis test. If the new model merits instantiation it is then added to the set of existing GP models. Algorithm 1 describes the algorithm of clustering GPs (Grande, Chowdhary, and J. How, 2013).

Algorithm 1 GP Clustering

Input: Initial data (X,Y), lps size l, model deviation η
Initialize GP Model 1 from (X,Y).
Initialize set of least probable points $S=\emptyset$.
while new data is available **do**
  Denote the current model by M_c.
  If data is unlikely with respect to M_c, include it in S.
  if $|S| == l$ **then**
    for each model M_i **do**
      Calculate log-likelihood of data points S using having been generated from current model M_i $\log(S|M_i)$, and find highest likelihood model M_h, making M_h current model.
      Create new GP M_S from S.
      if $\frac{1}{l}(\log(S|M_S)-\log(S|M_c)) > \eta$ **then**
        Add M_S as a new model.
      end if
    end for
  end if
end while

4.3 Markov Decision Processes

Markov Decision Processes (MDP) is the common framework for sequential decision making under uncertainty. A MDP is defined as a tuple (S,A,T,R,γ), where S indicates the set of possible states, A is the set of all possible actions that can be taken by the agent, T describes the stochastic transition by agent due to disturbances and R indicates the reward indicates the reward given to the agent for taking action at a given state. The factor γ is the discount factor which describes the weight given to the future consequences of current actions. The

goal of an agent operating in a MDP is to maximize it expected sum of discounted rewards it gathers by sequentially making decisions. Mathematically the agent maximizes the following function with the goal minimizing sampling efforts by the agent while maximizing the value of the information collected.

$$V(s) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)] \tag{4.4}$$

The solution to a MDP problem is a policy, $\pi : S \rightarrow A$, that provides a sequence of actions that maximizes the above equation. MDPs are solved by recursively solving the Bellman equation(2) that describes the value of taking an action a in a state s.

$$V^{\pi}(s) = E[R(s, a) + \gamma V^{\pi}(s')] \tag{4.5}$$

The optimal value of the Bellman equation, which represents a fixed point in the space of values, and the corresponding optimal policy are given as

$$V^*(s) = max_{a \in A}[E[R(s, a) + \gamma V^*(s')]] \tag{4.6}$$

$$\pi^* = argmax(V^*(s)) \tag{4.7}$$

4.3.1 Non-stationary MDP

If the reward $R(s, a)$ is not known fully, the MDP problem can be solved using reinforcement learning approaches (Sutton and Barto, 1998; Busoniu et al., 2010). In model-based reinforcement learning, a model $\hat{R}(s, a)$ needs to be learned. Once the model is learned, the policy π can be solved by using the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \hat{R})$. The approximate reward function $\hat{R}(s, a)$ is the predictive mean of a Gaussian process inference. Algorithm 2 describes the algorithm for solving a non-stationary Markov Decision Process using a model based reinforcement learning approach. This algorithm was inspired by our recent work (Allamraju et al., 2014).

In this method, the agent starts by exploring the environment based on the predictive variance of the GP. This variance indicate the confidence the GP has in the model and hence exploring based on the variance directs the agent towards states where predictive variance is high. Once the variance of the GP goes below a particular threshold the agent makes a decision to exploit its knowledge by using the predictive mean of the GP to solve the MDP. The first type is called an exploration MDP as it's reward function motivates exploration while the latter MDP is termed an exploitation MDP. This method currently defines the area susceptible to leaks as a fully observable Markov decision process and solves the problem according to the framework described above.

Algorithm 2 Nonstationary MDP Solver

Initialize: Initial data (X, Y), lps size l, model deviation η, GP parameters (σ, ω_n^2), space exploration threshold φ.
while new data (x_i, y_i) is available **do**
  Update GP cluster using Algorithm 1.
  Compute exploration reward $\hat{r}_{e_{t_i}}$
  Compute space explored s_e
  if $s_e > \varphi$ **then**
    Solve exploitation MDP $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \hat{r}_{t_i})$ to get path.
  else
    Solve exploration MDP $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \hat{r}_{e_{t_i}})$ to get path.
  end if
end while

4.4 Examples

Sensor identification of gas plumes were tested on a non-stationary plume model. The first set of figures (Fig. 4.1) indicate the different reward functions that are generated due to movement of the plume caused by unknown environmental effects or wind dynamics. Cost functions reward the detection of the actual plume and penalize the detection of the transient CO_2 variations. Consequently, the peak in the cost function indicates the position the gas in the atmosphere from a source. Valleys indicate reward given to the agent when it detects a false positive on the gas due to alternate sources. The agent gathers reward samples and builds an estimate of all the models. These models were generated by a gas plume simulator that artificially manufactures different plume models

The next set of figures (Fig. 4.2) are the models inferred by a Gaussian Process Clustering method. Using the method outlined in Algorithm 1 only six models out of the seven models tested were correctly predicted. The agent (UAV) acts on the predicted reward model and then plans path to maximize its reward. Such a path should lead the agent to the plume location in the quickest and most efficient manner, even when the plume location may not be exactly known.

Figure 4.3 indicate that the various paths taken by the agent from a random initial point to reach its destination, i.e. the position of the plume which indicates the leak. In the contour plots the exploration and exploitation paths are shown. The different paths serve to reduce the uncertainty in the model (exploration) and then use the learned model (exploitation) to find the plume. Using the current algorithm, the exploitation paths are computed when the agent has made the decision to stop exploring based on a threshold criterion and start using the predictive mean of the estimated model to plan the best path to the plume.

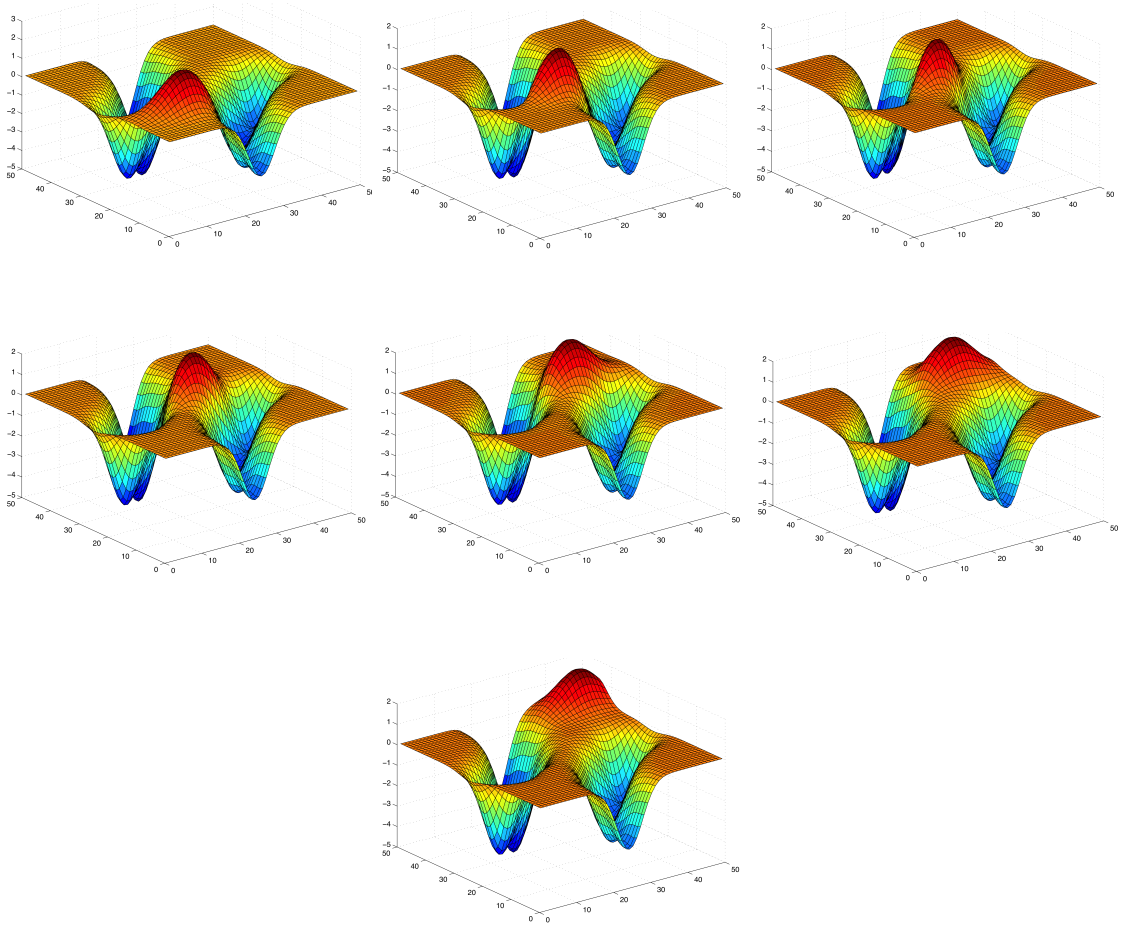


Figure 4.1: Reward function at different time instants used by the agent to gather data

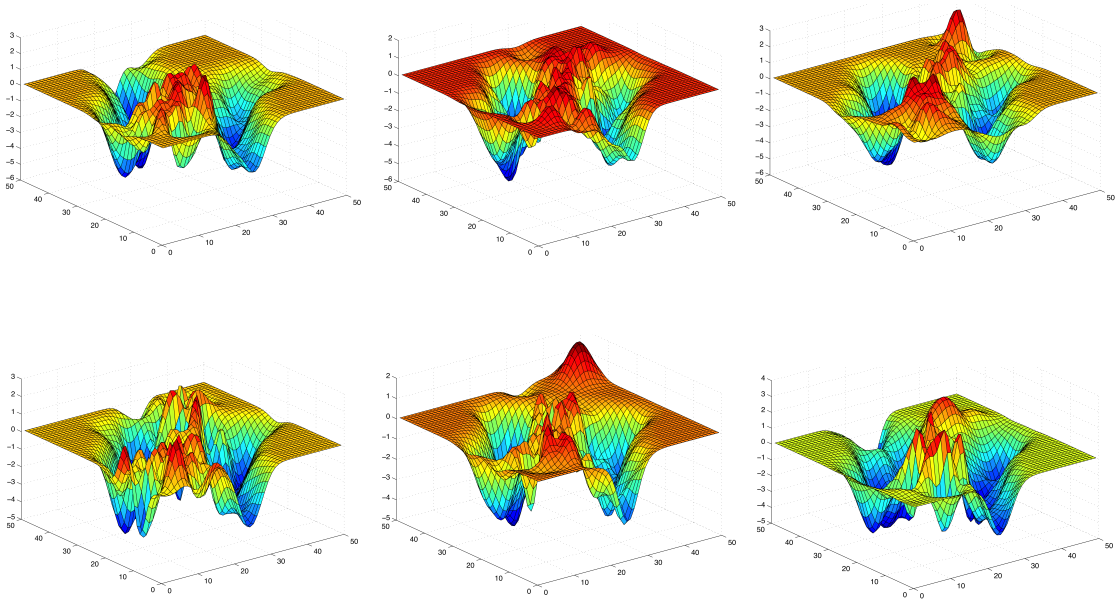


Figure 4.2: Models estimated by the agent

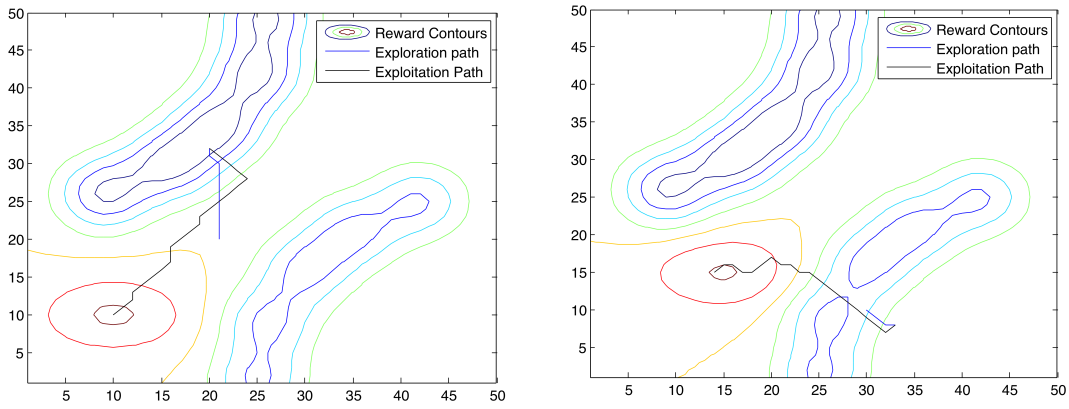
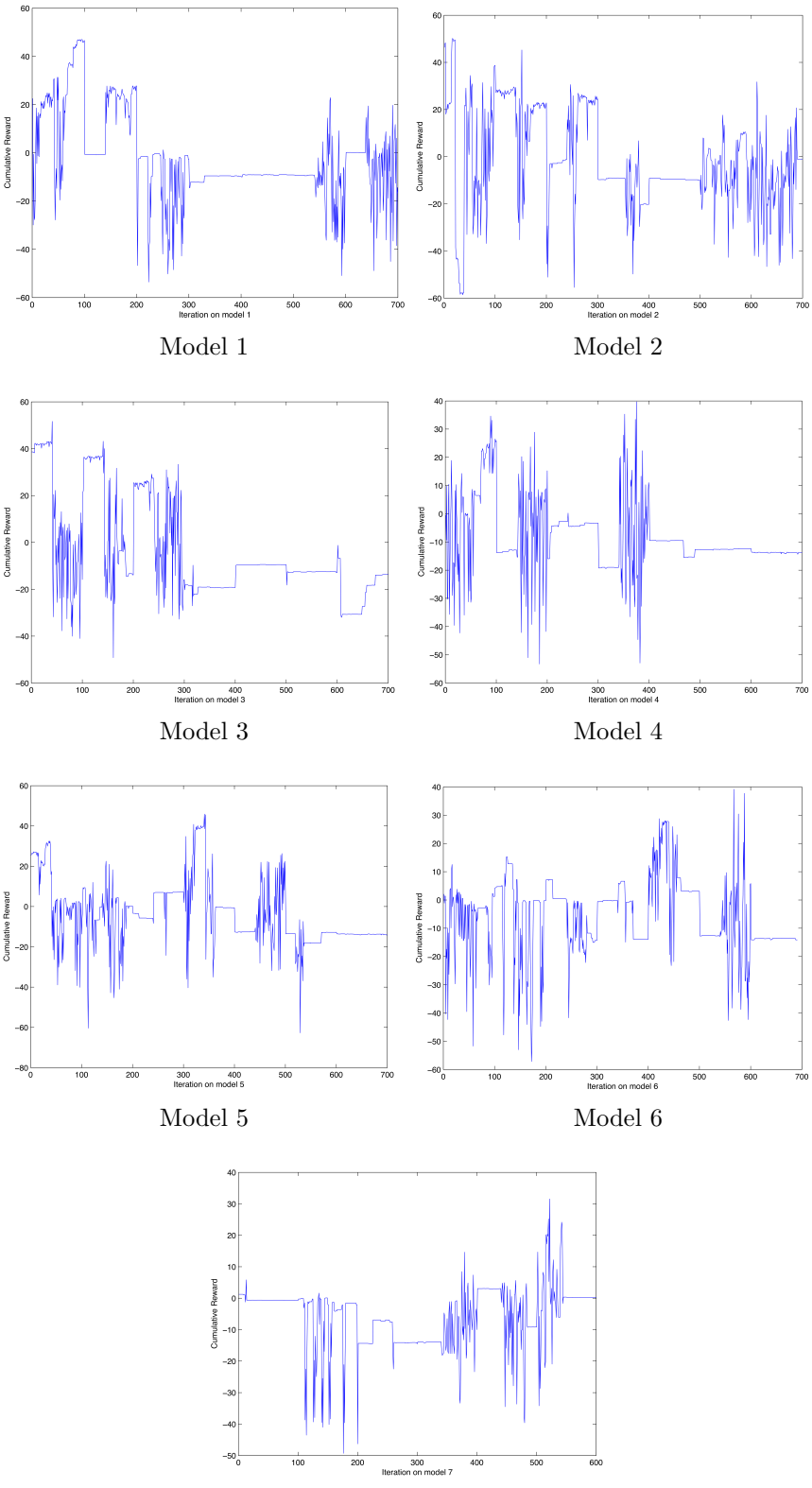


Figure 4.3: Planning waypoints for two different models

A performance measure of the agent is the cumulative rewards obtained when exploring the state space over several simulated runs. A higher cumulative reward indicates that the agent is able to exploit its available knowledge and optimize its planning strategies to quickly and efficiently find the plume. Figure 4.4 indicates the cumulative reward obtained by the agent for each model for the all the times in which that model led to. The performance measures indicate the feasibility of using the MDP and GP clustering based path planning approach to quickly detect the plumes. However, they also indicate that simple extensions of the existing algorithm are not enough to extract optimal performance.



Model 1 Model 2

Model 3 Model 4

Model 5 Model 6

Model 7

Figure 4.4: Cumulative rewards for each model

4.5 Model Refinement

In order to improve the performance of the models, novel mathematical tools to optimize the learning of the underlying CO_2 and CH_4 model have been developed. This allows for optimal sensor allocation and placement. First, methods were identified to facilitate data collection in a sensor's local environment, estimate the parameters that characterized the CO_2 and CH_4 local distribution, and share this information across a network in order to fuse information from all agents with "valuable" information, to estimate global parameters. The second part of this process involved developing techniques that leverage learning methods to perform UAV path planning in order to improve the speed and likelihood of detection of gas plumes by sampling areas with most likely high-value information.

4.5.1 Data fusion from multiple sensors using gaussian process fusion

The purpose of sensor allocation is to place sensors in locations where the probability of gas leakage is high. This is a challenging task in a multi-sensor setting as agents will only have access to local observations and coordination. It is more difficult when agents do not have a singular estimation on the plume model. The problem studied as a Markov Decision Process and the detection of gas plumes is modeled as a non-stationary Markov Decision Process. The goal is to optimize the sensor deployment by reducing uncertainties by using the non-stationary gas model. However, the land mass over which gas leakage detection is to be performed at the test location in the Farnsworth field is large. To cover that land mass, a number of sensors are needed. It is unrealistic for one UAV agent to explore the entire space to collect all the data, rather, techniques are needed to estimate the gas density model by using data obtained by a mixture of sensors. In most scenarios, stationary agents (i.e. ground sensors) are restricted to local areas where they identify the local parameters that make up the global model. In such a setting, agents communicate information to neighbors frequently to maintain a same set of global parameters and remain coordinated. However, these methods incur high communication costs, as agents have to frequently communicate each data point sampled in order to keep errors between their estimations bounded. To overcome this problem we recast the learning problem into a censoring based information fusion framework. In the GP fusion algorithm, agents communicate compressed generative models of the data that they have acquired to neighboring agents, instead of sending the data itself. This helps to mitigate the risk of communication clutter. The goal is to apply the GP Fusion framework to enable data from multiple ground sensors to be combined without overwhelming the communication channels. To validate the algorithm, a data set (Intel, 2014) of temporal-spatial variance of temperature on a grid for an experiment performed at Intel Berkeley Research Laboratory is used. The experiment was performed on a $40m \times 30m$ grid space with 54 sensors deployed strategically across the room. Fig. (4.5) shows the experimental grid setup with the position of the 54 sensors. All sensor positions are relative to top right corner of the grid. The experiment was conducted between February 28th and April 5th 2004 and Mica2Dot sensors with weather boards collected timestamped topology information,

along with humidity, temperature, light and voltage values once every 31 seconds. Data were collected using the TinyDB in-network query processing system, built on the TinyOS platform.

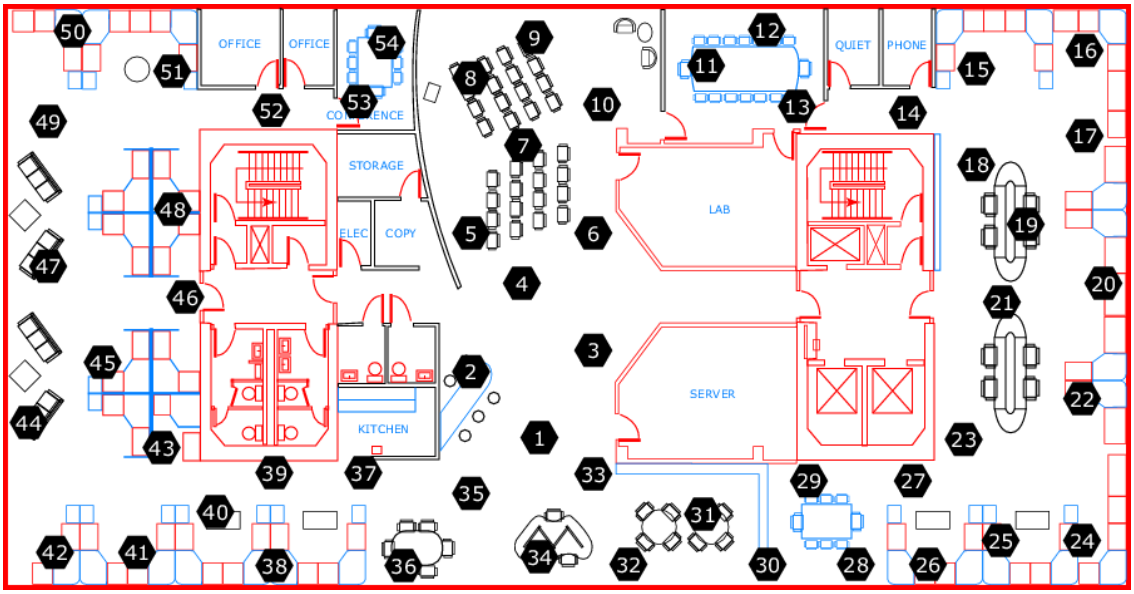


Figure 4.5: True Model of temperature distribution

The underlying principle of GP fusion is the consistency requirement of a Gaussian process. This property states that if given a GP with n-collection of random variables, then any subsets of n, m and (n-m) are draws from a GP whose parameters are subsets of the parameters of the main GP.

In this approach, a Sparse Online Gaussian Process approximation is used for calculations, as described by (Csató and Opper, 2002). The algorithmic design for achieving the information fusion is described in Algorithm 3 In the algorithm shown below the method for learning the global parameters of the model when agents are restricted from accessing global data is outlined.

Inputs to the algorithm are the set of agents, the time after which the process should terminate, the time increment, the type of graph in which the agents are linked, the cost threshold and the Environment object. The environment object holds information about the world states and the required function over which estimation is to be performed. Given these sets of inputs a network instance is created.
The functions described above perform the following tasks.

- *CollectnUpdate* function [Algorithm 4] causes the agents to pick up samples from the environment and update their GP's with the sampled data points.

Algorithm 3: Gaussian Process Fusion

```
Input  : Agent set A, Time Duration T, time increment Δt,Environment object,
         graph connectivity, cost
Output: Global Fused Predictive mean μ̂
begin
    Initialize: Network instance for all agents a ∈ A
    t ← 0
    while t < T do
        for a ∈ A do
            γ_a ← CollectnUpdate(a,Environment object)
            Voi = γ_a
            if Voi > V* then
                Transmit(a)
        if Networkcost < cost then
            V* = 0.9V*
        else if Networkcost >= cost then
            V* = 1.1V*
        t ← t + Δt
```

- *agenttype* function returns a flag which indicates whether an agent is a static or a dynamic type and random exploration is performed by the dynamic agent type.
- *explore* function updates the position of the agent to a new position based on the random action taken by the agent.

The Value of Information(VOI) is a measure that is used to inform the network about the contributions an agent can make to the network, based on any new observed data. Based on this measure, agents censor themselves from transmitting unnecessary information and prevent network cluttering. The VOI is calculated using the kernel independence test metric $\gamma = k(x_i, x_i) - k^T(X, x_i)K(X, X)^{-1}k^T(X, x_i)$, which describes the novelty of adding a new point.

Algorithm 4: CollectnUpdate

```
Input  : Agent a instance, Environment object
Output: Updated network instance, γ_a
if agenttype(a) is static then
    (x,y) ← Getdata
    a.UpdateGP(x,y)
    γ_a ← a.GP.getgamma
else if agenttype(a) is dynamic then
    agentposition ← a.Explore
    (x,y) ← Getdata
    a.UpdateGP(x,y)
    γ_a ← a.GP.getgamma
```

For a particular agent, if the VOI exceeds a particular threshold, then the agent transmits its GP object across the network to its neighbors. The neighboring agents sample the GP according to the sampling scheme described in Algorithm 5. This schema is selected such that sampling is directed into regions where predictive variance of the transmitting agent is below a threshold and the predictive variance of the sampling agent is above a different threshold. The samples are then used to update the sampling agent's GP.

4.5.2 Experiments

Since most data collected in real time is non-gaussian in nature, it was an ideal dataset to perform simulation tests to identify the potential for the method. Since these methods can be applied to any real-time data set, using this small set was used for initial testing, however, in the future the technique will be used to perform rigorous tests on the real-time gas data (USGS, 2013). Figure 4.6 shows the model of temperature distributions and Fig. 4.7 indicates the estimated mean of all the agents in a censored setting. Note that all estimations are similar to a model which is closer to the true value as indicated by the error plots in Fig. 4.8. The error is non-zero value as all agents do not fully span the entire world and hence do not have full set of global parameters of the model. Figure 4.9 describes the value of information of each agent. Agent 1-3 being a dynamic agent has a high value of information over time. Agent 4 contributes to the network only in the initial step as the data around the static sensor is mostly unchanging and future contributions of that information to the network is unnecessary. The experimental results presented below include one set of experiments with conditions as described below.

- Simulation run time was 20 seconds with a time step of 0.01 second.
- Agents are censored based on their Value of Information at each time step
- 3 Dynamic agents, 1 Static agent
- Initial positions: (3,28),(10,4),(30,6),(40,15)

Algorithm 5: Transmit

Input : Agent a instance, ζ_1, ζ_2
Output: Updates to GPs of neighbouring agent of a
for *a' ∈ a.connected* **do**
  $(\tilde{X}, \tilde{Y}) \rightarrow$ find($variance(a) < \zeta_1$ and $variance(a') > \zeta_2$) a'.UpdateGP($\tilde{X}, \tilde{Y}$)
Networkcost ← Networkcost + 1

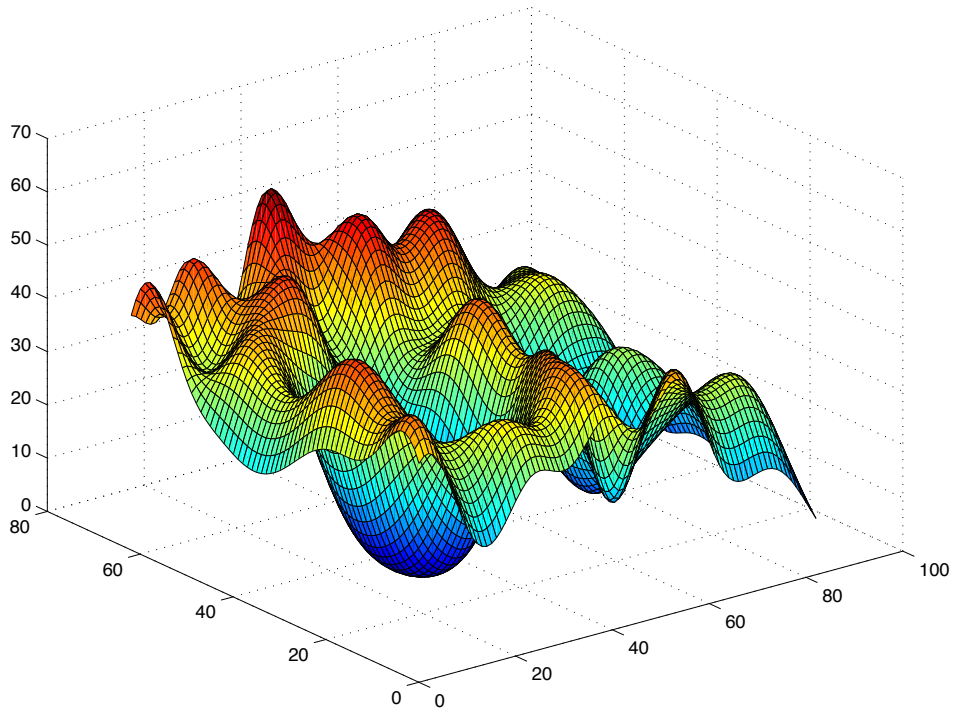


Figure 4.6: True model of temperature distribution

These results indicate the feasibility of Gaussian process information fusion for all agents in learning global parameters of the model when sensors have limited access to information.

To detect gas plumes, an array of static and mobile sensors are deployed in the test area. The mobile sensor measurements reduce the uncertainty of the model being built by the static sensors. Due to energy limitations and the size of the survey space, it is assumed that a UAV cannot visit all possible locations such as static sensor locations and regions of high uncertainty in one Hamiltonian path; i.e., a path that visits each vertex exactly once. The problem framework is illustrated in Fig. (4.10). The following preliminary formulation seeks to plan a series of Hamiltonian paths, $\underline{\eta}$, each of which visits some subset, η_t, of the possible location set, H, with the objective of maximizing a score function. For simplicity, each subset η_t is limited to κ out of K possible vertices; where $\eta \in \mathbb{R}^{\kappa}$, $t = 1, 2, ..., \infty$ and $H \in \mathbb{R}^{K}$. The score function of node $x_i \in H$ is defined as

$$f(x_i(t+1)) = \begin{cases} 0 + \epsilon, & x_i \in \eta_t \\ f(x_i(t)) + \epsilon, & f(x_i(t)) < 1 \\ 1, & \text{t=0 or otherwise,} \end{cases} \tag{4.8}$$

where the subset η_t is defined using the binary variable, $\gamma \in \mathbb{R}^{K}$, as

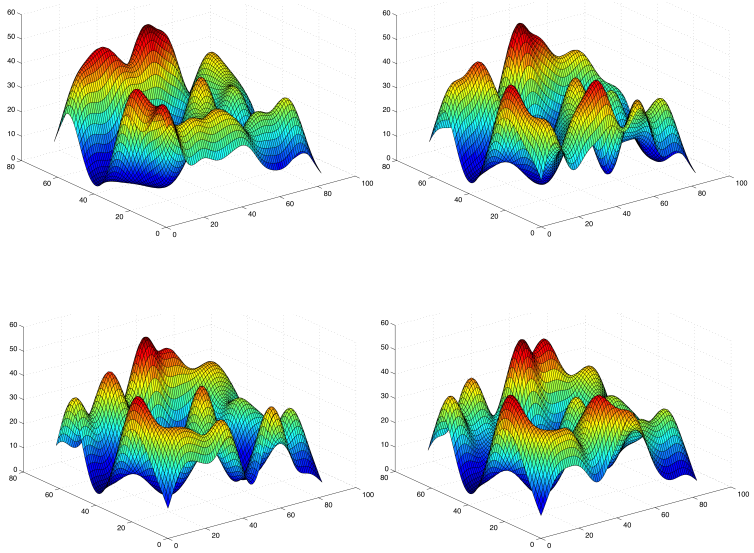


Figure 4.7: Estimated Model for each agent

$$\eta_t = \underset{x_i}{\operatorname{argmax}} \sum_{i=1}^{K} f(x_i(t))\gamma_i$$
$$s.t. \sum_{i=1}^{K} \gamma_i \leq \kappa, \tag{4.9}$$

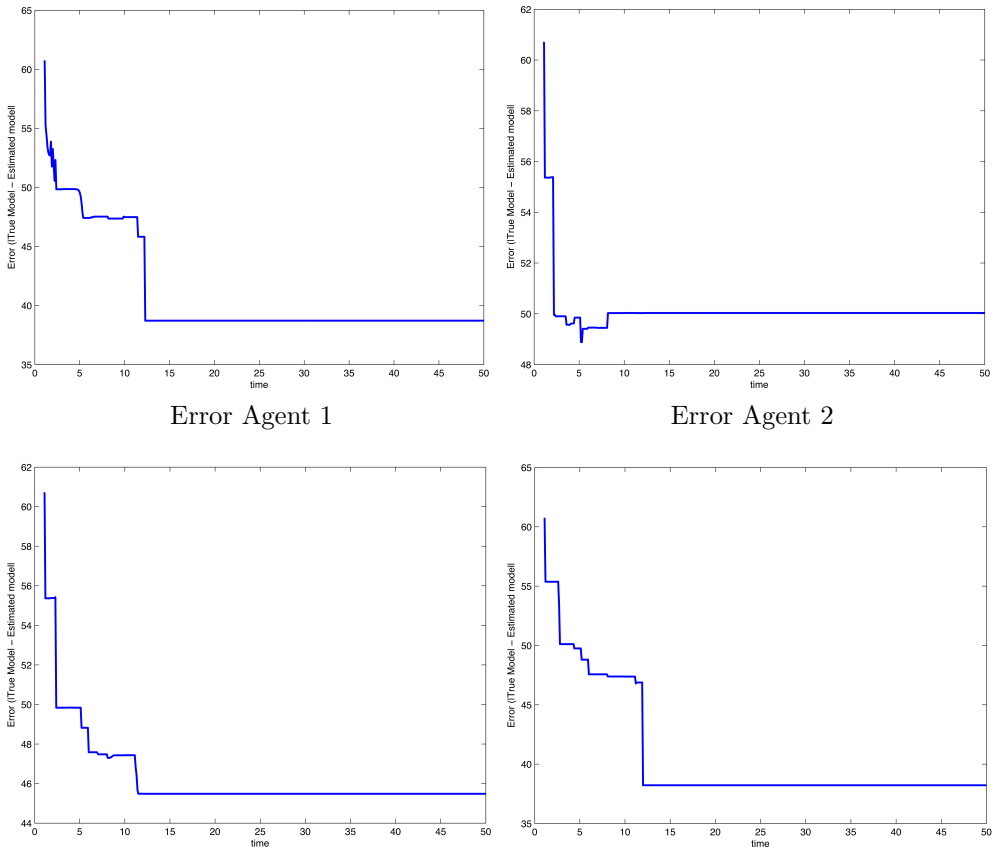
which yields the planned paths $\underline{\eta} = [\eta_1, \eta_2, ..., \eta_\infty]$.

4.5.3 Data preparation

In order to deal with large data sets with unknown errors, data preparation algorithms remove outliers, detect and mitigate inconsistencies in data, and model the noise properties in the data. These algorithms were validated on real CO_2 datasets obtained from the Farnsworth unit as well as the Berkeley temperature dataset, and satellite weather data.

4.5.4 Predictive representations

Algorithms for a communication-efficient sensor network have been developed to facilitate the optimization of CO_2 distribution models and sensor deployment to predict the source of potential CO_2 and CH_4 emisions. New and improved methods have been developed to learn the local models of CO_2 distribution, share information models under limited bandwidth and energy constraints and fuse information in an efficient manner to make predictions. Relevant data sets, which include distributions of atmospheric carbon dioxide from various



Error Agent 1 Error Agent 2

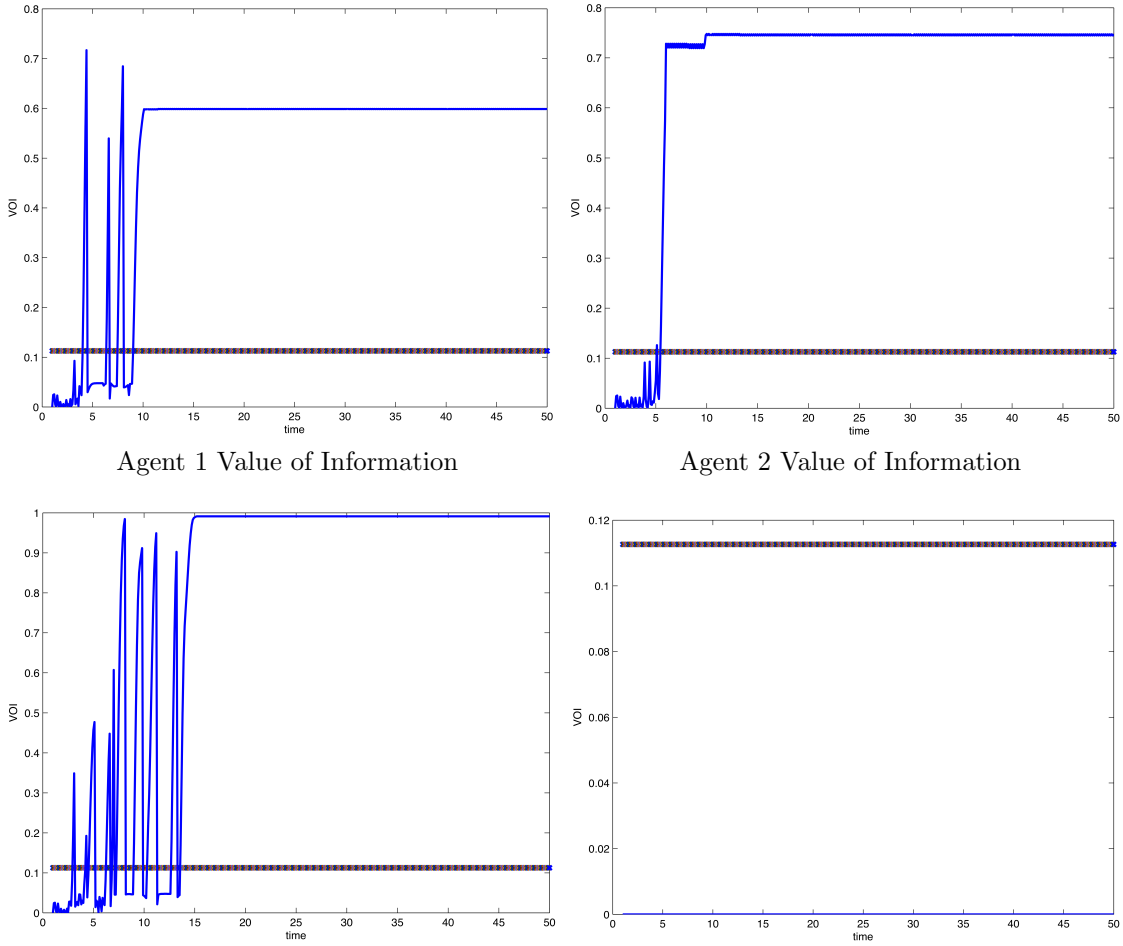
Error Agent 3 Error Agent 4

Figure 4.8: Estimated error for each agent

sources across the US and temperature and humidity distribution from Intel Research Labs in Berkeley were used to run simulations and test the efficiency of the sensor network.

4.5.5 Information fusion with reduced communication overhead

The objective of sensor allocation is to deploy sensors to locations that optimize the detection of gas leaks. This is a challenging task in multi-sensor setting as sensors will only have access to local observations and deploying sensors independently based on local information will not only result in suboptimal sensor placements but also incur high costs as well. It is mandatory for sensors to communicate their measurements to other sensors and jointly perform inference about the underlying unknown process. This is, however, a non-trivial task as communication between agents requires frequent message passing between agents, which is a highly resource intensive task. The challenges are further augmented by increasing size of sensor network and communication costs scale with the size of sensor networks in large domains, such as detection of CO_2 sources over a large area. Such problems can be approached by using graphical models



Agent 1 Value of Information

Agent 2 Value of Information

Agent 3 Value of Information

Agent 4 Value of Information

Figure 4.9: Estimated VOI for each agent

for distributed inference, which are powerful techniques to solve distributed inference class problems such as distributed sensing, parameter estimations and target tracking.

The problem approach involves using graphical models as they can easily capture the structure of a sensor network and provide a suitable framework for the development of communication constrained algorithms. Highly efficient and scalable algorithms for generating graphical models, which can easily be applied to large sensor networks have been developed. By combining graphical models and GP-Fusion techniques, a distributed inference method for large sensor networks was developed that can be generically applied to learn any process. The data sets that were used for the simulations involve learning of various physical phenomena and that list includes atmospheric carbon dioxide distribution (Energy(DOE) National Energy Technology Transfer (NETL), n.d.), methane distribution (2011) temperature and humidity distributions(Intel, 2014) over a large areas. Our approach shows significant improvements in the communication efficiency without sacrificing accuracy of estimation. Our methods,

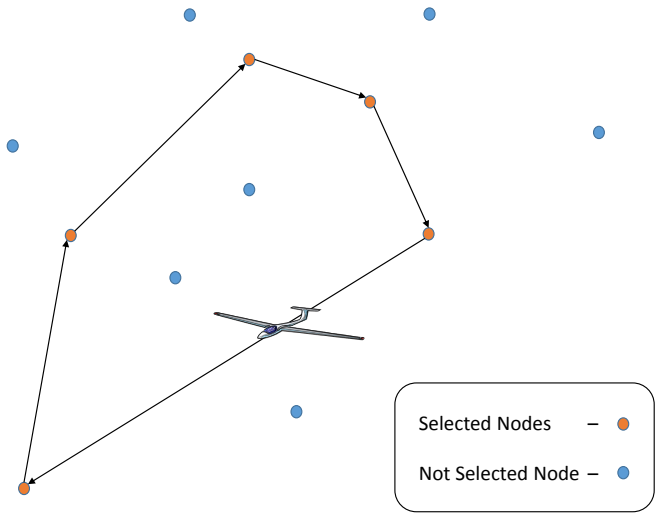


Figure 4.10: A UAV must plan a path that visits most informative nodes over the area to collect the most valuable information, however, it cannot visit all nodes due to fuel constraints. Therefore, the problem that must be solved is: find the most informative subset of nodes to visit that can be reached through one Hamiltonian path within UAV's fuel constraints

which we describe as 'Efficient Information Fusion using Gaussian Processes', is compared against state-of-the-art communication efficient algorithms and is shown to have significant improvements over them.

For the test simulations, the first data set was spatial distributions of stationary sources of atmospheric carbon dioxide from the National Carbon Sequestration Database and Geographic Information System(NATCARB)(Energy(DOE) National Energy Technology Transfer (NETL), n.d.). This database includes stationary sources of atmospheric carbon dioxide emissions across USA and Canada. Since the dataset was huge we chose a small set of sources within the regions of Oklahoma and Texas and tested our sensor network within this domain. The results, shown in Fig. 4.11, show accuracy of estimation and the cumulative cost incurred by the network for our methods compared against state-of-the-art algorithms. Figure 4.12 shows the algorithm performed with methane distributions from the EDGAR data set and Fig. 4.13 indicates the results of running the algorithm on the Intel Berkeley sensor network labs.

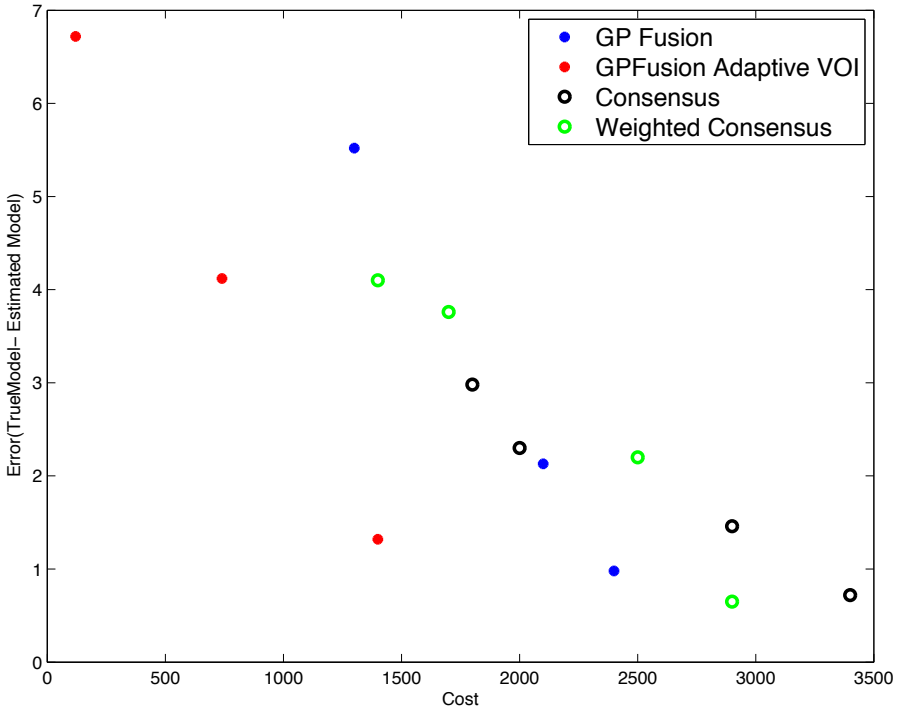


Figure 4.11: NATCARB carbon dioxide data

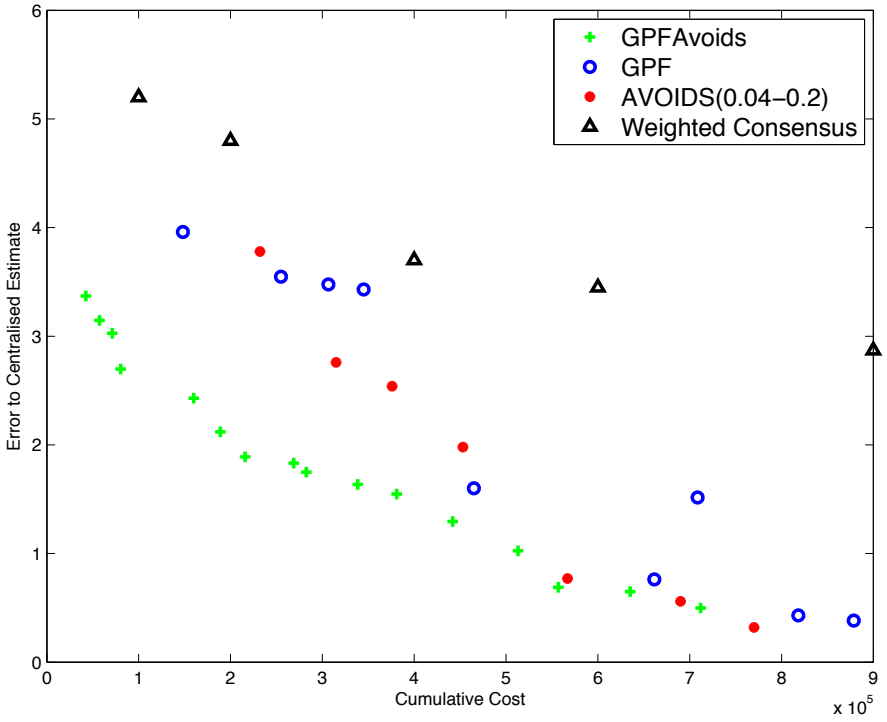


Figure 4.12: EDGAR Methane data

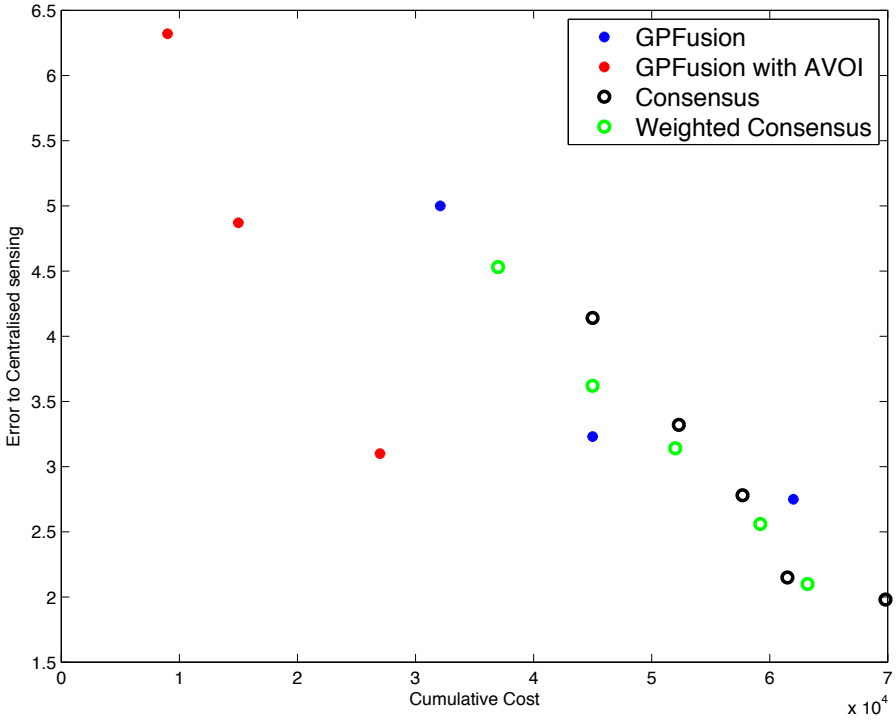


Figure 4.13: Berkeley sensor network temperature data

4.5.6 System optimization

Algorithms were developed to optimize the Value of Information (VOI) harvested by a set of mobile agents from a set of static surface agents over large spaces. This resulted in the N-Sample Data Ferrying algorithm which plans paths for information harvesting mobile agents (UAS in our case) as it learns which sensor locations are likely to have the highest Value of Information (VOI) in a time-varying environment.

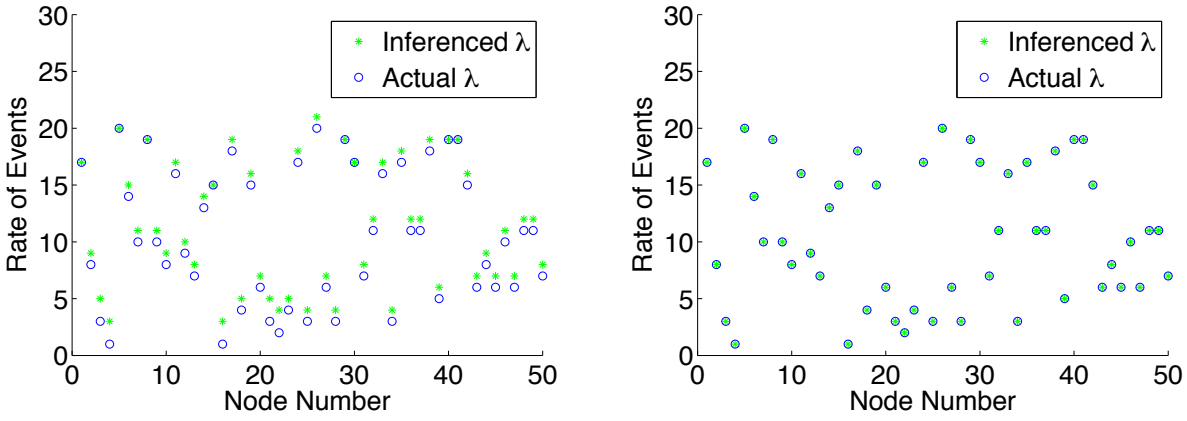
UAS allocation for optimal information harvesting

The arrival rate of the Value of Information (VOI) at the static agent locations is denoted henceforth as λ. The increment λ_i, is treated as a Poisson parameter and the likelihood is defined using the Poisson distribution. Since the likelihood of an independant identically-distributed (iid) Poisson distribution is equivalent to $t - \Upsilon_i$ samples, where Υ_i is the most recent time step at which static agent x_i was last visited, from a Poisson distribution and since the Gamma distribution is the conjugate prior of a Poisson distribution, the Gamma distribution is used as the prior distribution.

Simulation study

The algorithm was validated on a simulated monitoring mission over an area and setup representative of the Texas Andarko basin. The goal was to enable the UAS to ferry data from the ground sensors, as well as visit interesting locations on the map for monitoring. It was assumed that the UAV, is able to visit a subset of up to 6 static agents per episode; i.e., $\kappa = 6$. The mobile agent first explores the data available from each of the 50 deployed static agents ($K = 50$) due to an initially high assumed data feature arrival rate ($\hat{\lambda}_i = 100, \alpha_i = 100, \beta_i = 1 \; \forall i$). Bayesian inference is used to optimistically allocate the mobile agent to the static agent subset based on an optimistic prediction of the static agents' VOI. The mobile agent ferries data for a total of 1000 episodes. The optimistic allocation converges to the optimal allocation on average when the actual λ parameter is inferenced. Graphical results are presented in Fig. 4.14.

These results indicate the feasibility of this algorithm for sensor placement optimization and UAV deployment over the test section in the Farnsworth field.



Single-Sample Parameter Estimation

N-Sample Parameter Estimation

Figure 4.14: Parameter Estimation

4.6 Distributed information maximization

In order to develop predictive representations of underlying CO_2 and CH_4 changes using data sampled at various locations, algorithms in which a distributed network of sensors can build accurate representations of the CO_2 and CH_4 models with minimal communication between sensor nodes is needed. Using these predictive representation new and improved algorithms to make decisions in identifying regions of high information to reduce the cost of sampling. Techniques which plan optimal paths to highly informative regions while building predictive models online in a closed loop structure have been developed. Adaptive reconfiguration of the sensor networks is being explored in an effort to minimize the number of sensors needed to adequately sample a given area. This is accomplished by decomposing the study area into partitions such that sensing agents are locally constrained to smaller regions. This ensures that paths planned by agents in one region are independent of the paths considered by the other agents thereby reducing the computational penalty in planning and improving coverage of the operating domain. The goal is to tessellate the area such that static and UAV based sensors are best positioned to quickly detect any hydrocarbon leakage. In many standard coverage problems explored in the planning literature, a standard Voronoi tessellation is directly initialized to create subregions in the domain, based on assumed prior knowledge of a potential distribution. Agents are constrained to sense within these subregions (Pavone et al., 2009, M. Pavone, 2011). Many methods to accomplish this using Vornoi tessellation have been previously studied in the coverage control literature. These methods, however, are unable to adaptively reallocate the area to which a sensor is reliable in response to the sensor's capability, the available information in that area, or spatiotemporal effects such as wind. It is expected that these methods may not perform reliably while estimating spatiotemporally varying CO_2 and CH_4 distributions. A new adaptive sensor network reconfiguration paradigm was developed termed *Data Adaptive Voronoi Tessellation*, that uses the predictive confidence

of an agents model to adaptively reconfigure the area for which they are responsible. Gaussian process information fusion methods are used to infer parameters of interest locally and use this as a measure to identify assignment of agents to regions which they can locally maximize. The proposed architecture for the distributed adaptive Voronoi tessellation planning is shown in Fig. 4.15. As the figure indicates, agents sample the global environment at their current locations and build up local predictive models, which they share in the network based on the value of information of each model. Once agents receive models of their neighbors, they compare and identify local areas that they should be assigned to.

In order to test the effectiveness of this sensing method, a massive multi-agent simulation algorithm that emulates real time physics of different platforms was utilized. The simulation environment, which we call "MAGE", is described below:

4.6.1 Multiagent game environment

The MAGE (Multi-agents Game Environment) multiplayer online game is a highly scalable simulation system, designed by the Distributed Autonomous Systems Lab at OSU. MAGE can emulate a large scale collaborative sensor network consisting of dynamic as well as static sensors. MAGE is build from various open software like JSBSim, Flightgear and RLPy which provide internal support for running various tasks in the game engine. It is designed as a Massive Multi-player online game, where the players are autonomous intelligent agents that can interact with each other. Therefore, we expect MAGE to be of great use in validating and developing algorithms for monitoring large spatiotemporally varying environments of DOE interest. Currently, MAGE is used for sensor network emulation for a group of decentralized autonomous agents that are trying to build predictive models of various physical quantities (such as hydrocarbon plumes) in an operating environment in order to generate plans for optimal sensor allocation. The architecture for the autonomous sensor agents emulation is described in Fig. 4.16. Agents operate in the central environment that is hosted by one of the players and other players connect to the host. It should be noted that autonomous agents in architecture can be replaced by human players if needed.

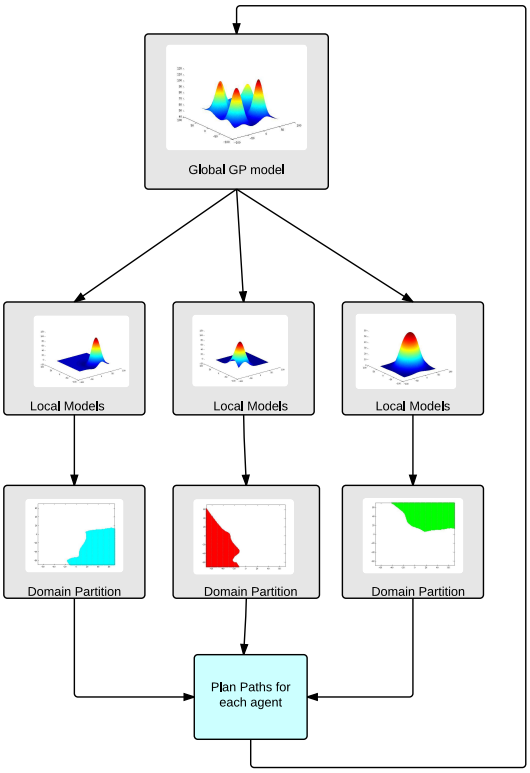


Figure 4.15: Flow chart indicating tessellation of domain with local models

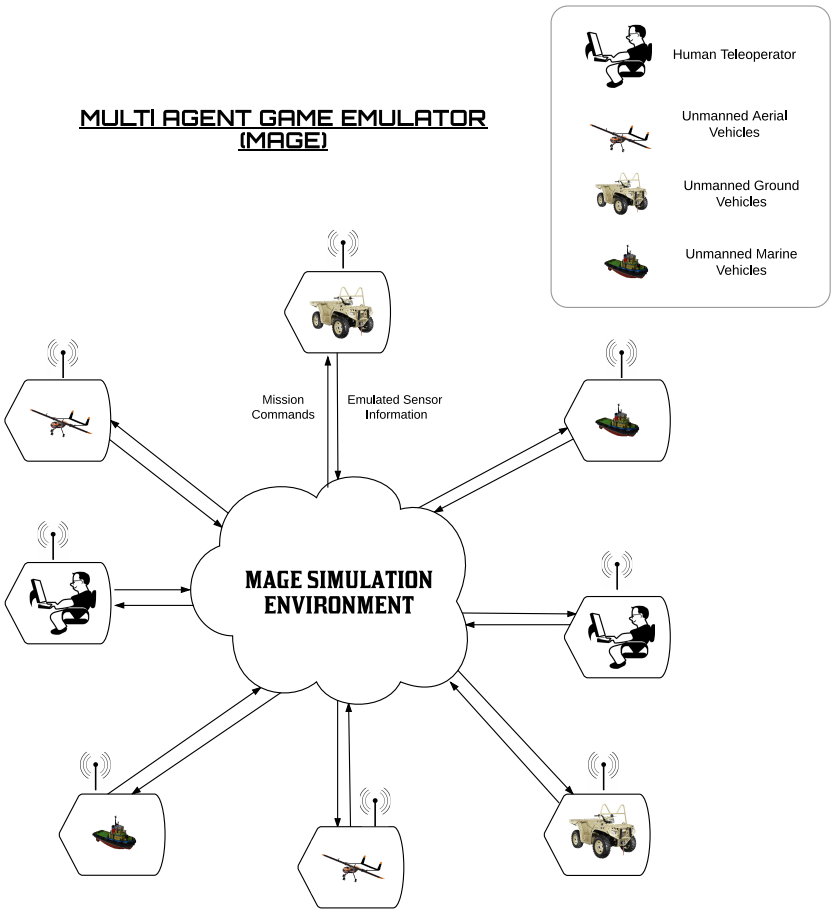


Figure 4.16: Architecture of DAS-MAGE, a Massive Multi-Player online simulation environment for collaborative autonomy research

4.6.2 System optimization through models on expectation of change

UAS allocation for optimal information harvesting

To study the problem, a scenario is considered wherein an endurance constrained UAS (battery or fuel) is ferrying data across a spatially distributed network of unmanned ground sensors (UGS). The objective is to optimize the VOI collected from each sensor by deciding which sensor to query and how often a sensor should be queried. To implement this, a model was developed based on the expected VOI: the VOI at location i during the next time period is modeled as

$$V_i^+ = \begin{cases} \lambda_i, & i \in \eta \\ V_i + \lambda_i, & \text{Otherwise,} \end{cases} \tag{4.10}$$

where λ_i is defined as

$$\begin{aligned} &\lambda_i \sim Pois(\bar{\lambda}_i(t)) \\ &\textstyle\int_{\tau_i^n}^{t} \bar{\lambda}_i(t)dt \sim GP(m_i(t), k_i(t,t')), \end{aligned} \tag{4.11}$$

where $Pois(\cdot)$ is the Poisson Process, and GP is a Gaussian Process with the mean m_i and covariance kernel $k(.,.)$. The resulting architecture is termed, the Cox Gaussian Process (CGP). The key benefit of this model is that the GP can evolve to accommodate a changing number and distribution of UGS in the sensor network. Since the Poisson parameter must be strictly greater than zero, the Fog of War (FOW) described in (Allamraju et al., 2014), which encourages exploration to learn the Poisson parameter, is used as the baseline estimate of $\bar{\lambda}_i(t)$). The predictive covariance of the GP is used to transition estimate from the FOW to the GP estimate of $\bar{\lambda}_i(t)$). A novel data-based variant of the FOW is modeled as

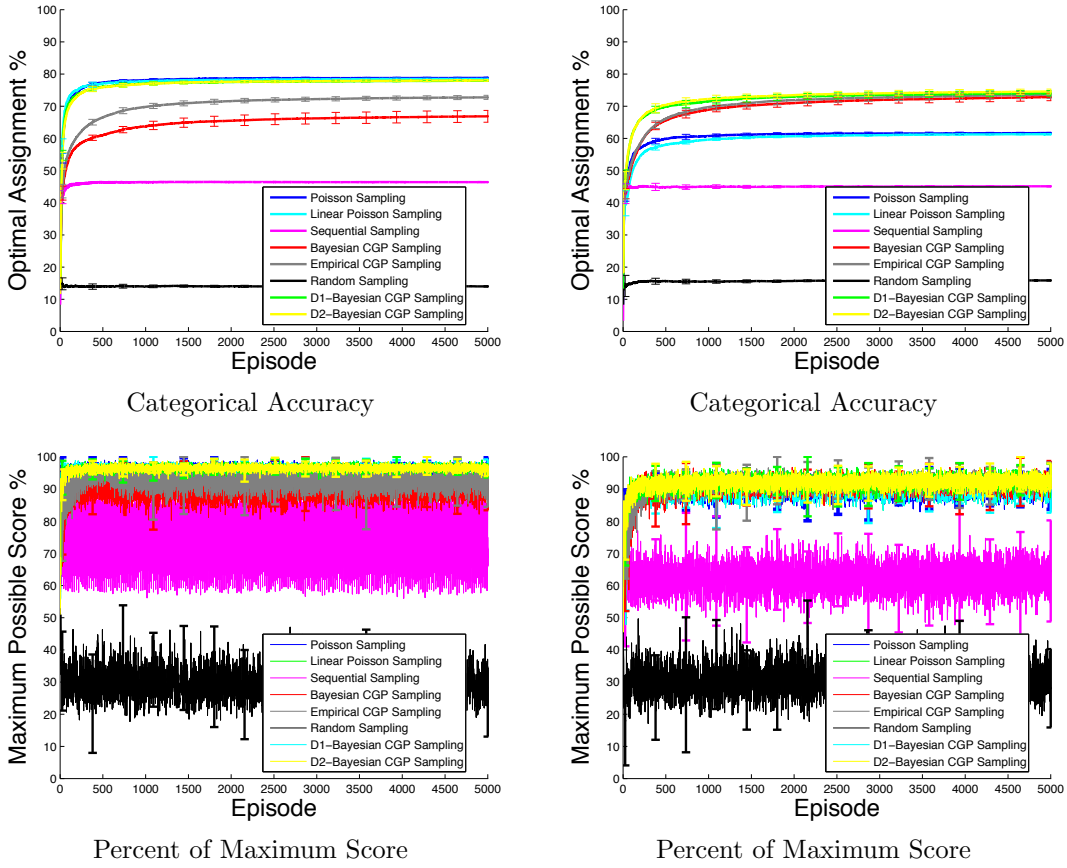
$$\lambda_i \sim Pois(\bar{\lambda}_i(t)), \tag{4.12}$$

where $\bar{\lambda}_i(t)$ is estimated using a homogeneous Poisson processes, thus providing a fully Bayesian CGP for estimating the accumulation of VOI.

4.6.3 Simulation study

Several variations of EIEIO are implemented. The Poisson sampling method assumes that the generative Poisson model is homogeneous. In addition, the linear Poisson sampling method likewise assumes homogeneity, but further assumes that the observed VOI is generated by M independent identically distributed (iid) samples. The empirical Cox process method uses the FOW from (Allamraju et al., 2014) to train a Gaussian process on the difference between the observed VOI and the VOI reported by the FOW. In contrast, the Bayesian Cox process method uses the FOW from (**Allamraju:14:ICRA**) as a baseline so, as the

predictive confidence of the GP increases, the estimation transitions from FOW-reported VOI to GP-estimated VOI. A dual Bayesian Cox process uses a homogeneous Poisson process to implement a data-driven FOW term. The first dual Bayesian Cox process implemented uses the Poisson sampling method. The second dual Baeysian Cox process uses the Linear Poisson sampling method. In the simulation, the aforementioned algorithms were compared to random and sequential sampling as baselines in the following simulation. It is assumed that the UAS, is able to visit a subset of up to 6 UGS per episode; i.e., $\kappa = 6$. A mobile agent first explores the data available from each of the 50 deployed UGS ($K = 50$) due to an initially high assumed data feature arrival rate ($\hat{\lambda}_i = 10, \alpha_i = 10, \beta_i = 1 \; \forall i$). The mobile agent ferries data for a total of 5000 episodes. And then, the graphical results are presented in Fig. 4.17.



Categorical Accuracy

Categorical Accuracy

Percent of Maximum Score

Percent of Maximum Score

Figure 4.17: Figures 4.17a and 4.17b showcase the categorical accuracy of the Poisson-based methods in the presence of significant noise. Likewise, Figs. 4.17c and 4.17d showcase the efficacy of the Poisson-based methods in optimally reducing the available VOI. Notice that although the categorical accuracy is between 70% and 76% for the Poisson-based methods, the entropy reduction of the methods is between 85% and 96% as effective as a system with perfect knowledge.

These results indicate the feasibility of this algorithm for VOI-based sensor network optimization and active querying over a sparsely distributed hydrocarbon sensing networks in the Farnsworth field.

4.6.4 System optimization based on expectation of change models

In order to study the problem of locally changing sensor readings, the Exploitation by Informed Exploration between Isolated Operatives (EIEIO) algorithm that plans paths based on anticipated VOI for data ferrying UAS in a time-correlated non-stationary environment was developed.

4.6.5 Problem definition

The general problem we are interested in is distributed inference and monitoring over a spatio-temporally varying measure y that changes with spatial variable $x \subset \mathbb{R}^2$ and temporal variable $t \in \mathbb{T}$. It is assumed that a set of independent resource constrained unattended ground sensors (UGS) S, indexed by the variable i, provide measurements of y at various locations $x_i \subset \mathbb{R}^2$. The total number of sensors in the network is denoted by K. The measurements at each of these locations is denoted by the random variable y_i, which generates a temporally evolving stochastic process Y_t across all of the nodes. The stochastic process is time-varying, however, its rate of change need not be the same across all nodes. Consequently, not all nodes will have new information at all times. It was assumed that the Value-of-Information, V_i of a node can be captured by an information-theoretic metric such as Kullback-Leibler (KL) divergence or Renyi divergence (Mu, Chowdhary, and J. How, 2014). Note that due to spatio-temporal variations, V_i is a temporally dependent random variable that takes positive values. The UGS can leverage in-situ resources to operate over an extended duration of time, but it is assumed that they have a limited range of communication. Therefore, it is assumed that the ground sensors do not have sufficient power to communicate with a central hub, which leads to clusters of ground sensors that may be able to talk to each other but do not form a completely connected network. Instead, a data-ferrying agent, such as a UAV, is needed to physically ferry the data between the ground sensors. However, the data-ferrying agent itself has a limited endurance, and can only visit a subset $\eta \subset S$ consisting of κ of the ground sensors in any given flight sortie. Qualitatively, the problem we are interested in solving is stated as follows:

Problem 1 Determine the subset of nodes η – the most informative set of UGS to visit in each flight sortie – and a path connecting them, such that the remaining VOI after each flight sortie is minimized.

It is in general difficult to determine η because the expected VOI at each node is not known. Furthermore, without visiting each node it is not possible to glean what the VOI at each node might be. A proactive planning strategy for this problem is to build a model on the expected accumulated VOI at each node, and utilize this model to plan *anticipated change* at each of the sensing locations. In order to cast the problem in a form that can be analyzed, the following assumptions are made. The first assumption simply states that the information being gathered at one node can be assumed to be independent of that being gathered at another node. This is a reasonable assumption if the nodes are placed sufficiently far away from each other.

Assumption 1 The VOI accumulated at each node is assumed to be mutually independent.

It should be noted that mutual dependency occurs either due to spatial dependence, or because the sensors can communicate with each other over a cluster to infer a joint local model of y (Fraser et al., 2012; Mu, Chowdhary, and J. How, 2014). Mutually dependent

UGS may be accounted for by visiting only one UGS in the cluster, allowing each cluster to be effectively treated as one independent UGS. Therefore, the cumulative VOI collected in a flight sortie over the set of nodes η, denoted by $f(\eta)$ can now be represented as a sum over the individual V_i.

$$f(\eta) = \sum_{i \in \eta} V_i. \tag{4.13}$$

Next we make the relatively harmless assumption that the time to transmit data from the node to the UAV can be ignored. This is a reasonable assumption in the sense that depending on the time of transmission, the variable κ denoting the size of the informative set η will change. Furthermore, we assume that the nodes have sufficient memory to store the information from the last visit by the data-ferrying agents:

Assumption 2 The UGS can store and instantaneously transmit sensing history since the last visit to the data-ferrying agent. The UGS can store and transmit sensing history since the last visit to the data-ferrying agent in some finite time that is significantly less than iteration time.

Let τ_i^n denote the n^{th} time that the node i was visited by the data-ferrying agent. It follows that once the information in a node is retrieved, its VOI should be reset to zero:

Assumption 3 The next instantaneous sample at each UGS has no informatic value; i.e.,

$$V_i^+(t) = 0 \quad \forall t = \tau_i^n. \tag{4.14}$$

Using these assumptions, the problem in Problem 1 reduces to an optimization problem. Let Υ denote a binary vector, $c^{(i)}$ denote the time-cost of operating a UGS, and $k^{(i)}$ denote the cost-conversion coefficient. The problem can then be recast as the following mathematical programming problem:

Problem 2 Determine the subset η – the most informative set of UGS to visit in each flight sortie – by solving the following mathematical program

$$\begin{aligned} &\eta(t) = \underset{\Upsilon(t)}{\operatorname{argmin}} \textstyle\sum_{i=1}^{K} \mathbb{E}(V_i(t)) - \mathbb{E}(V_i(t))\Upsilon^{(i)}(t) \\ &s.t. ||\Upsilon(t)||_1 = \kappa, \quad \Upsilon^{(i)}(t) \in \{0,1\}, \quad \gamma \geq \textstyle\sum_{i=1}^{K} \Upsilon^{(i)}(t) c_1^{(i)} k^{(i)}. \end{aligned} \tag{4.15}$$

In order to solve this problem, we need a model that can predict the expected VOI of a node as a function of the time since last visit, and the node's location. The model of the expected VOI must be positively valued, with the exception of simultaneously-collected samples, and must model the expected VOI with respect to the time since last visit to each UGS. A notable random process with strikingly similar constraints is the Poisson process,

and the similarities provide the motive to treat observations of VOI as the likelihood of a Poisson process; i.e.,

$$\mathbb{E}(V_i(t)) = \mathbb{E}(\int_{\tau_i^n}^{t} \sum_{j}^{r} s_j(\lambda_{j,i}(t) - \tilde{\lambda}_{j,i}(t))dt), \tag{4.16}$$

where $\lambda_{j,i}(t)$ and $\tilde{\lambda}_{j,i}(t)$ are positively valued random variables denoting the Poisson arrival and departure rates of an informative event type j at UGS i worth s_j bits per net informative arrival. Note that there are r types of informative events, which relates to the notion of embodied cognition, in that only the types of sensors used may affect what informative events can be observed, and r need not be known. This model is inspired by the queueing-theoretic application of Poisson process priors in determining the arrival rate of packets in communication networks (Karagiannis et al., 2004). Although the Poisson process is a well-known discrete random process, the likelihood of the Poisson process may be a continuous quantity, thus we are permitted a continuous model for the available VOI in a large set of UGS. A general model for spatio-temporally varying Poisson processes is the Cox process. In the Cox process, $\hat{\lambda}_i$ is drawn from a stochastic process. Recall that existing Cox process models require *a priori* output scaling or domain specification (Møller, Syversveen, and Waagepetersen, 1998; Zhou et al., 2014; Adams, Murray, and MacKay, 2009; Gunter et al., 2014). Since an upper bound on λ or the number of UGS may not be known known a-priori, we introduce a new Bayesian Nonparametric model termed *Cox-Gaussian Process (CGP)*, which models the accumulated VOI at a location V_i using a Gaussian process prior:

$$\begin{array}{l} \int_{\tau_i^n}^{t} \sum_{j}^{r} s_j(\lambda_{j,i}(t) - \tilde{\lambda}_{j,i}(t))dt \sim Pois(V_i(t)) \\ V_i(t) \sim GP_i(m_i(t), k_i(t,t')), \end{array} \tag{4.17}$$

where $Pois(\cdot)$ is the Poisson process and $GP_i(\cdot,\cdot)$ is a Gaussian process with the mean m_i and covariance kernel $k(.,.)$. Another key benefit of this model is that the GPs can evolve to accommodate a changing number and distribution of UGS in the sensor network.

4.6.6 Validation on the real-world data sets

Using the Intel Berkeley Research Lab spatio-temporal temperature data set (Bodik et al., 2004) and a concatenated variant of the Global Historical Climatology Network data set, the EIEIO can be tested for viability. The performance of EIEIO on the concatenated data set, where discrete changes in the VOI model is encountered, are shown in Fig. 4.18 and Fig. 4.19. Although the data from the data sets are not Gaussian-distributed, the Central Limit Theorem is used to treat the observations at each UGS as belonging to a Gaussian-distributed likelihood; i.e., $y_i \sim N(\mu, \sigma_1^2)$. For simplicity, each UGS models local observations using the Gaussian distribution as the conjugate prior (Gelman et al., 2013). The Bayesian update is

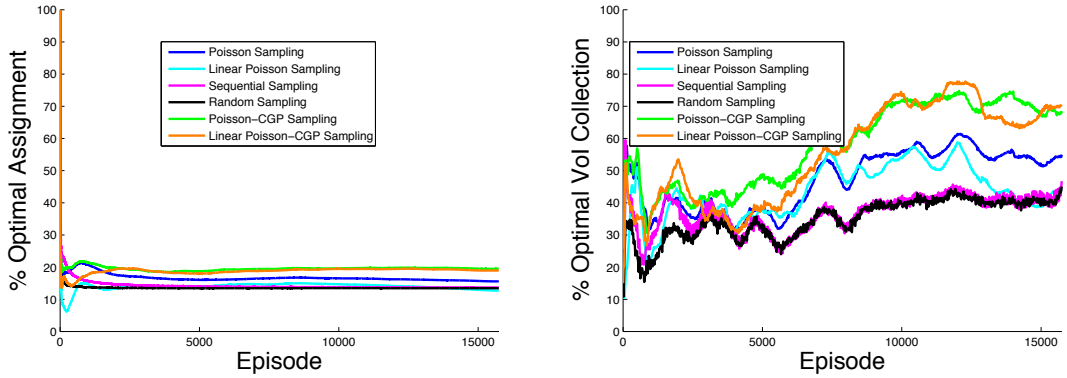
$$\mu_{\hat{q}} = \frac{\frac{\sigma_3^2}{M}\mu_{\hat{p}} + \sigma_{\hat{p},i}^2 \bar{y}_i}{\frac{\sigma_3^2}{M} + \sigma_{\hat{p},i}^2}, \tag{4.18}$$

and

$$\sigma_{\hat{q}}^{-2} = (\frac{\sigma_3^2}{M} + \sigma_{\hat{q}}^2)^{-1}, \tag{4.19}$$

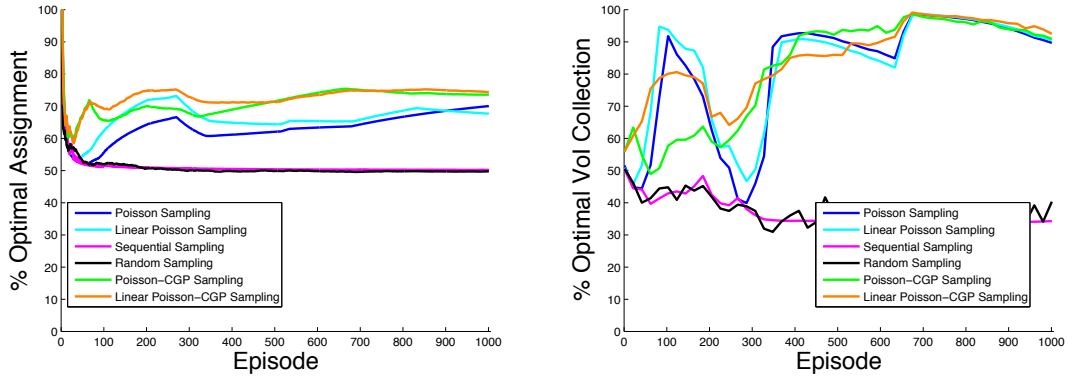
where $\mu_{\hat{p}}$ is the prior mean, $\mu_{\hat{q}}$ is the posterior mean, $\sigma_{\hat{p},i}^2$ is the prior variance, and $\sigma_{\hat{q},i}^2$ is the posterior variance. When the UAS visits each UGS, the local model of the UGS is received and the VOI is calculated using KL divergence (Hershey and Olsen, 2007). The KL divergence, D_{KL}, for scalar normal distributions (i.e., $d = 1$) is

$$D_{KL}(\hat{q}||\hat{p}) = 0.5[log(\frac{\sigma_{\hat{p}}}{\sigma_{\hat{q}}}) + Tr[\sigma_{\hat{p}}^{-1}\sigma_{\hat{q}}] - d + (\mu_{\hat{q}} - \mu_{\hat{p}})^T\sigma_{\hat{p}}^{-1}(\mu_{\hat{q}} - \mu_{\hat{p}})]. \tag{4.20}$$



Categorical Accuracy Percent of Maximum Score

Figure 4.18: The Intel Berkeley Research lab data set using Equation (4.35)(Bodik et al., 2004).



Categorical Accuracy

Percent of Maximum Score

Figure 4.19: The Global Historical Climatology Network dataset test using Equation (4.35)(Lawrimore et al., 2011).

4.6.7 Identifying regions of maximum information

Testing Effectiveness of Algorithms Using a Simulated Platform

Building predictive representations of underlying CO_2 and CH_4 concentration changes using data sampled at various locations. Algorithms in which a distributed network of vehicles can build accurate representations of the CO_2 and CH_4 models through minimal communication between various agents have been developed. Using the MAGE platform outlined previously, a sensor network was immulated to test the effectiveness of the algorithms.

4.6.8 Exploitation of change modeling results

A set of real-world simulations using MAGE were used to experimentally validate the presented solution approaches and each experiment includes four Monte Carlo runs to generate reproducible results. The estimation techniques are applied on the Intel Berkeley data set (Bodik et al., 2004). Experiments were conducted on multiple time-scales, as the Intel Berkeley data set is known to be a slowly evolving system (Mu, Chowdhary, and J. P. How, 2014). The fastest timescale simulates a data ferrying agent that collects data from a subset of sensing locations for thirty-one second intervals, or episodes, across the five days of data collection. The performance of EIEIO on compressed timescales where, for instance, we simulate a data ferrying agent collecting data from a subset of sensors for approximately five minute episodes across the five days of data collection is also examined. Additional timescales to simulate a data ferrying agent collecting data from a subset of sensors for approximately ten minute, 30 minute, and sixty minute episodes across the five days of data collection reported.

The slowest timescale, where each sample is collected in sixty minute intervals, exhibits the noisiest change in the model per time step and 15,740 consecutive samples are used, which is over five days of continuous measurements. It is in this particularly noisy slow timescale, where one episode is carried out each hour, that the baseline methods are expected to be the most competitive with the proposed Poisson process methods. Although the data from the Intel Berkeley Research Lab is not Gaussian-distributed, the Central Limit Theorem can be used to treat the observations at each UGS as belonging to a Gaussian-distributed likelihood; i.e., $y_i \sim N(\mu, \sigma_1^2)$. For simplicity, each UGS models local observations using the Gaussian distribution as the conjugate prior (Gelman et al., 2013). The Bayesian update is

$$\mu_{\hat{q}} = \frac{\frac{\sigma_1^2}{M}\mu_{\hat{p}} + \sigma_{\hat{p},i}^2 \bar{y}_i}{\frac{\sigma_1^2}{M} + \sigma_{\hat{p},i}^2}, \tag{4.21}$$

and

$$\sigma_{\hat{q}}^{-2} = \left(\frac{\sigma_1^2}{M} + \sigma_{\hat{q}}^2\right)^{-1}, \tag{4.22}$$

where $\mu_{\hat{p}}$ is the prior mean, $\mu_{\hat{q}}$ is the posterior mean, $\sigma_{\hat{p},i}^2$ is the prior variance, and $\sigma_{\hat{q},i}^2$ is the posterior variance.

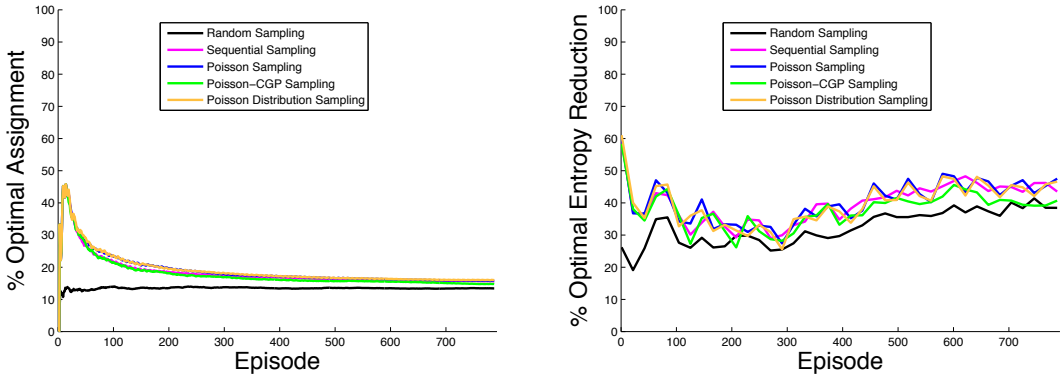
When the UAS visits each UGS, the local model of the UGS is received and the VOI is calculated using KL divergence. The KL divergence, D_{KL}, for scalar normal distributions (i.e., $d = 1$) is

$$D_{KL}(\hat{q}||\hat{p}) = 0.5[log\left(\frac{\sigma_{\hat{p}}^2}{\sigma_{\hat{q}}^2}\right) + Tr\left[\left(\sigma^2{}_{\hat{p}}\right)^{-1}\sigma^2{}_{\hat{q}}\right] - d + (\mu_{\hat{q}} - \mu_{\hat{p}})^T\left(\sigma^2{}_{\hat{p}}\right)^{-1}(\mu_{\hat{q}} - \mu_{\hat{p}})]. \tag{4.23}$$

4.6.9 Comparing proposed solutions to baselines

For both Poisson distribution sampling and EIEIO, two sequential searches are performed to initialize the inferencing technique to learn to predict the KL divergence measures of each sensing location. The Poisson distribution method, where the mean of the observed KL divergence measure is used, has some interesting performance qualities for time-varying data sets. If the initially obtained KL divergence measures are lower than subsequent KL divergence measures, then the Poisson distribution method will fail to visit other sensor locations as evinced in Fig. 4.21. On the other hand, if the initially obtained KL divergence measures are higher than subsequent KL divergence measures, then the Poisson distribution method will perform comparably to sequential sampling as shown in Fig. 4.20; moreover when using a prior of $\alpha_0 = 1,000,000,000$ and $\beta_0 = 1$ for the Poisson process methods, the same result is achieved.

By contrast, in Fig. 4.21 we use $\alpha_0 = 0$ and $\beta_0 = 1$ and find that EIEIO consistently outperforms sequential sampling. It is interesting to note that when the initially sampled KL divergence measures by the Poisson distribution method during two iterations of sequential sampling coincides with the mean of the time series of KL divergence measures observed, then the Poisson distribution method outperforms EIEIO in time-varying environments.



Categorical Accuracy

Percent of Maximum Score

Figure 4.20: The Intel Berkeley Research lab data set (Bodik et al., 2004) using Equation (4.35). The percentages are generated with respect to a decision-making framework with perfect situational awareness. For this simulation, all Poisson-based methods were initialized with a very large prior on the entropy rate.

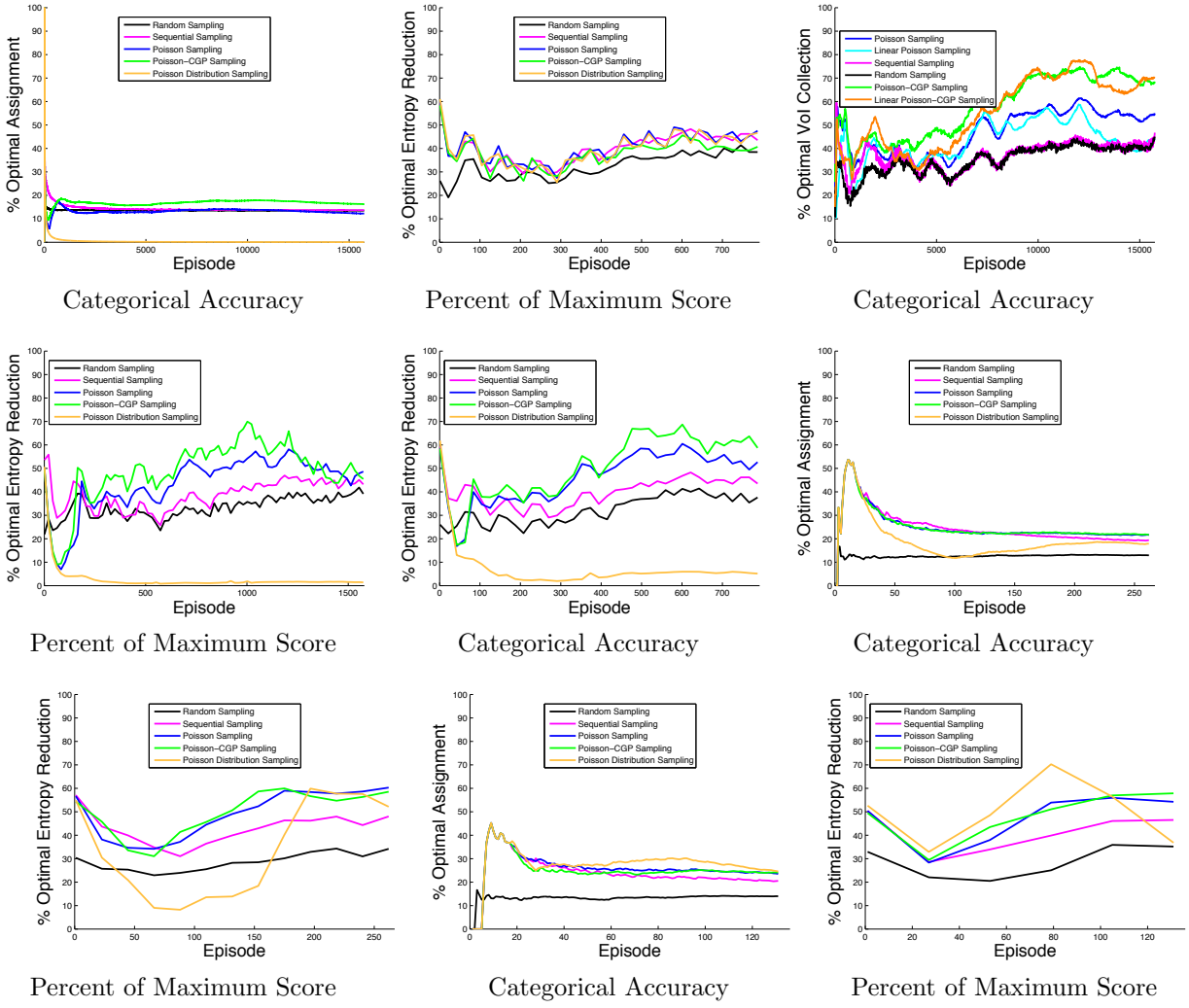


Figure 4.21: Data-ferrying performance at multiple time episode durations (top to bottom: 31 sec, 5 min, 10 min, 30 min & 60 min) using the Intel data is shown.

4.6.10 Identifying regions of maximum information

As discussed previously, algorithms have been developed to maximize accumulated information from a sensor field. This is particularly important in wide area monitoring where access to sensors is difficult and/or time consuming.

Testing the effectiveness of algorithms on a simulation platform

Algorithms in which a distributed network of vehicles can build accurate representations of the CO_2 and CH_4 models through minimal communication between various agents have been developed to minimize the data collection effort while maximizing the usefulness of the data.

Using MAGE (Multi-agent Game Environment) a sensor network which emulates real time physics of different vehicle platforms, provided the platform to test the effectiveness of the algorithms.

4.6.11 Modeling of atmospheric dispersion

Modeling atmospheric dispersion is the cornerstone of finding contamination sources by ground sensors and UAVs. Therefore, the mathematical description of contaminant transport in the atmosphere was investigated and simulated. The term dispersion is used to describe the mixture of diffusion and advection that occurs within the air near the ground surface. Concentrations of a contaminant released into the air is described in (4.29) by the advection-diffusion equation which is a second-order partial differential equation.

$$\frac{\partial C}{\partial t} + \nabla \cdot (C\overrightarrow{u}) = \nabla \cdot (K \nabla C) + S, \tag{4.24}$$

where C is mass concentration, K is diffusion coefficient, $\overrightarrow{u}$ is the wind velocity and S is source or sink term. A solution is presented by the Ermak formulation (4.30) in which it is assumed that contaminant particles are more massive than air and so they tend to settle out of the atmosphere.

$$\begin{aligned} C = \frac{Q}{4\pi u r} \exp\left(\frac{-y^2}{4r}\right) \exp\left(\frac{-\omega_{set}(z-H)}{2K} - \frac{\omega_{set}^2 r}{4K^2}\right) \times \left[\exp\left(\frac{-(z-H)^2}{4r}\right) + \right. \\ \left. \exp\left(\frac{-(z+H)^2}{4r}\right) - \frac{2\omega_0\sqrt{\pi r}}{K} \exp\left(\frac{\omega_0(z+H)}{K} + \frac{\omega_0^2 r}{K^2}\right) + erfc\left(\frac{z+H}{2\sqrt{r}} + \frac{\omega_0\sqrt{r}}{K}\right)\right] \end{aligned} \tag{4.25}$$

where Q is the emission rate, ω_{set} is settling velocity and ω_{dep} is the deposition velocity and $\omega_0 = \omega_{dep} - 0.5\omega_{set}$. Also the stack height is shown by H and r is defined in (4.31).

$$r = \frac{1}{u} \int_0^x K(\zeta) d\zeta \tag{4.26}$$

Results from the Ermak solution are illustrated in Fig. 4.24 for a zinc smelter site (Teck-Cominco Ltd. In Trail, BC, Canada) that shows the total ground-level steady-state contaminant concentration (mg/m^3) of zinc caused by four sources S (red circles) and sensed by nine receptors R (green rectangles). Distribution of contamination is affected by wind with a speed of 1 (m/s) from left to right in the horizontal direction. It is assumed that the emission rate from the sources S1 to S4 are 15, 10, 5, and 5 (kg/s) respectively.

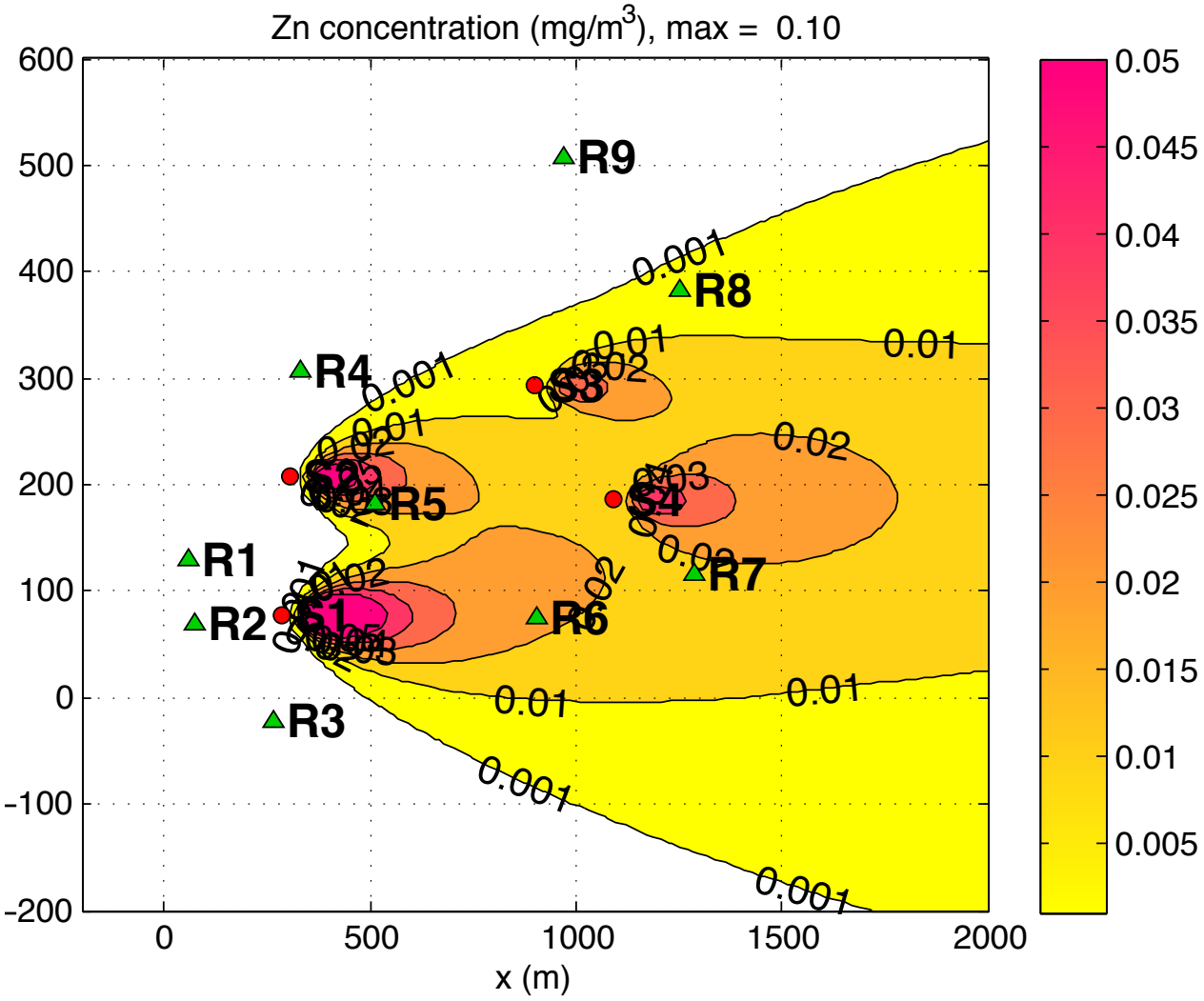


Figure 4.22: Contaminant concentration ($mg/m3$) of zinc caused by four sources S (red circles) and sensed by nine receptors R (green rectangles) modeled by Emark solution.

4.6.12 Learning to predict the time-varying value of information

Let K denote the cardinality of the state space. We assume that the KL divergence at each state i may be modeled as an independent stochastic process $\mathcal{Z}_i$. Random variable $Z_i(\tau_i^n)$ denotes the KL divergence obtained at state i, where τ_i^n is the time of the n$^{\text{th}}$ visit at state i. In each episode the agent visits a subset of the states, which are denoted by $\eta \subset \{1, 2, \ldots, K\}$. The second assumption is that each episode is executed instantly, i.e., all the states in η are

assumed to be visited at the time of the sortie. Finally, to model the dynamic constraints of the agent, we allow the agent to visit at most κ states at each sortie. The objective is to maximize the total expected value of information gained in each sortie. Hence, we aim to solve:

$$\begin{aligned} \eta^* := \arg \max_{\eta \subset \{1,\ldots,K\}} \quad & \mathbb{E}\Big[\sum_{i \in \eta} Z_i\Big] \\ \text{subject to} \quad & \text{Car}(\eta) \leq \kappa, \end{aligned}$$

where $\text{Car}(\eta)$ denotes the cardinality of the index set η. Ideally, the associated cost, say c_i, for visiting the i^{th} state, and ensure that the following constraint is satisfied:

$$\sum_{i \in \eta} c_i \leq C,$$

where C is a constant. For now, we consider the special case when $C = \kappa$ and $c_i = 1$ for all $i \in \{1, 2, \ldots, K\}$ so that the results focus on how well the process $\mathcal{Z}$ is learned and exploited.

Poisson Exposure Process Model Similar to Kim, Nefian, and Broxton, 2009; Kim, Nefian, and Broxton, 2010, it is assumed that the continuous observations (i.e., $z = D_{KL}(\hat{q}||\hat{p}) \sim \mathcal{Z}$) are generated by an unknown monotonic transformation, $g(\cdot)$, of some draw from an unobserved discrete Poisson process. If λ_y were observable, approaches where Poisson processes are used to learn the rate at which informative events (such as the arrival or departure of the target at each grid in a gridworld) occur in the environment, could otherwise be used. Instead, our model is a Hidden Markov Model (HMM) where the state of the HMM is the number of events (such as the number of photons striking a sensor) drawn from an unobserved discrete Poisson process ($Pois$); i.e.,

$$\begin{aligned} y &\sim Pois(\lambda_y) \\ x &\sim \mathcal{X}|y \\ D_{KL}(\hat{q}||\hat{p}) &\sim Pep(\lambda|y), \end{aligned} \tag{4.27}$$

where λ_y is the arrival of unobserved informative events, $\lambda = g(y)$ and Pep is the Poisson exposure process, a continuous variant of the Poisson process.

Learning the Pep describing the available KL divergence allows us to learn about the underlying HMM, which is described as a Poisson-arrival process of informative (i.e., entropy-injecting) events.

Simulation results

Pep is applied on the Intel Berkeley Research Lab spatiotemporal temperature data set Bodik et al., 2004. Although the data from the Intel Berkeley Research Lab is not Gaussian-distributed, the Central Limit Theorem is used to treat the observations at each state as

belonging to a Gaussian-distributed likelihood; i.e., $x_i \sim N(\mu, \sigma_1^2)$. For simplicity, each state is modeled using the Gaussian distribution as the conjugate prior Gelman et al., 2013. The Bayesian update is

$$\mu_{\hat{q}} = \frac{\frac{\sigma_1^2}{M}\mu_{\hat{p}} + \sigma_{\hat{p},i}^2 \bar{y}_i}{\frac{\sigma_1^2}{M} + \sigma_{\hat{p},i}^2} \quad \text{and} \quad \sigma_{\hat{q}}^{-2} = \left(\frac{\sigma_1^2}{M} + \sigma_{\hat{q}}^2\right)^{-1},$$

where $\mu_{\hat{p}}$ is the prior mean, $\mu_{\hat{q}}$ is the posterior mean, $\sigma_{\hat{p},i}^2$ is the prior variance, and $\sigma_{\hat{q},i}^2$ is the posterior variance. The KL divergence, D_{KL}, for scalar normal distributions (i.e., $d = 1$) is

$$D_{KL}(\hat{q}||\hat{p}) = 0.5[log\left(\frac{\sigma_{\hat{p}}^2}{\sigma_{\hat{q}}^2}\right) + tr\left[\left(\sigma^2{}_{\hat{p}}\right)^{-1}\sigma^2{}_{\hat{q}}\right] - d + \left(\mu_{\hat{q}} - \mu_{\hat{p}}\right)^T \left(\sigma^2{}_{\hat{p}}\right)^{-1}\left(\mu_{\hat{q}} - \mu_{\hat{p}}\right)]. \quad (4.28)$$

4.6.13 Modeling of atmospheric dispersion

The term dispersion is used to describe the mixture of diffusion and advection that occurs within the air near the ground surface. The concentration of a contaminant released into the air is described in (4.29) by the advection-diffusion equation which is a second-order partial differential equation.

$$\frac{\partial C}{\partial t} + \nabla \cdot (C\overrightarrow{u}) = \nabla \cdot (K\nabla C) + S, \quad (4.29)$$

where C is mass concentration, K is diffusion coefficient, $\overrightarrow{u}$ is the wind velocity and S is source or sink term. The solution is presented by the Ermak formulation (4.30) in which it is assumed that contaminant particles are more massive than air and so they tend to settle out of the atmosphere.

$$C = \frac{Q}{4\pi ur}\exp\left(\frac{-y^2}{4r}\right)\exp\left(\frac{-\omega_{set}(z-H)}{2K} - \frac{\omega_{set}^2 r}{4K^2}\right) \times [\exp\left(\frac{-(z-H)^2}{4r}\right) + \exp\left(\frac{-(z+H)^2}{4r}\right) - \frac{2\omega_0\sqrt{\pi r}}{K}\exp\left(\frac{\omega_0(z+H)}{K} + \frac{\omega_0^2 r}{K^2}\right) + erfc\left(\frac{z+H}{2\sqrt{r}} + \frac{\omega_0\sqrt{r}}{K}\right)] \quad (4.30)$$

where Q is the emission rate, ω_{set} is settling velocity and ω_{dep} is the deposition velocity and $\omega_0 = \omega_{dep} - 0.5\omega_{set}$. Also the stack height is shown by H and r is defined in (4.31).

$$r = \frac{1}{u} \int_0^x K(\zeta) d\zeta \tag{4.31}$$

The result of Ermak solution is illustrated in Fig. 4.24 for a CO_2 site (Ochiltree County, TX) which shows the total ground level steady state contaminant concentration (mg/m^3) of CO_2 caused by one source (red circles) and sensed by 77 receptors R (green triangles). This contamination is affected by a periodic wind speed of 5 to 10 (m/s) from left to right in the horizontal direction. It is assumed that the emission rate from the source S was 0.0032 (kg/s).

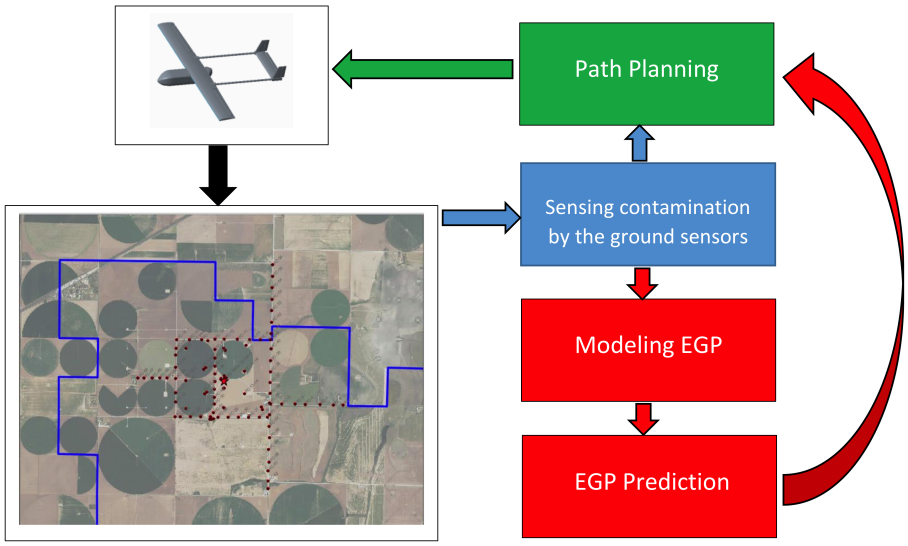


Figure 4.23: Path planning framework to locate CO_2 leakage source

Locating the source of a plume

A framework for locating the source of a contaminate plume as depicted in Fig. 4.23 has been developed. To begin with, the atmospheric dispersion model for denser-than-air-releases by (Ermak, 1990) discussed previously, to model the dispersion of a contaminate leaking from a storage site and forming plumes in the atmosphere. Concentration of the contaminate in space is modeled as a spatial function that evolves in time. And thus, the approach detailed in)Kingravi, 2015) to monitor the evolution of contaminate concentration in space, with time can be used. As a first step, a Kernel Observer model is constructed using the assumption that the evolution of contaminate contamination is known, this forms the training set for the model. Training set was generated for the source location at $(-4000, 0)$ shown in Fig. 4.24.

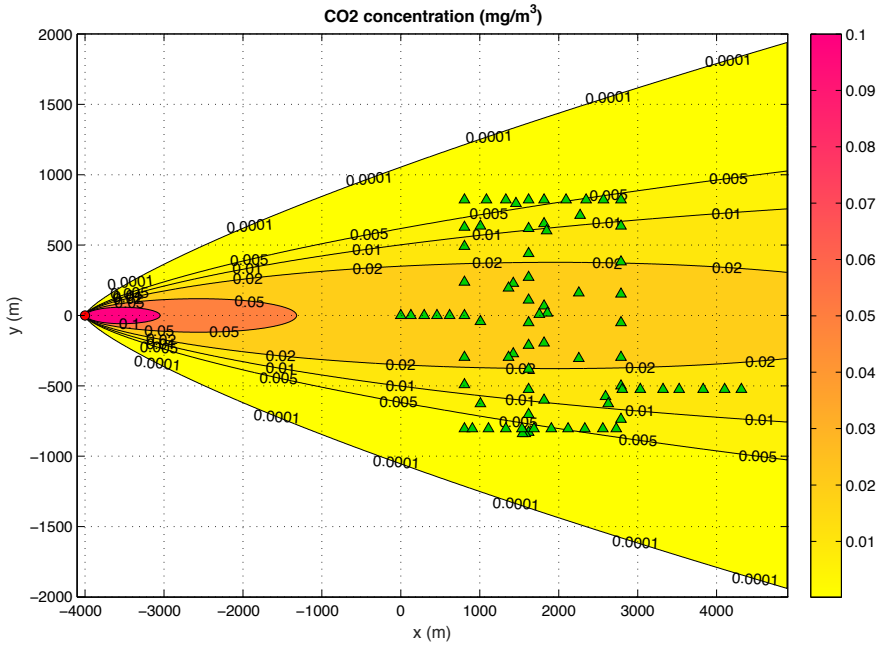


Figure 4.24: Contaminant concentration ($mg/m3$) of CO_2 caused by source S (red circles) and sensed by receptors R (green rectangles).

The Kernel Observer Model captures the evolution trend of the contaminate plume, in the form of a linear transition operator $\hat{A}$. This model can then be used to predict the evolution of a plume with a different source location. This was validated for five different plume sources located at Pos-1 (-1000,500), Pos-2 (300,200), Pos-3 (200,-400), Pos-4 (0,100), and Pos-5 (500,500). The result of the validation is presented in Figs. 4.25 and 4.26. These figures show that the learned model had a good prediction accuracy except for source location at Pos-2. The periodic nature of time-series error is due to the periodicity in wind speed.

The simulation results indicate the possibility of modeling a contaminate plume using Kernel Observer model. The first step of the path planning framework depicted in Fig. 4.23. In the next step of the project, prediction from Kernel Observer model will be integrated with the path planning algorithm (Allamaraju et al., n.d.), to identify the source location.

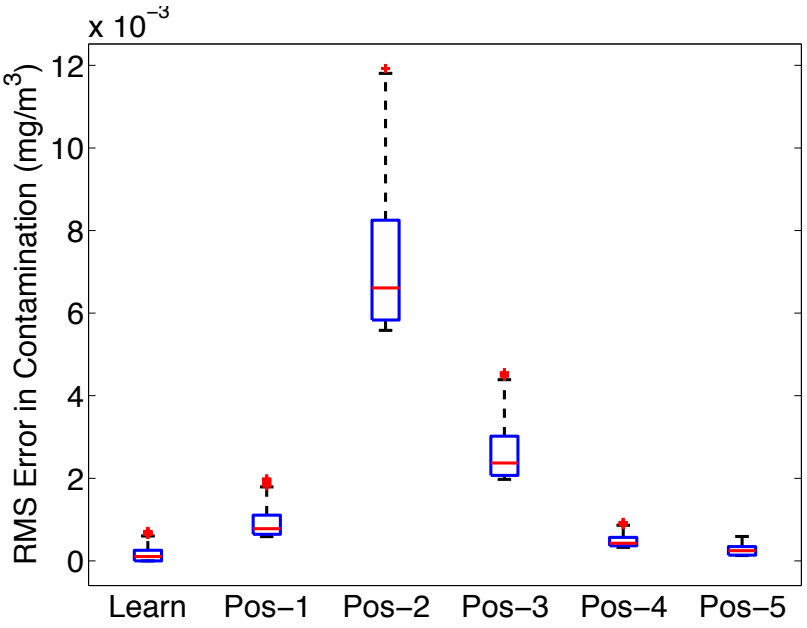


Figure 4.25: Error (Boxplot)

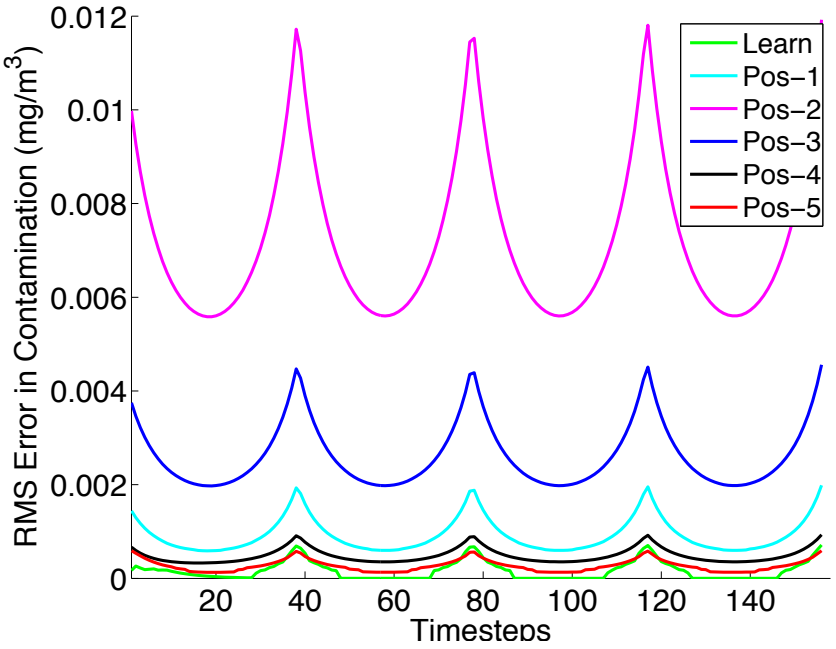


Figure 4.26: Error (timeseries)

4.6.14 Results

The PEP method was applied on the Intel Berkeley Research Lab spatiotemporal temperature data set (Bodik et al., 2004). Although the data from the Intel Berkeley Research Lab is not Gaussian-distributed, the Central Limit Theorem can be used to treat the observations at each state as belonging to a Gaussian-distributed likelihood; i.e., $x_i \sim N(\mu, \sigma_1^2)$. For simplicity, each state is modeled using the Gaussian distribution as the conjugate prior (Gelman et al., 2013). The Bayesian update is

$$\mu_{\hat{q}} = \frac{\frac{\sigma_1^2}{M}\mu_{\hat{p}} + \sigma_{\hat{p},i}^2 \bar{y}_i}{\frac{\sigma_1^2}{M} + \sigma_{\hat{p},i}^2} \quad \text{and} \quad \sigma_{\hat{q}}^{-2} = \left(\frac{\sigma_1^2}{M} + \sigma_{\hat{q}}^2\right)^{-1},$$

where $\mu_{\hat{p}}$ is the prior mean, $\mu_{\hat{q}}$ is the posterior mean, $\sigma_{\hat{p},i}^2$ is the prior variance, and $\sigma_{\hat{q},i}^2$ is the posterior variance.

The KL divergence, D_{KL}, for scalar normal distributions (i.e., $d = 1$) is

$$D_{KL}(\hat{q}||\hat{p}) = 0.5[log\left(\sigma_{\hat{p}}^2 \sigma_{\hat{q}}^{-2}\right) + tr\left[\sigma_{\hat{p}}^2 \sigma_{\hat{q}}^{-2}\right] - d + (\mu_{\hat{q}} - \mu_{\hat{p}})^T \sigma^{-2}{}_{\hat{p}} (\mu_{\hat{q}} - \mu_{\hat{p}})]. \quad (4.32)$$

Implementation in embedded hardware However, a problem of real-time scalability emerges for conventional processors as the number of independent sensing locations increases. Fortunately, the independence condition makes the problem amenable to real-time parallel processing for our decision-making problem. Hence, we will calculate the information dynamics of many independent sensing locations in a time-efficient manner using the Jetson TK1 mobile embedded system board, developed by NVIDIA. The Jetson TK1 board has an on-board graphics processing unit and adds a manageable 142 g to our UAV payload. The low payload of the board, in addition to relatively low power consumption and parallel processing capability, makes the Jetson TK1 is ideal for our application.

4.7 Modeling Using an Evolving Gaussian Process Algorithm

A new class of machine learning algorithms termed the Evolving Gaussian Process (E-GP) that were designed specifically to model spatiotemporally evolving dynamic phenomena that are particularly amenable to Bayesian inference of the latent state of spatiotemporal processes with very few sensor measurements. Results indicate the feasibility of utilizing very few sensor measurements to quickly determine the source of a contaminate dispersion, furthermore, the results also provide direct guidance on UAS mission planning, by indicating where measurements need to be made.

4.7.1 Learning to Predict the Time-varying Value of Information

Formulating the proactive sensor scheduling problem as a multi-play n-armed restless bandit problem provides a route to a solution. The degradation of model-fidelity as a function of time is modeled using the Poisson exposure process (PEP).

Poisson Exposure Process Model

Definition 1
The Poisson exposure process (PEP) is defined as

$$f(z|\Lambda(t)) = C_{\Lambda(t)} \frac{(\Lambda(t))^z e^{-\Lambda(t)}}{\Gamma(z+1)}, \tag{4.33}$$

where $\Lambda(t) = \lambda t$ is a homogeneous PEP with an exposure rate of λ, C_λ is the normalizing constant, and $Z(\tau^i)$ is termed the i^{th} Poisson exposure trial. When $\Lambda(t) \neq \lambda t \; \forall \; t$, the PEP is termed inhomogeneous. We term the PEP as a Poisson exposure distribution (PED) when $\Lambda(t)$ is a constant, as in (Kim, Nefian, and Broxton, 2009; Kim, Nefian, and Broxton, 2010).

Probabilistic guarantees on the accuracy of our PEP regression were theoretically derived and used to dictate when actions to bolster model-fidelity are competently warranted.

Theorem 1 (Exposure Inequality)
Let $\Delta Z(\tau_1), ..., \Delta Z(\tau_n)$ be independent Poisson exposure process increments so that $Pr(\Delta Z(\tau_i)) = p_i$. Let $\bar{Z}(\tau_n) = \frac{\sum_{j=1}^{n} \Delta Z(\tau_j)}{n\tau_n}$ and $\lambda = \mathbb{E}[\bar{Z}(\tau_n)]$. Then,

$$Pr\left(\left|\bar{Z} - \lambda\right| \geq \lambda^{\frac{3}{4}}\right) \leq \frac{1}{n\beta\sqrt{\lambda}}. \tag{4.34}$$

4.7.2 Results

Using the Intel Berkeley Research Lab spatiotemporal temperature data set (Bodik et al., 2004) to test the methods. Although the data from the Intel Berkeley Research Lab is not Gaussian-distributed, the Central Limit Theorem is used to treat the observations at each state as belonging to a Gaussian-distributed likelihood; i.e., $x_i \sim N(\mu, \sigma_1^2)$. For simplicity, each state is modeled using the Gaussian distribution as the conjugate prior (Gelman et al., 2013). The Bayesian update is

$$\mu_{\hat{q}} = \frac{\frac{\sigma_1^2}{M}\mu_{\hat{p}} + \sigma_{\hat{p},i}^2 \bar{y}_i}{\frac{\sigma_1^2}{M} + \sigma_{\hat{p},i}^2} \quad \text{and} \quad \sigma_{\hat{q}}^{-2} = \left(\frac{\sigma_1^2}{M} + \sigma_{\hat{q}}^2\right)^{-1},$$

where $\mu_{\hat{p}}$ is the prior mean, $\mu_{\hat{q}}$ is the posterior mean, $\sigma^2_{\hat{p},i}$ is the prior variance, and $\sigma^2_{\hat{q},i}$ is the posterior variance. The information gain, uniquely denoted by the Kullback-Leibler (KL) divergence $D_{KL}(\text{Posterior}||\text{Prior})$ (Hobson, 1969), for scalar normal distributions (i.e., $d = 1$) is

$$D_{KL}(\hat{q}||\hat{p}) \quad = \quad 0.5[log\left(\sigma^2_{\hat{p}}\sigma^{-2}_{\hat{q}}\right) \; + \; tr\left[\sigma^2_{\hat{p}}\sigma^{-2}_{\hat{q}}\right] \; - \; d \; + \; \left(\mu_{\hat{q}} - \mu_{\hat{p}}\right)^T \sigma^{-2}{}_{\hat{p}} \left(\mu_{\hat{q}} - \mu_{\hat{p}}\right)]. \quad (4.35)$$

The PEP to model the growth in the expected information gain as a function of the time between visits to a sensor location. We summarize the end-results of 50 experiments on the Intel Berkeley Dataset.

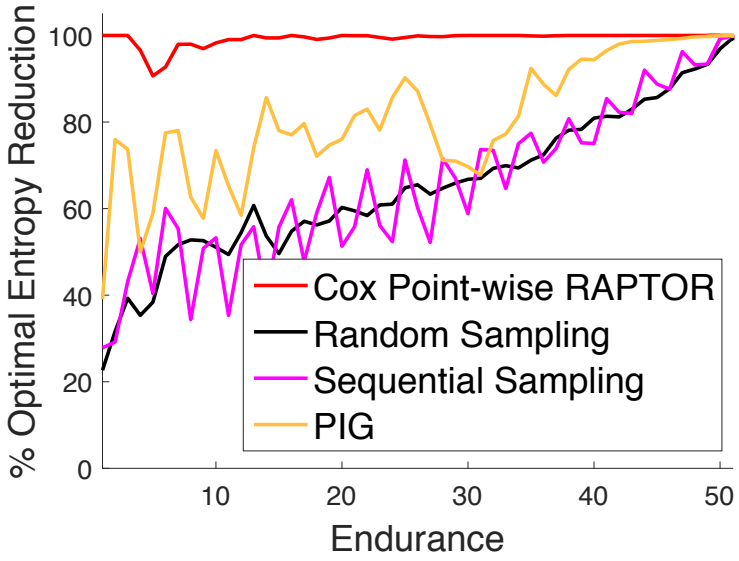


Figure 4.27: Intel Temperature Data

4.7.3 Implementation in embedded hardware

However, a problem of real-time scalability emerges for conventional processors as the number of independent sensing locations increases. Fortunately, the independence condition makes the problem amenable to real-time parallel processing for the decision-making problem. This requires the addition of processing power to the UAV in order to provide real-time route updating. Fortunately NVIDIA makes a mobile embedded system that is capable of calculating the information dynamics of many independent sensing locations in a time-efficient manner. The Jetson TK1 board has an on-board graphics processing unit and adds a manageable 142 g to our UAV payload. The low payload weight of the board, in addition to relatively low power consumption and parallel processing capability, makes the Jetson TK1 ideal for this application. Implementation of this subsystem is beyond the scope of the current project.

4.8 Summary

When the proposal for the overall project was being prepared, it was determined that it would be necessary to collect data from each of the sensor nodes either by visiting each node manually or by using a UAV to collect the data. Because the choice of the UAV was less labor intensive, the work described in this section was proposed. After the proposal was selected for funding late in the design phase it was determined that using wireless technology and cell modems, the data collection process could be automated without seriously impacting the cost per node. As a backup, this task was continued until it reached the point that it could be implemented. By that time, the wireless communication system was working, so the task was brought to a logical conclusion and ended.

5 Summary and Conclusions

The main goal of the project was to show that low cost carbon dioxide and methane sensors could be used to construct a sensor array capable of detecting leaks at a CO_2 injection and storage site. Orginaly, the proposal specified a sensor array that stored data at each node and required periodic manual retrieval. During the design stage it was determined that for a small extra cost per unit, node to node communication could be added. This coupled with the development of a communication node that collectsdata from the sensor nodes and sends the data to a remote computer via a cellular modem. The communication capability lowers the manpower needed to monitor the site and provides near realtime access to the data. In an industrial situation, the lower manpower requirement coupled with rapid access to data provides a significant benefit.

A second goal was to design and test an airborne system using inexpensive sensors and a small unmanned aerial platform. The most difficult problem that had to be overcome was creating an adequate airflow through the sensors to minimize the lag time between readings. While an idea solution has not been found, the lag time was minimized by modifying the flow system. This made correcting the sampling location relatively straightforward. Testing of the UAV system has been carried out at the flight center, Farnsworth, and over controlled burns. Flight center testing was carried out with point sources of carbon dioxide. Testing at the Farnsworth site was somewhat problematic due to the sporadic nature of the carbon dioxide releases (Fig. 2.27). None of the observed carbon dioxide events lasted long enough to get to the Farnsworth site from Stillwater in time to make measurements. Testing at the flight center and over controlled burns proved to be a much better measure of success.

For the first time, the fracture network at the Farnsworth site has been characterized. This information is important for determining the potential for leaks and the locations of where the leaks are most likely to occur. When CO_2 injection in the Farnsworth field expands to the whole field, this information will be necessary to determine the optimal placement of sensor nodes and design of flight patterns.

This project set out to prove that low cost carbon dioxide and methane sensors could be used to detect short and long term gas releases. While only short-term releases were observed, the sensors proved to be robust and sufficiently accurate. Leak detection is a much different problem than that involved in measuring carbon dioxide levels for scientific purposes. Leaks that have a geologic origin will result in a long-term baseline shift, while leaks caused from equipment failure will have much shorter duration. The units deployed at the Farnsworth

site proved that they were capable of detecting short-term leaks. Geologic leaks should be easier to identify since a major baseline shift should result. With a wider deployment and shorter sampling intervals, it might be possible to narrow down the location of a leak.

List of Figures

List of Tables

Bibliography

(2011). URL: http://edgar.jrc.ec.europa.eu/overview.php?v=42.

Adams, Ryan Prescott, Iain Murray, and David JC MacKay (2009). "Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities". In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, pp. 9–16.

Ahlers, S., G. Müller, and T. Doll (2005). "A Rate Equation Approach to the Gas Sensitivity of Thin Film Metal Oxide Materials". In: *Sensors and Actuators B: Chemical* 107.2, pp. 587–599. ISSN: 0925-4005. DOI: 10.1016/j.snb.2004.11.020.

Allamaraju, Rakshit et al. (n.d.). "Human Aware UAS Path Planning in Urban Environments using Nonstationary MDPs". In:

Allamraju, Rakshit et al. (2014). "Human Aware Path Planning in Non Stationary Urban Environment using Model Based Reinforcement Learning". In: *International Conference of Robotic Automation*. submitted.

Ball, Mahlon M., Mitchell E. Henry, and Sherwood E. Frezon (1991). *Petroleum Geology of the Anadarko Basin Region, Province (115), Kansas, Oklahoma, and Texas*. USGS Survey 88-450W. USGS.

Bamberger, I. et al. (2014). "Spatial Variability of Methane: Attributing Atmospheric Concentrations to Emissions". In: *Environmental Pollution* 190, pp. 65–74. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2014.03.028.

Barsan, N., Markus Schweizer-Berberich, and Wolfgang Göpel† (1999). "Fundamental and Practical Aspects in the Design of Nanoscaled SnO2 Gas Sensors: A Status Report". In: *Fresenius' Journal of Analytical Chemistry* 365.4, pp. 287–304. ISSN: 1432-1130. DOI: 10.1007/s002160051490.

Black, R.R. et al. (2012). "Formation of Artefacts While Sampling Emissions of PCDD/PCDF from Open Burning of Biomass". In: *Chemosphere* 88.3, pp. 352–357. ISSN: 0045-6535. DOI: 10.1016/j.chemosphere.2012.03.046.

Blasing, T (2016). *Recent Greenhouse Gas Concentrations*. Tech. rep. U. S. Department of Energy, CDIAC.

Bodik, Peter et al. (2004). *Intel Lab Data*. Tech. rep.

Busoniu, Lucian et al. (2010). *Reinforcement Learning and Dynamic Programming Using Function Approximators*. CRC Press.

Carter, L.S. et al. (1998). "Heat flow and thermal history of the Anadarko basin,Oklahoma". In: *AAPG Bulletin* 82, pp. 291–316.

Chayes, F. (1949). "A simple point counter for thin section analysis". In: *American Mineralogist* 34, pp. 1–11.

— (1956). *Petrology modal analysis*. New York: John Wiley and Sons Inc.

Chiu, Shih-Wen and Kea-Tiong Tang (2013). "Towards a Chemiresistive Sensor-Integrated Electronic Nose: A Review". In: *Sensors* 13.10, pp. 14214–14247. ISSN: 1424-8220. DOI: 10.3390/s131014214.

Chung, Wan-Young and Seung-Chul Lee (2008). "A Selective AQS System with Artificial Neural Network in Automobile". In: *Proceedings of the Eleventh International Meeting on Chemical Sensors IMCS-11IMCS 2006IMCS 11* 130.1, pp. 258–263. ISSN: 0925-4005. DOI: 10.1016/j.snb.2007.07.138.

Csató, Lehel and Manfred Opper (2002). "Sparse on-line Gaussian processes". In: *Neural computation* 14.3, pp. 641–668.

Cunningham, B. J. (1961). "Stratigraphy of Oklahoma Texas panhandles:oil and gas fields of the Texas and Oklahoma panhandles Amarillo". In: *Panhandle Geological Society*, pp. 45–60.

Currie, L. A. (1997). "Detection: International Update, and Some Emerging Di-Lemmas Involving Calibration, the Blank, and Multiple Detection Decisions". In: *Chemometrics and Intelligent Laboratory Systems* 31.1, pp. 151–181. DOI: https://doi.org/10.1016/S0169-7439(97)00009-9.

Dlugokencky, Ed (2016). *Trends in Atmospheric Methane*. Tech. rep. NOAA/ESRL.

Dlugokencky, Ed and Pieter Tans (2016). *Trends in Atmospheric Carbon Dioxide*. Tech. rep. NOAA/ESRL.

Energy(DOE) National Energy Technology Transfer (NETL), US Dept. of (n.d.). *NATCARB Sources(v1303)*. English. Online. Vector digital data. Version 1303. US Dept. of Energy(DOE). URL: http://www.netl.doe.gov/technologies/carbon_seq/natcarb (visited on).

Ermak, D.L. (1990). *User's manual for SLAB: An atmospheric dispersion model for denser-than-air-releases*. Technical Report UCRL-MA-105607 ON: DE91008443. Lawrence Livermore National Lab., CA (USA).

Eugster, W. and G. W. Kling (2012). "Performance of a Low-Cost Methane Sensor for Ambient Concentration Measurements in Preliminary Studies." In: *Atmospheric Measurement Techniques* 5.Copyright (C) 2015 American Chemical Society (ACS). All Rights Reserved. Pp. 1925–1934. ISSN: 1867-1381. DOI: 10.5194/amt-5-1925-2012.

Folk, R (1980). *Petrology of Sedimentary Rocks*. Austin,Texas: Hemphill.

Fraser, Cameron S.R. et al. (2012). "A Hyperparameter Consensus Method for Agreement Under Uncertainty". In: *Automatica* 48.2, pp. 374–380.

Gelman, Andrew et al. (2013). *Bayesian data analysis*. CRC press.

Grande, Robert, Girish Chowdhary, and Jonathan How (2013). "Online Nonstationary Model Detection and Learning using NonBayesian Hypothesis Tests". In: *International Conference on Machine Learning*. submitted.

Guerriero, V. et al. (2010). "Quantifying uncertainties in multi-scale studies of fractured reservoir analogues:implemented statistical analysis of scan line data". In: *Journal of Structural Geology* 32, pp. 271–278.

Gunter, Tom et al. (2014). "Efficient Bayesian Nonparametric Modelling of Structured Point Processes". In: *arXiv preprint arXiv:1407.6949*.

Guohua, Hui et al. (2012). "Study of Grass Carp (Ctenopharyngodon Idellus) Quality Predictive Model Based on Electronic Nose". In: *Sensors and Actuators B: Chemical* 166–167, pp. 301–308. ISSN: 0925-4005. DOI: 10.1016/j.snb.2012.02.066.

Gustavson, T.C. (1986). "Geomorphic development of the Canadian River Valley,Texas Panhandle:an example of regional salt dissolution and subsidence". In: *GSA Bulletin* 97, pp. 459–472.

Gustavson, T.C. and D.A.Winkler (1988). "Depositional facies of the Miocene Pliocene Ogallala Formation, northwestern Texas and eastern New Mexico". In: *Geology* 16, pp. 203–206.

Gustavson, T.C. and R.T.Budnik (1985). "Structural influences on geomorphic processes and physiographic features,Texas Panhandle:Technical issues in siting a nuclear-waste repository". In: *Geology* 13, pp. 173–176.

Gustavson, T.C. and V.T.Holliday (1999). "Eolian sedimentation and soil development on a semiarid to subhumid grassland,Tertiary Ogallala and Quaternary Blackwater Draw Formations,Texas and New Mexico High Plains". In: *Journal of Sedimentary Research* 69, pp. 622–634.

Hershey, John R and Peder A Olsen (2007). "Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models." In: *ICASSP (4)*, pp. 317–320.

Ho, Clifford K. et al. (2001). *Review of Chemical Sensors for Real-Time In-Situ Sensing*. Tech. rep. SAND2001-0643. Sandia National Laboratories.

Hobson, A. (1969). "A new theorem of information theory". In: *Journal of Statistical Physics* 1.3, pp. 383–391.

Hooker, J. N., S. E. Laubach, and R. Marrett (2013). "Fracture-aperture size – frequency, spatial distribution, and growth processes in strata-bounded and non-strata-bounded fractures, Cambrian Mesón Group, NW Argentina". In: *Journal of Structural Geology* 54, pp. 54–71.

Intel (2014). *Mica2Dot Sensors with weather boards*. Online. URL: http://db.csail.mit.edu/labdata/labdata.html (visited on).

Johnson, K. S. et al. (1989). "Geology of the southern midcontinent". In: *Geological Survey Special Publication* 89.2, p. 53.

Karagiannis, Thomas et al. (2004). "A nonstationary Poisson view of Internet traffic". In: *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*. Vol. 3. IEEE, pp. 1558–1569.

Karunanithi, S. et al. (2009). "Performance of Labscale Solar Powered Wireless Landfill Monitoring System". In: *Energy and Environment, 2009. ICEE 2009. 3rd International Conference On*, pp. 443–448. DOI: 10.1109/ICEENVIRON.2009.5398610.

Kennedy, C. L. et al. (1982). *The deep Anadarko basin*. Tech. rep. Tulsa, OK: Petroleum Information Corporation.

Kim, Taemin, Ara V Nefian, and Michael J Broxton (2009). "Photometric Recovery of Ortho-Images Derived from Apollo 15 Metric Camera Imagery". In: *Advances in Visual Computing*. Springer, pp. 700–709.

Kim, Taemin, Ara V Nefian, and Michael J Broxton (2010). "Photometric recovery of Apollo metric imagery with Lunar-Lambertian reflectance". In: *Electronics letters* 46.9, pp. 631–633.

Kingravi, Hassan A. (2015). "Kernel controllers: A systems-theoretic approach for data-driven modeling and control of spatiotemporally evolving processes". In: *Decision and Control (CDC), 2015 IEEE 54th Annual Conference*. IEEE. IEEE.

Kintel, M and C Clifford Wolf (2011). "OpenSCAD, The Programmers Solid 3D CAD Modeller". In:

Klusman, R. W. (2005). "Baseline studies of surface gas exchange and soil-gas composition in preparation for CO2 sequestration research: Teapot Dome, Wyoming". In: *AAPG Bulletin* 89, pp. 981–1003.

Krause, R.G. and T.A.M. Geijer (1987). "An improved method for calculating the standard deviation and variance of paleocurrent data". In: *Journal of Sedimentary Petrology* 57, pp. 779–780.

KSWO (2015). *Weather History for KSWO from July 1, 2014 to July 31, 2015 [Dataset]*.

— (2016). *Weather History for KSWO from January 1, 2016 [Dataset]*.

Lai, D. Y. F. et al. (2012). "The effect of atmospheric turbulence and chamber deployment period on autochamber CO_2 and CH_4 flux measurements in an ombrotrophic peatland". In: *Biogeosciences* 9.8, pp. 3305–3322. DOI: 10.5194/bg-9-3305-2012. URL: http://www.biogeosciences.net/9/3305/2012/.

Lawrimore, Jay H et al. (2011). "An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3". In: *Journal of Geophysical Research: Atmospheres (1984–2012)* 116.D19.

Long, Gary L. and J. D. Winefordner (1983). "Limit of Detection A Closer Look at the IUPAC Definition". In: *Analytical Chemistry* 55.07, 712A–724A. ISSN: 0003-2700. DOI: 10.1021/ac00258a724.

M. Pavone E. Frazzoli, F. Bullo (2011). "Adaptive and Distributed Algorithms for Vehicle Routing in a Stochastic and Dynamic Environment". In: *IEEE Transactions on Automatic Control* 56.8, pp. 1834–1848.

McGinn, S. M. et al. (2014). "Cattle Methane Emission and Pasture Carbon Dioxide Balance of a Grazed Grassland". English. In: *Journal of Environmental Quality* 43.3, pp. 820–828. DOI: 10.2134/jeq2013.09.0371. URL: http://dx.doi.org/10.2134/jeq2013.09.0371.

Mehta, S., A.E. Fryar, and J.L. Banner (2000). "Controls on the regional-scale salinization of the Ogallala aquifer, Southern High Plains, Texas, USA". In: *Applied Geochemistry* 15, pp. 849–864.

Mocak J. et al. (2009). "A Statistical Overview of Standard (IUPAC and ACS) and New Procedures for Determining the Limits of Detection and Quantification: Application to Voltammetric and Stripping Techniques (Technical Report)". In: *Pure and Applied Chemistry* 69.2, p. 297. ISSN: 13653075. DOI: 10.1351/pac199769020297. (Visited on 01/03/2017).

Modeling, Environmental (2012). *Berlin Heidelberg: Springer-Verlag*. Berlin Heidelberg: Springer-Verlag.

Møller, Jesper, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen (1998). "Log gaussian cox processes". In: *Scandinavian journal of statistics* 25.3, pp. 451–482.

Mu, Beipeng, Girish Chowdhary, and Jonathan How (2014). "Efficient distributed sensing using adaptive censoring-based inference". In: *Automatica*.

Mu, Beipeng, Girish Chowdhary, and Jonathan P How (2014). "Efficient distributed sensing using adaptive censoring-based inference". In: *Automatica*.

Munson, T. W. (1990). *Farnsworth (Upper Morrow) Field*. Ed. by Field book committee. Panhandle (Texas) Geological Society.

Nelson, R. A. (2001). *Geologic Analysis of Naturally Fractured Reservoirs*. Boston, MA: Gulf Professional Publishing.

Neri, Giovanni (2015). "First Fifty Years of Chemoresistive Gas Sensors". In: *Chemosensors* 3.1, pp. 1–20. ISSN: 2227-9040. DOI: 10.3390/chemosensors3010001.

NETL (2012). *Basin-Scale Leakage Risks from Geologic Carbon Sequestration: Impact on Carbon Capture and Storage Energy Market Competitiveness*. Project Fact Sheet DE-FE0000749 DE-FE0000749. U. S. Department of Energy, National Energy Technology Laboratory.

— (2013). *Southwest Regional Partnership Farnsworth Unit EOR Field Project Development Phase*. Project Fact Sheet FC26-05NT42591. U. S. Department of Energy, National Energy Technology Laboratory.

Ortega, O.J., R.A. Marrett, and S.E. Laubach (2006). "A scale-independent approach to fracture intensity and average spacing measurement". In: *AAPG* 90.2, pp. 193–208.

Pavone, Marco et al. (2009). "Equitable partitioning policies for robotic networks". In: *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, pp. 2356–2361.

Pering, T.D. et al. (2014). "High Time Resolution Fluctuations in Volcanic Carbon Dioxide Degassing from Mount Etna". In: *Journal of Volcanology and Geothermal Research* 270, pp. 115–121. ISSN: 0377-0273. DOI: 10.1016/j.jvolgeores.2013.11.014.

Rasmussen, C.E. and C.K.I. Williams (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press. ISBN: 026218253X.

Reeves, C.C. (1970). "Origin, classification, and geologic history of caliche on the southern high plains, Texas and eastern New Mexico". In: *Journal of Geology* 78, pp. 352–362.

— (1972). "Tertiary-Quaternary stratigraphy and geomorphology of west Texas and southeastern New Mexico". In: ed. by V. Kelley and F.D. Trauger. Vol. 24. New Mexico Geological Society Guidebook, Guidebook for east-central New Mexico. New Mexico Geological Society.

Seni, S. J. (1980). *Sand-body geometry and depositional systems, Ogallala Formation, Texas*. Report of Investigations 105. exas Bureau of Economic Geology.

Shendell, Derek G. et al. (2012). "Outdoor Near-Roadway, Community and Residential Pollen, Carbon Dioxide and Particulate Matter Measurements in the Urban Core of an Agricultural Region in Central CA". In: *Atmospheric Environment* 50, pp. 103–111. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2011.12.056.

Solis, J. L. et al. (2005). "Fluctuation-Enhanced Multiple-Gas Sensing by Commercial Taguchi Sensors". In: *IEEE Sensors Journal* 5.6, pp. 1338–1345. ISSN: 1530-437X. DOI: 10.1109/JSEN.2005.857882.

Somov, Andrey et al. (2013). "Deployment and Evaluation of a Wireless Sensor Network for Methane Leak Detection". In: *Selected Papers from the 26th European Conference on Solid-State Transducers Kraków, Poland, 9-12 September 2012* 202, pp. 217–225. ISSN: 0924-4247. DOI: 10.1016/j.sna.2012.11.047.

Sutton, R. and A. Barto (1998). *Reinforcement learning: An introduction*. MIT Press.

Tarnawski, M. G. et al. (1994). "Diel changes of atmospheric CO2 concentration within, and above, cryptogam stands in a New Zealand temperate rainforest". In: *New Zealand Journal of Botany* 32.3, pp. 329–336. ISSN: 0028-825X. DOI: 10.1080/0028825X.1994.10410474. URL: http://dx.doi.org/10.1080/0028825X.1994.10410474.

Turner, A. J. et al. (2016). "A Large Increase in U.S. Methane Emissions over the Past Decade Inferred from Satellite Data and Surface Observations". In: *Geophysical Research Letters* 43.5, pp. 2218–2224. ISSN: 1944-8007. DOI: 10.1002/2016GL067987.

Tycon Power Systems (2014). *RemotePro Data Sheet*. URL: http://tyconpower.com/products/files/RemotePro_Spec_Sheet.pdf.

USGS (2013). *Century - Net Ecosystem C Balance ER6.2 CGCM A1B Version 1*. Online. URL: http://www.usgs.gov/climate_landuse/land_carbon/data.asp (visited on).

Wang, D. H. and R. P. Philp (1997). "Geochemical study of potential source rocks and crude oils in the Anadarko basin, Oklahoma". In: *AAPG Bulletin* 81.2, pp. 249–275.

White, M.D. et al. (2014). "Numerical Simulation of Carbon Dioxide Injection in the Western Section of the Farnsworth Unit". In: *12th International Conference on Greenhouse Gas Control Technologies, GHGT-12* 63, pp. 7891–7912. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2014.11.825. URL: http://www.sciencedirect.com/science/article/pii/S187661021402640X.

Williams, T. and C. Kelley (2016). *Gnuplot 5.0: An Interactive Plotting Program*.

Won, Wangyun and Kwang Soon Lee (2012). "Nonlinear Observer with Adaptive Grid Allocation for a Fixed-Bed Adsorption Process". In: *Computers & Chemical Engineering* 46, pp. 69–77. ISSN: 0098-1354. DOI: 10.1016/j.compchemeng.2012.07.001.

Woods, A. (2010). "Turbulent Plumes in Nature". In: *Annual Review of Fluid Mechanics* 42, pp. 391–412.

Yang, Zheng et al. (2014). "A Systematic Approach to Occupancy Modeling in Ambient Sensor-Rich Buildings". In: *SIMULATION* 90.8, pp. 960–977. DOI: 10.1177/0037549713489918.

Yi, Pan, L. Xiao, and Y. Zhang (2010). "Remote Real-Time Monitoring System for Oil and Gas Well Based on Wireless Sensor Networks". In: *Mechanic Automation and Control Engineering (MACE), 2010 International Conference On*, pp. 2427–2429. DOI: 10.1109/MACE.2010.5535870.

Yun, Seok-In et al. (2013). "High-Time Resolution Analysis of Diel Variation in Methane Emission from Flooded Rice Fields". In: *Communications in Soil Science and Plant Analysis* 44.10, pp. 1620–1628. ISSN: 0010-3624. DOI: 10.1080/00103624.2012.756510. URL: http://dx.doi.org/10.1080/00103624.2012.756510.

Zhou, Zhengyi et al. (2014). "A Spatio-Temporal Point Process Model for Ambulance Demand". In: *arXiv preprint arXiv:1401.5547*.

Acronyms

ACES Aircraft for Climate and Environmental Sensing

ArcGIS Geographic information system for working with maps

E-BP Evolving Gaussian Process

CAD Computer Aided Design

CDMA Code Division Multiple Access

CFD Computational Fluid Dynamics

COA Certificate of Authorization

COTS Commercial Off-The-Shelf

DAQ Data Acquisition

DEM Digital Elevation Model

DIY Do It Yourself

E-GP Evolving Gaussian Process

EIEIO Exploitation by Informed Exploration between Isolated Operatives

EVDO Evolution-Data Optimized

FAA Federal Aviation Administration

FPV First-Person Video

GP Gaussian Process

GPC or GP Clustering Gaussian process clustering

GPF Gaussian Process Fusion

GPS Global Positioning System

HMM Hidden Markov Model

HVAC Heating, Ventilation and Air Conditioning

HITL Hardware-In-The-Loop

IUPAC International Union of Applied Chemistry

LIDAR Light Imaging Detection and Ranging

LiPo Lithium Polymer

LOS Line Of Site

MDP Markov Decision Process

MOU Memorandum Of Understanding

Nas National air space

NDIR Nondispersive Infrared Sensor

NETL National Energy Technology Laboratory

NIOSH National InSstitute for Occupational Safety and Health

NOTAM Notice to Airmen – Written notification issued to pilots before a flight

OSHA Ocupational Safety and Health Administration

OSU Oklahoma State University

ppm Parts Per Million

PCB Polychlorinated biphenyl

PIC Pilot In Command

PID Proportional Integral Derivative

PED Poisson Exposure Distribution

PEP Poisson Exposure Process

RISC Reduced Instruction Set Computer

RMSE Root-Mean-Square-Error

SD Secure Digital

SUAS Small Unmanned Aerial Systems

SWP Southwest Regional Carbon Sequestration Partnership

UAV Unmanned Aerial Vehicle

UAFS Unmanned Aircraft Flight Station

UAS Unmanned Aerial System

UGS Unmanned Ground Sensor

VMC Visual Meteorological Conditions

VOI Value of Information

Appendicies

Appendix 1

Presentations and Publications

Presentations

Jacob, Jamey, Rakshit Allamraju, Allan Axelrod, Calvin Brown, Girish Chowdhary, and Taylor Mitchell, "Airborne Detection and Tracking of Geologic Leakage Sites," 67th Annual Meeting of the APS Division of Fluid Dynamics, Volume 59, Number 20, November 23–25, 2014; San Francisco, California.

Taylor Mitchell and Jamey Jacob "Airborne Detection of CO2," 3 Minute Thesis Competition, College of Engineering Winner, April 16th, 2015, Stillwater, OK.

Jamey Jacob , Taylor Mitchell , Wes Honeycutt , Nicholas Materer , Tyler Ley , Peter Clark, "Monitoring of Carbon Dioxide and Methane Plumes from Combined Ground-Airborne Sensors," 69th Annual Meeting of the APS Division of Fluid Dynamics, Volume 61, Number 20, November 2022, 2016; Portland, Oregon.

Meeting Papers and Presentations

Taylor, M., Brown, C. and Jacob, J. "CO2 Detection Using UAS," (AIAA 2015-1459) AIAA 53rd Aerospace Sciences Meeting, Orlando, FL, 2015.

Mitchell, T., Kidd, J., and Jacob, J. D "Plume Tracking and Dynamics Using UAS," AIAA 2016-1459, AIAA Aerospace Sciences Meeting, California, CA, 2016.

Materer, N. "Sensor arrays for environmental detection of carbon dioxide and methane in remote and/or hazardous locations," invited presentation at the 6th ICC 2016 "New Trends in Chemistry" Conference held in Riyadh,Saudi Arabia.

Papers

Meng, J., Pashin, J.C., Clark, P., 2017, "Structural Architecture of the Farnsworth Oil Unit: Implications for Geologic Storage of Carbon Dioxide," *Environmental Geosciences*, v. 24, no. 2, p. 73-94.

Taylor, M., and Jacob, J. "CO2 Detection Using UAS," Submitted to *Atmospheres.*

Appendix 2

Long-Term Carbon Dioxide Concentration

Data presented in Fig. A.2.1 chronicles approximately 320 days of continuous CO_2 monitoring. The baseline for CO_2 is 400 ppm and the peaks represent ppm above the baseline. Each time step represents 15 minutes. The data begins on 4/16/2016 and terminates on 3/2/2017. By the middle of September (approximately 15,000 time units) the level and frequency of CO_2 spikes decreased. The exact reason for this change is not known, but was observed at all of the measurement stations. Potential sources for the CO_2 include farming activity, fires, petroleum powered vehicles and intermittent releases from injection/compression operations. Most of which should decrease as the weather cools. The one exception is probably CO_2 generated from vehicles.

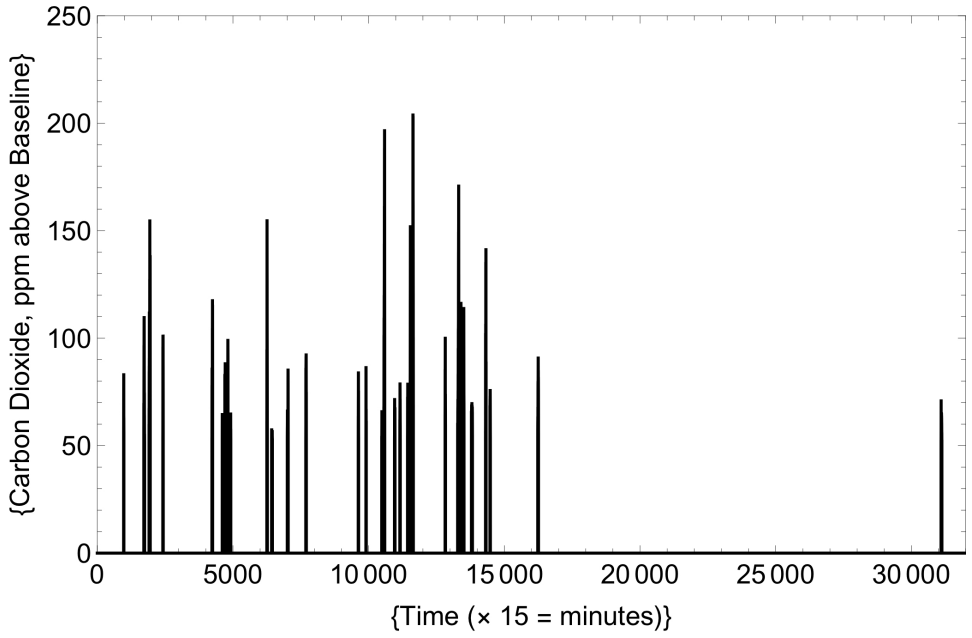


Figure A.2.1: Approximately 320 days of CO_2 data.

Appendix 3

Data Filtering

As part of the data analysis process, the data was filtered to remove random noise, short duration spikes, and peaks that were less than 2σ above the baseline. Each time unit on the x-axis is fifteen minutes. Peaks were only retained if they persisted eight time units or more. Figure A.3.1 the unfiltered data is shown in blue and the filtered data in red. Looking at the blue peak around 11,000 time units, it is obviously the tallest peak. It is, however, only one time unit in duration while the red peak (second tallest peak) next to it lasted for almost a day. Unfortunately, when approximately 32,000 points are plotted on the same graph, the width of the peak does not reflect the elapsed time to any degree of accuracy.

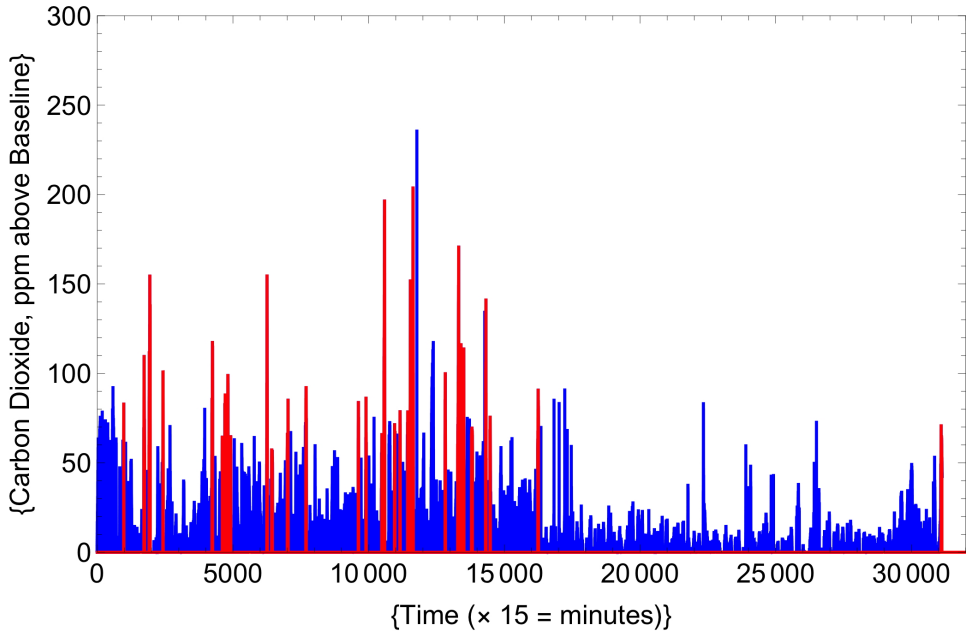
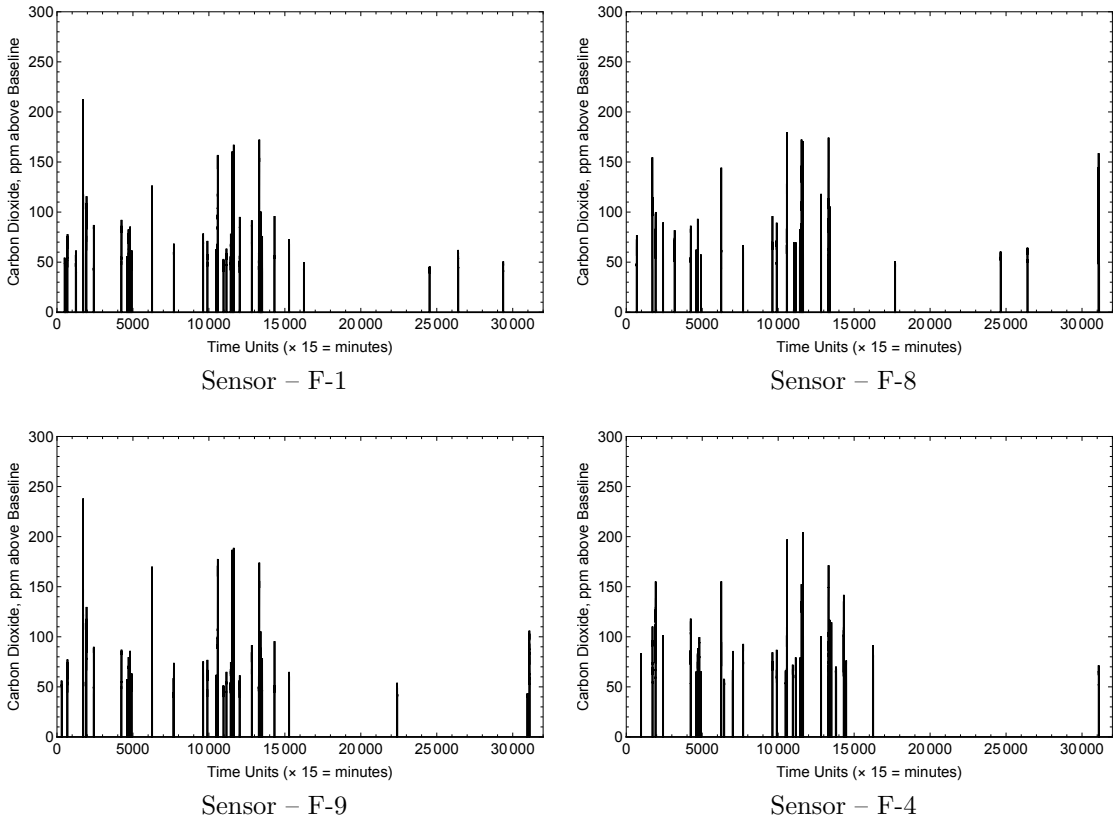


Figure A.3.1: Filtering example

Appendix 4

Examples of CO_2 and CH_4 Data From Different Sensor Locations



Sensor – F-1 Sensor – F-8

Sensor – F-9 Sensor – F-4

Figure A.4.1: Longterm CO_2 Sensor Data From Different Sensors Located in the Same General Vicinity

The plots shown in Fig. A.4.1 are representative of the data produced by the sensors. All of the sensors come from the blue set of locations shown in Fig. 2.24. These sensors were selected because they were the first to be deployed and therefore have the largest data set.

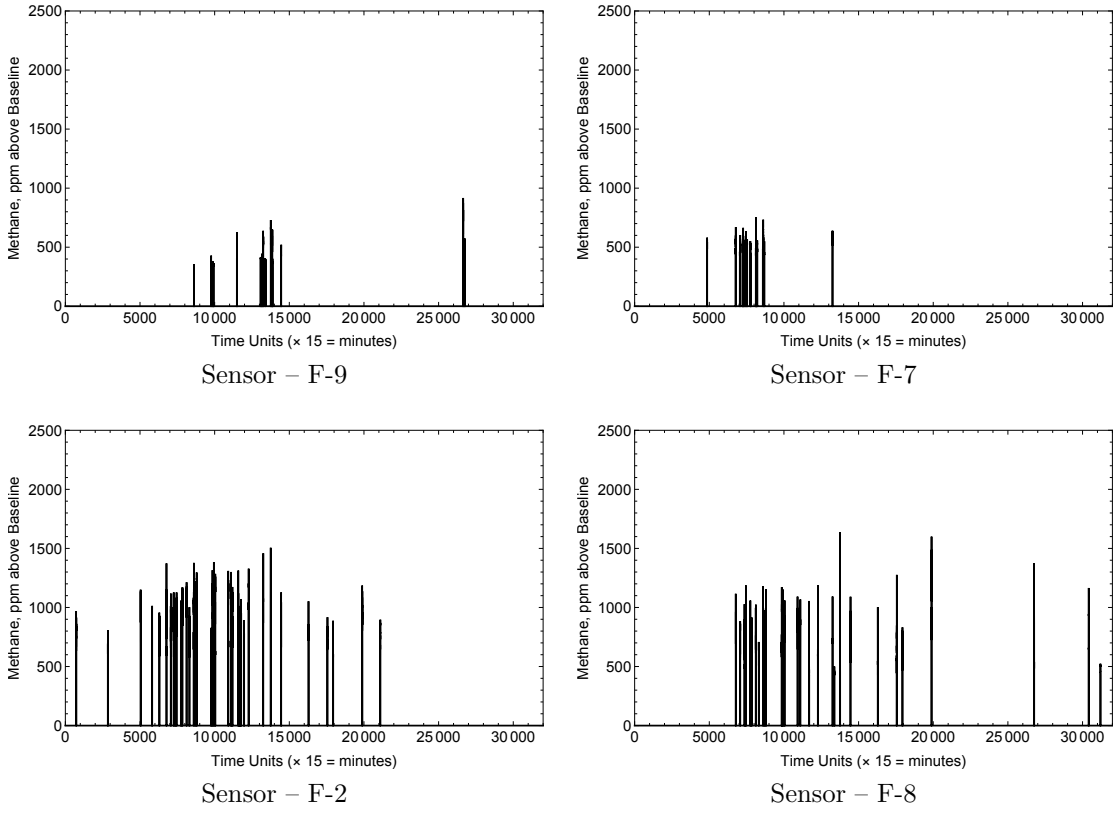
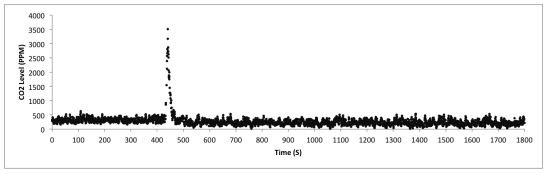


Figure A.4.2: Longterm CH_4 Sensor Data From Different Sensors Located in the Same General Vicinity

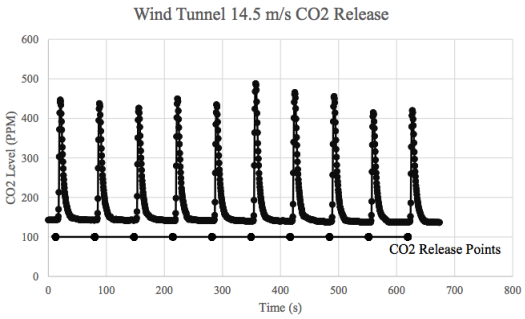
Data in all of the plots shown in Fig. A.4.2 were derived from individual sensor output that was processed in the same manner as the CO_2 data.

Appendix 5

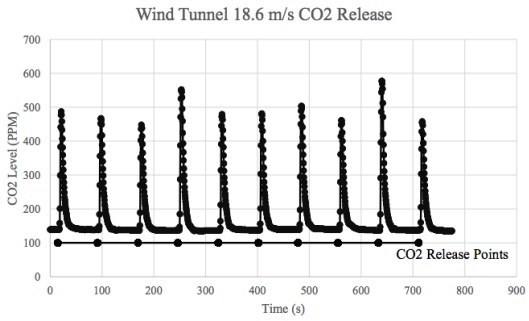
Flight Data



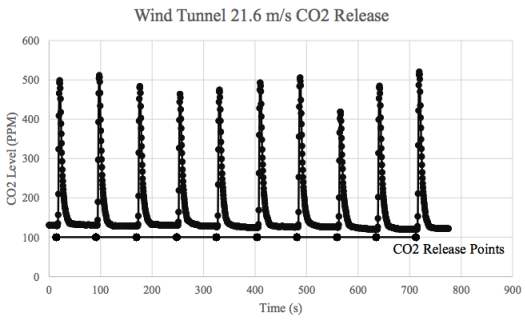
Burst Response



Low Speed – 14.5 m/s



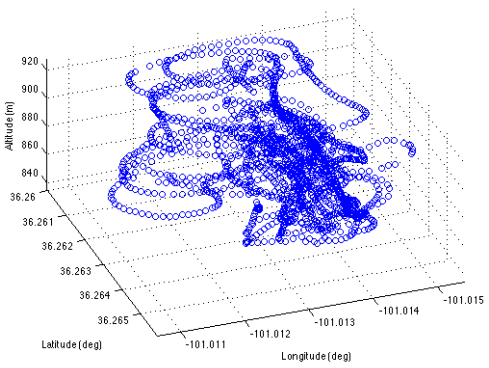
Medium Speed – 18.6 m/s



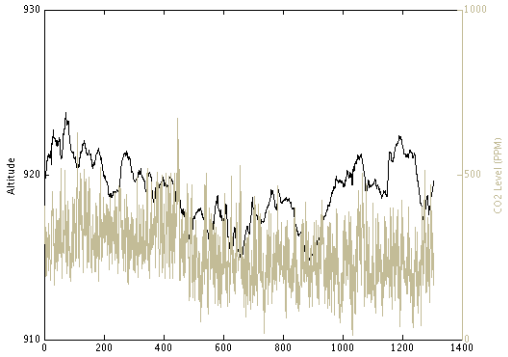
High Speed – 21.6 m/s

Figure A.5.1: Airborne Sensor Wind-Tunnel Calibration

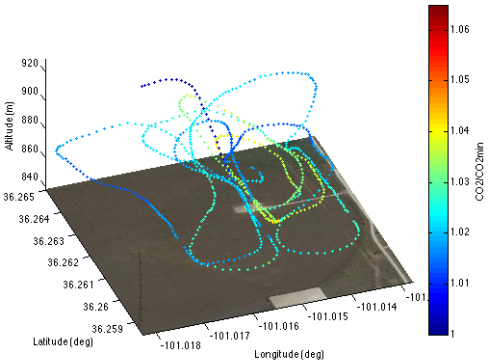
The plots shown in Fig. A.5.1 indicate that the CO_2 sensor used in the airborne sensor package has an excellent response time and good reproducibility at the same and different air speeds.



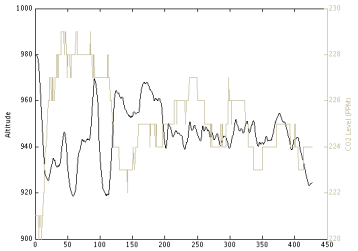
Flight Path – 5/25



Normalized CO_2 Concentration – 5/25



Flight Path – 8/17



Normalized CO_2 Concentration

Figure A.5.2: Flight Path and Sensor Data From Farnsworth

Fig. A.5.2 shows the flight paths and CO_2 concentrations as a function of position for two different runs. The CO_2 sensor used to acquire the data on 5/25 was somewhat noisy. Data acquired on 8/17 showed a sharply reduced noise component.

None of the flight data produced much new information above that acquired using the ground-base system. As pointed out earlier, plots of the ground-base sensor nodes give the impression that the CO_2 spikes are close together, but because there are almost 32,000 data points on most of the plots temporal resolution is misleading. If there should be a geologic failure that resulted in a leak, the airborne system would be useful for locating the problem.

Appendix 6

Cost Analysis

Hardware costs are summarized in Table 2.6. Units produced commercially would be three to seven times more expensive. The multiplier will depend on the size of the market for the units as the cost of components and assembly drops rapidly as the number of units per year increases.

Installation costs will vary with travel time. A crew of two can install four or five sensors on power poles per hour depending upon separation between the poles. Installed on stakes as shown in Fig. 2.17 the number of sensors installed per hour would increase to ten or so. The installation costs will depend on the labor cost. For this project the cost was rather inexpensive since we used graduate students.

Operating costs were small since the data was reported through a cellular network. The cellular bill was approximately fifty dollars per month. With the proper software, regular reports or leak detection warnings could be generated with little or no effort. Currently, the data is stored in an SQL type database. The computer needed to run the whole project would cost less than $1,000.

The figure shown below is a section of Fig. 2.27 showing the near simultaneous response of the sensors to a CO_2 event at the Farnsworth site. Since most of the sensors were mounted on power poles, the distance between the sensors were a function of the pole placement. Separation distance varied from a few tens of feet to over one mile. Data were taken during fifteen minute intervals, so the temporal resolution is not fine enough to resolve movement of the event across the section, but the measurement intervals were short relative to the duration of the event marked by the read dashed line in the figure.

Given the near simultaneous detection of a CO_2 event over a large area, dense placement of sensors over a given area is probably not necessary. A leak resulting from a geologic failure in the rock seal would in all likelihood result in an event marked by a persistent baseline shift not a short term spike in CO_2 concentration above baseline. This would be true of a major pipeline failure. The figure below shows placement of four sensor nodes and one communication node in a four square mile (2.59×10^6 m^2) area. Doubling the area to eight square miles is less than twice the cost because two of the sensor nodes can be shared. As in any long-term monitoring situation, the manpower cost to analyze the data will in all likelihood overwhelm the hardware costs. Much of this effort could be automated if the goal was limited to leak detection.

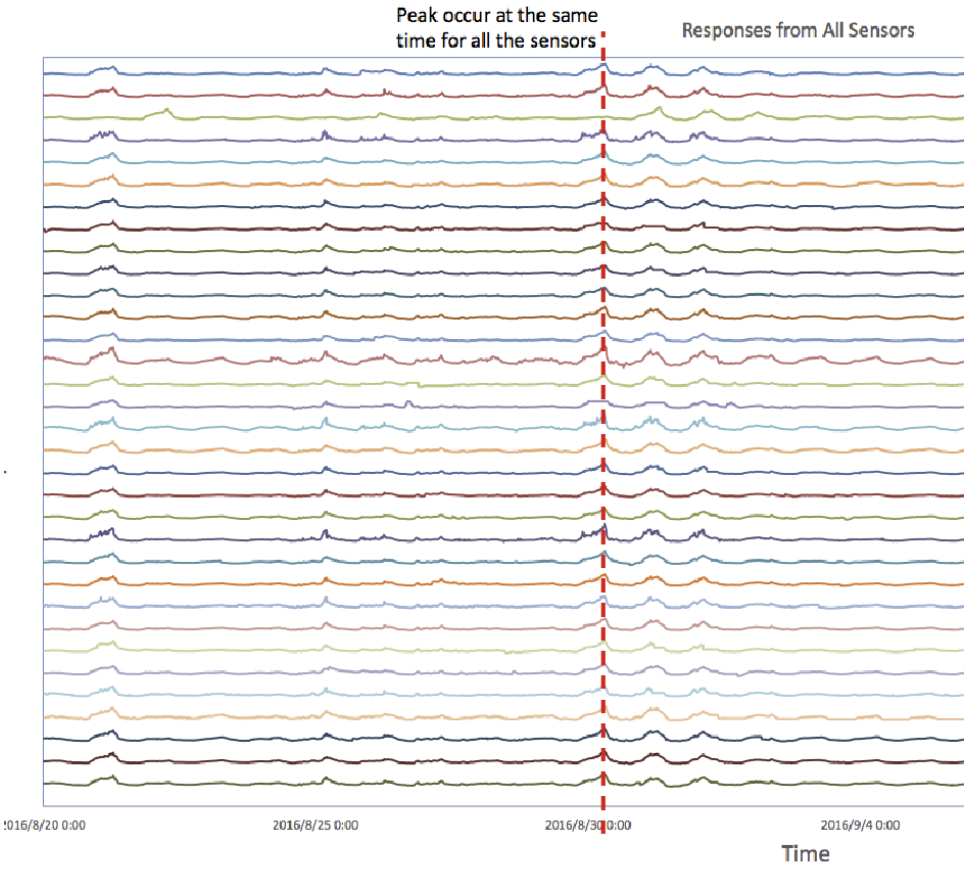


Figure A.6.1: Segment of Figure 2.27

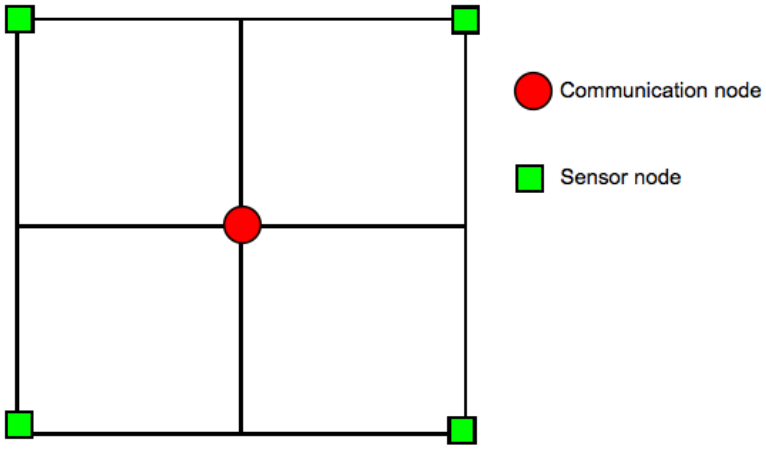


Figure A.6.2: Sensor placement in a four square mile area.