Procedia Computer Science
Volume 51, 2015, Pages 1633–1642
ICCS 2015 International Conference On Computational Science

T…an Integrated Cyberinfrastructure for S…riven Monitoring, Dynamic Prediction… Resilience of Wildfires

Ilkay Altintas[1*], Jessica Block[2], Raymond de Callafon[3], Daniel Crawl[1], Charles Cowart[1], Amarnath Gupta[1], Mai Nguyen[1], Hans-Werner Braun[1], Jurgen Schulze[2], Michael Gollner[4], Arnaud Trouve[4] and Larry Smarr[2]

[1]*San Diego Supercomputer Center, University of California San Diego, U.S.A.*
[2]*Qualcomm Institute, University of California San Diego, U.S.A.*
[3]*Dept. of Mechanical and Aerospace Engineering, University of California San Diego, U.S.A.*
[4]*Fire Protection Engineering Dept., University of Maryland, U.S.A.*

Abstract
Wildfires are critical for ecosystems in many geographical regions. However, our current urbanized existence in these environments is inducing the ecological balance to evolve into a different dynamic leading to the biggest fires in history. Wildfire wind speeds and directions change in an instant, and first responders can only be effective if they take action as quickly as the conditions change. What is lacking in disaster management today is a system integration of real-time sensor networks, satellite imagery, near-real time data management tools, wildfire simulation tools, and connectivity to emergency command centers before, during and after a wildfire. As a first time example of such an integrated system, the WIFIRE project is building an end-to-end cyberinfrastructure for real-time and data-driven simulation, prediction and visualization of wildfire behavior. This paper summarizes the approach and early results of the WIFIRE project to integrate networked observations, e.g., heterogeneous satellite data and real-time remote sensor data with computational techniques in signal processing, visualization, modeling and data assimilation to provide a scalable, technological, and educational solution to monitor weather patterns to predict a wildfire's Rate of Spread.

Keywords: Cyberinfrastructure, Data Assimilation, Workflows, Wildfire Modeling, Scientific Data Integration

1 Introduction

Fire is critical for healthy ecosystems in much of the world. However, our current urbanized existence in these environments, in conjunction with exotic vegetation growth, imported water resources, and global climate changes, is inducing the ecological balance to evolve into a different dynamic; a different climatological system of rainfall, wind, seasons, and thus fire seasons. In the 21st century California, Arizona and Texas have seen their biggest fires in recorded history. In October 2003, a series of massive Santa Ana-driven wildfires erupted in Southern California. In San Diego (SD) County the Cedar fire [1] burned 280,278 acres, 2,820 buildings and killed 15 people, including one firefighter. In 2007, Santa Ana winds created an even larger set of wildfires, leading to the evacuation of over half a million people in SD County, the largest fire evacuation in U.S. history, and causing damages over $1 billion [2]. The Wallow Fire in 2011 was the largest in Arizona's history [3].

* Correspondence should be sent to: altintas@sdsc.edu. This work was supported mainly by NSF-1331615 under CI, Information Technology Research and SEES Hazards programs, and in part by NSF-112661, NSF-1062565 and NSF- 0941692.

© The Authors. Published by Elsevier B.V.

doi:10.1016/j.procs.2015.05.296

Wind speeds and directions affecting the spread of a fire can change instantly, and first responders can only be effective if they take action as quickly as the conditions change. To deliver information needed, we must capture the details of these conditions to understand environmental processes. SD County is uniquely positioned to monitor and analyze these dynamics through our research sensor networks, namely, the High Performance Wireless Research and Education Network (HPWREN) [4]. It was not until the last decade that we have the capacity to feed field measurements into simulations and visualizations at high resolutions, and it is rarely available in real-time to those that need it, in this case environmental modeling efforts.

Significant measurement instrumentation at the HPWREN and partner sites facilitates the collection of large dimensional heterogeneous data of disparate environmental sensors that include meteorology, vision, audio and hydrology. Integration of real-time sensor telemetry data can provide better situational awareness integral to decision making processes for emergency response situations. However, it is critical to provide only relevant data for environmental awareness to the recipient to avoid "data-overload" and "sensor failure". As the quantity of sensors in wide-area multiple-domain environments increase, it is imperative to provide a systematic and easy-to-maintain programming environment for data analysis in which large dimensional heterogeneous sensor data can be reduced in real-time to a lower dimensional representation. In particular, data assimilation and parameter estimation techniques can be used to address the need to reduce large dimensional data sets to a parametric lower dimension suitable for analysis, interpretation and alert purposes. Such systematic data reduction to a lower dimensional representation allows a more efficient (mobile) communication of events for decision-making and crisis management, but the following questions need to be addressed for effective environmental data analysis:

- How can large dimensional heterogeneous sensor data of the natural environment be analyzed systematically to a (lower dimensional) format useful for information processing, real-time monitoring and visualization?
- How can such data be combined with existing scientific models to allow for prediction of propagating wildfires and potential future events to prepare within regions of highest risk?
- What quality and density of real-time sensors is necessary to improve both the predictive and preventive capabilities of current fire models?
- How can such information processing be easily configured, programmed and computed by users with various skill levels to formulate actual real-time data-driven environmental alerts?

In particular, *what is lacking in disaster management today is a system integration of real-time sensor networks, satellite imagery, near-real time data management tools, wildfire simulation tools, and connectivity to emergency command centers before, during and after a firestorm.*

Contributions. This paper describes the initial design of our end-to-end integrated cyberinfrastructure, called WIFIRE, to *catalyze new thinking paradigms and practices* for wildfire research and response. The presented near real-time WIFIRE software infrastructure integrates networked observations, e.g., heterogeneous satellite data and real-time remote sensor data with computational techniques in signal processing, visualization, modeling and data assimilation to provide a scalable, technological, and educational solution to monitor weather patterns to predict a wildfire's Rate of Spread. We also present our early efforts and findings on data management, programmability, scalability and visualization of wildfire data and workflows.

2 WIFIRE System Architecture

WIFIRE cyberinfrastructure is architected for solutions with pathways that enable joint innovation for wildfire management and collaboration between its diverse users. WIFIRE software products integrate wide-area multiple-domain sensor telemetry data. New data assimilation and parameter estimation techniques reduce these large telemetry data sets to parametric lower dimension models to

predict the Rate of Spread (ROS) of wildfires in a constantly changing environment. Scientific workflows are used for integration of developed techniques and as a distributed programming and execution model that supports interfacing to different components of the cyberinfrastructure and heterogeneous computing platforms.

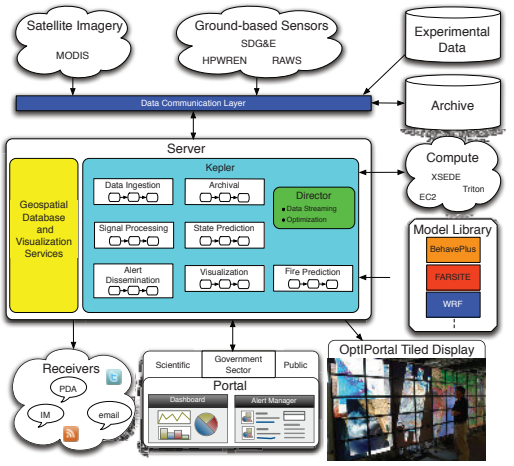


Figure 1. Integrated real-time data processing and programming environment in WIFIRE.

Improving the data processing functionality for the monitoring, modeling and prediction of wildfire spread is accomplished in WIFIRE by integrating three main components as indicated in Figure 1: **(i)** a central organization of Scientific and Engineering modules within the open-source Kepler scientific workflow system [5] that coordinate the execution of real-time data processing and fire propagation tools on distributed computing environments, **(ii)** a data communication layer with links to archives, experimental data, modeling products and heterogeneous sensor data from a diverse set of data sources, and **(iii)** portals for dissemination of data to different end users that include scientists, first responders and public notification of user-defined real-time alerts via various receivers and Web 2.0-based public systems. The scientific workflow component includes all real-time data processing tools for data assimilation of different wildfire spread models, recursive parameter estimation, and real-time prediction of wildfire spread. Visualization interfaces include the Optiportal Tiled Display, i.e., HIPerspace Wall, and STAR CAVE currently placed at Calit2 Qualcomm Institute.

Due to the central organization around scientific and engineering modules within the open-source Kepler scientific workflow system, the WIFIRE cyberinfrastructure is built inherently scalable to handle both large sets of heterogeneous sensor data [6] and execute computations in parallel on distributed computing environments [7], e.g., XSEDE. The Data Communication Layer can be extended to include different kind of data sets without altering these Kepler modules.

3 WIFIRE Subsystems

3.1 Data Communication

WIFIRE's *Data Communication Subsystem* provides a layer above all data sources needed for wildfire modeling. It handles the ingestion and integration of a diverse set of ground-based and airborne sensor data and satellite imagery along with other experimental, modeling and geopolitical datasets. The use of low-latency sensor data delivery mechanisms for field-deployed sensors supports scalability and rapid availability of raw or pre-processed data. This scalable approach allows for the

simultaneous availability of such data to many processing modules, without creating an excessive load on the networking or computing substrates. The three sensed dataset sources that are integrated and streamed via the data communication subsystem can be summarized as follows:

- **Meteorology Stations:** Meteorology, a.k.a. weather stations, are collections of sensors that record and transmit multiple environmental metrics, including wind speed and direction, air temperature, barometric pressure, and relative humidity. WIFIRE receives near real-time data from HPWREN's nineteen stations and receives updates every ten minutes from over 150 stations operated by San Diego Gas & Electric (SDG&E). In addition, WIFIRE collects smaller datasets from providers including the National Park Service, and San Diego State University's Field Stations Program.
- **Still-Image Cameras:** 110 individual cameras collect color and grayscale images from 35 separate locations throughout SD County. Still-images are valuable for remotely observing fire, smoke, and other phenomena; moreover, the locations of these phenomena can be triangulated using multiple cameras and detection algorithms. Geospatial view sheds on the collected camera images are being built for near-real time smoke detection and querying of the fire location.
- **Satellite and Aerial Data Products:** Data taken from the Terra and Aqua units of the MODIS satellites are used to generate one-kilometer resolution fire and smoke detection maps of the SD County region four times a day. The WIFIRE project has modeled this data, and is extending its software to serve it. In the future, WIFIRE will incorporate data from additional satellites such as AVHRR, VIIRS, GOES, Landsat, and aerial data from multiple sources.

The data communication layer includes REST-based services for ingesting sensor data and derived time-dependent parameters into geospatial databases and archives so that the data can be extracted into any Open Geospatial Consortium compatible data format. The ingestion process connects to the HPWREN sensor multicast streams and tests for conditional ETL rules and routes to appropriate databases for sub-processing. Doing all the processing for incoming satellite data in RAM keeps the processing time very low, and the data is written to permanent archives after processing.

3.2 Wildfire Behavior Modeling and Data Assimilation

For modeling wildfire growth and computing the Rate of Spread (ROS) of a wildfire, one may distinguish between studies that focus on computational fluid dynamics (CFD) of flame-wind interactions and the semi-empirical Rothermel model [8] or it's equivalent. CFD studies of flame-wind interactions typically feature a problem size of several hundred meters and a spatial resolution on the order of one meter. The CFD studies provide a description of the fire-driven modifications of atmospheric conditions. Although powerful in describing wildfire dynamics and ROS, their computational cost is high, limiting flame-scale CFD to off-line analysis of wildfire events.

Examples of operational wildfire spread models used in the US include BehavePlus [9] and FARSITE [10]. Operational semi-empirical models can also be found in the McArthur Fire Danger Meter in Australia [11] and the Fire Behavior Prediction System (FBPS) in Canada [12]. Spatial information on topography, fuel content and moisture along with regional weather and wind input is used to drive the semi-empirical models. This allows the fire spread model to adopt a regional scale perspective and simulate a wildfire as a propagating front for long time periods under heterogeneous conditions of terrain, fuel parameters, and weather conditions. It is clear that spatial information on terrain, fuel parameters and weather conditions provide the *a priori* information needed to drive the operational wildfire spread models. However, the dense weather sensors and remote Internet connectivity available in WIFIRE provides a big asset to update the *a posteriori* information during the computation of wildfire behavior by virtue of Data Assimilation (DA). Real-time measurements to correct simulation errors by DA techniques have been explored in wildfire and building fire applications [13-14] along with oceanographic and geophysical fluid flow applications [15-16].

The DA system of WIFIRE is integrated with an operational semi-empirical fire model and allows the implementation of parameter and state estimation techniques that can serve two purposes. First, a DA technique can be used to implement a recursive parameter estimation technique that adjusts the *a priori* fuel parameters and measured or predicted *a priori* wind conditions. Such parameter estimation techniques only make adjustments to the (input) parameters of the operational wildfire spread model. Secondly, a DA technique can be used to implement a recursive state estimation technique that adjusts the simulated fire front location with an a posteriori update/measurement of the actual fire front location. The recursive state estimation updates the initial conditions (e.g. the states) of the operational wildfire spread model, to provide a simulation of the fire front that is closer to the measured fire front location. Recent examples of wild fire DA techniques for (input) parameter estimation can be found in [14] where a spatially-uniform correction of biomass fuel and wind parameters is used. State estimation to sequentially update the two-dimensional coordinates of markers along a discretized fire front can be found in [17]. Input and state estimation can also be combined in a single DA technique as developed in [18] with applications to flow field estimation.

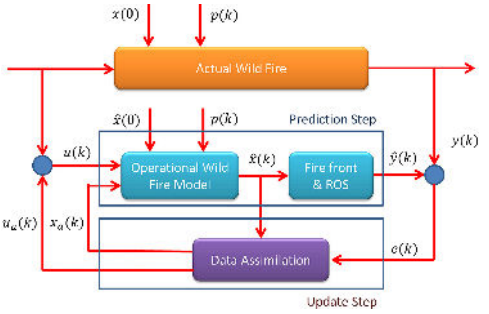


Figure 2. Illustration of prediction and update steps via separation in models for operational wild fire modeling with fire front/ ROS simulation and Data Assimilation algorithm.

The DA system used in WIFIRE builds upon existing and well-established recursive algorithms formulated by the Extended Kalman Filter (EKF) and the Ensemble Kalman Filter (EnKF). In addition, extensions are made via a Bayesian framework to formulate recursive algorithms to jointly estimate a perturbation $x_a(k)$ to the simulated state $\hat{x}(k)$ and a perturbation $u_a(k)$ to the input $u(k)$ for the operational semi-empirical wild fire model. To allow the implementation of a DA via (Kepler) workflow components that run sequentially in time, the algorithms for the DA system in WIFIRE will be separated in two parts, as illustrated in Figure 2. The first part is a *prediction* step, in which a previous estimate of the state is evolved forward in time to the time of a new observation of the fire front using the perturbation $x_a(k)$ and modified input $u(k)$. The second step is an *update* step, where the evolved estimate $\hat{x}(k)$ of the state is updated using information from the fire front observations.

The separation between a prediction and update step is standard in the EKF and EnKF, and allows an existing operational fire model such as FARSITE to be implemented as a separate workflow component for the *prediction* step. Explicit expressions for the evolution of the error covariance (as done in the EKF) or statistical approximation of the error covariance (as done in the EnKF) can be implemented as separate workflow components. In addition, the *update* step involving the error covariance and measurement covariance matrix based on an EKF, EnKF or Bayesian joint input and state estimation algorithm can also be implemented as separate workflow components. In the update step, workflow components in Kepler must pull sensor measurements to allow for real-time updates for the operational fire model. More details on the workflow management and components in WIFIRE will be given in the following section.

3.3 Workflow Management

A workflow is comprised of steps or tasks linked together by data dependencies. Workflows are an integral part of the overall WIFIRE architecture to ensure system integration and programmability, computational scalability and reproducibility of the developed data access and modeling tools.

We have chosen the Kepler Scientific Workflow System [5] and APIs to build our workflow design and execution specific extensions. Kepler provides a graphical user interface for designing workflows composed of a linked set of extensible and configurable components called *Actors* that may execute under a rich set of different Models of Computations (MoCs). Actors are the implementation of specific functions that need to be performed and communication between actors takes place via tokens that contain both data and messages. The MoCs are implemented by *Directors*, which specify how the communication between the actors is achieved, when actors execute, and when the overall workflow execution stops. The designed workflows can then be run through the user interface or in batch mode from other applications. In addition, Kepler provides a provenance framework [20] that keeps a record of chain of custody for data and process products within a workflow design and run.

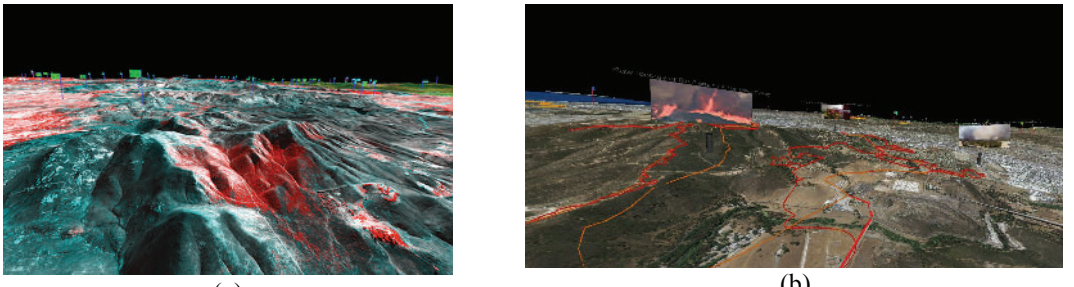
3.4 Data Mining

Data mining is an interdisciplinary field concerned with analyzing large amounts of data to discover patterns. Data mining techniques are being applied to WIFIRE data to gain insight into environmental conditions affecting fire behavior, e.g., analysis of data from weather stations can determine patterns of weather data associated with Santa Ana conditions. This information can then be used to alert firefighters of specific regions experiencing conditions susceptible to wildfires. Alerts of changing weather conditions can be especially useful for areas surrounding an existing wildfire. Results from data mining tasks can also feed into the fire models to provide specific data about current environmental conditions for more accurate modeling of fire behavior.

New Kepler components for data mining are being implemented. Existing machine learning tools and libraries such as R and MLlib [20] will be leveraged, as well as Kepler's Distributed Data Parallel (DDP) capability to provide a framework for scalable processing of WIFIRE data.

3.5 Visualization and Communication of Results

The WIFIRE research in visualization is building products in virtual reality, mobile data acquisition, and web mapping. Each of these addresses important needs for fire research and response.



(a) (b)

Figure 3. (a) osgEarth screenshot of a preserved vegetation patch of young chaparral (red) near Ramona, CA. The burned landscape is in dark gray-green. Data resolution: 0.5-m for imagery and 2-m for topography. Flagpoles represent weather sensor showing wind direction, wind velocity (flag length) and local air temperature (color)**. (b)** Geolocated Twitter images and fire perimeters are drawn to provide near real-time awareness of fire progression.

Virtual Scientific Visualization. Immersive visualizations are important for understanding complex processes in fire behavior and for testing best practices for visualizing fire event data [21].

Our prototype uses the osgEarth, an open source geospatial graphics platform that is part of OpenSceneGraph (OSG). The OSG is capable of being run on multi-screen tile display walls in mono and in stereo. Its architecture is also optimized to ingest real-time data and tile hierarchies for fast serving and consumption over the web, and can natively read standard GIS and volumetric model file formats [22]. Using tile display walls, the resolution of the data viewed is scalable from megapixels to tens of megapixels. The accessibility of this high-resolution display with the data has proven useful when working with fire agencies in Southern California to understand the effects of past fires on the landscape (see Figure 3). The high-resolution imagery and topography highlights the intricate influence of topography on fire behavior in historic events. Taking the highest resolution imagery and topography freely available, we used the osgEarth platform to display SD County burn scars from the 2003 Cedar fire. Figure 3 shows screenshots of the data in the Calit2 NexCAVE immersive environment, focused in on the burned wildland-urban interface where the fire hazards remain high.

The prototype immersive virtual environment is being expanded to view multiple new datasets. To test the ingestion of mobile photos, we have tested with twitter images taken during the most recent firestorms in San Diego. Figure 3(b) shows three twitter images posted as billboards on the San Diego landscape. Red lines on the ground indicate fire perimeters.

Mobile Data Acquisition. We are developing a mobile app that allows users to take and upload pictures of active fires. The app uses the phone orientation and geographic location to accurately place a collected image so as to recreate the view perspective of the photographer if the photo is placed in a virtual 3D environment. We can then use those photos to plot in 3D and see the active fire in the virtual environment. Further development will use computer vision techniques to identify the horizon if it appears in the picture, which can then be used to geolocate the flames or smoke. If enough images are taken, we can create dynamic estimates of a fire perimeter as a fire is traveling.

Web Maps. Web maps have been developed to make the environmental data accessible and interactive. Users can now monitor environmental conditions and view workflow results. Although much of the data is publicly available, they have not previously been integrated in the same interface, e.g., HPWREN cameras can be located by the map interface and the user can correlate the location with what the cameras can see. Model outputs (see Section 3.2) can also be viewed in this interface.

4 Early Results

4.1 WIFIRE Data Model

Management of *information heterogeneity* is the focal theme in the design of the data architecture of the WIFIRE system. Weather stations produce a time-series of vector of measurements together with the metadata of the sensors themselves; satellite data providers like NASA produce time-sequences of arrays of multi-band rasters, where some of the processed data may have their own data dictionary; ground topography, an important factor in fire simulation is often represented as digital elevation maps; fire simulation models like FARSITE produce temporally evolving fire perimeters (temporal polygons); wind simulators like WindNinja produce time and space-varying vector fields; human observers and monitoring cameras produce geolocated images of fire. The data modeling task in WIFIRE is to capture all metadata, observed data and computed data assimilated by the system into a single common framework. To this end, we have developed a *semantic semistructured data model* that covers both data and metadata. The structure of the model follows from the observation that regardless of the source and nature of the data, all data elements are essentially spatiotemporal and can be semantically associated with an ontology of observables. We list the primary elements of our model through the following exposition.

- The data model assumes the existence of an ontology which itself may be constructed from terms and relationships from several contributing ontologies. An ***observable*** is defined as an

ontological element that corresponds to an entity observed by a sensor. It is possible to have a more complex observable like wind, whose properties are sensed and stored as data. In this case the metadata system models the entity as well as its properties.

- For all measured entities or entity properties, the data model captures their spatial and temporal resolutions.
- For some data sources, the data model is associated with a data dictionary that specifies the domain of a variable captured by the model, e.g., NASA's Aqua MODIS satellite creates a thermal anomaly model that identifies regions that have fire. Our data model recognizes geolocated arrays as a basic data type, where a cell of the array may contain a vector of values.
- Since our essential data model is spatiotemporal, we capture data sources that do not output the actual data but rather produces an aggregate of the data over a window of observations.

A novel aspect of our metadata model stems from the requirement for enabling simulation engines with data and capturing the output of these engines. In such cases, we treat a simulation engine as a function that accepts a complex semistructured object and returns another semistructured object.

REST Service Interface to the Wildfire Data Model. WIFIRE currently exposes its metadata catalog through an industry-standard REST-based interface. We provide data at minute-level resolution from close to 170 stations around San Diego County. The interface returns data and metadata in XML, incorporating ontologies from NASA SWEET and OGC. In addition to XML, WIFIRE supplies selected data in additional formats such as GeoJSON, Comma-Separated-Values (CSV), and WindNinja's native input format. Our objective is to make it easier and faster to search for and push data directly to software from the fire-modeling and other communities.

4.2 Workflow Use Cases

To facilitate access to geospatial data and usage of the accessed data in scientific analysis tools, WIFIRE has implemented new Kepler GIS actors to read and write GIS files such as Shapefile, KML, and GeoJSON. These actors can read, write, transform, and perform various operations on both vector and raster data, and are built using GeoTools (geotools.org), an open-source Java GIS Toolkit. Additionally, two example data-driven use case applications have been created as Kepler workflows.

4.2.1. Use Case 1: Detection of Areas Affected by Santa Ana Conditions

A use case application was created to determine areas within SD County experiencing severe fire-weather conditions called *Santa Ana winds*, which can lead to very dangerous fire conditions. A Santa Ana wind can be defined as a combination of values for the wind direction, wind speed, and relative humidity. The HPWREN system currently monitors these values measured by HPWREN and SDG&E weather stations, and sends an email alert when a station experiences Santa Ana conditions. However, this alert only denotes that a specific point source (the weather station) is experiencing Santa Ana conditions, and not the size of the area. It would be useful to know how large the area is surrounding the station experiencing Santa Ana winds, e.g., how many homes are affected.

A Kepler workflow was created to determine the area around each station experiencing Santa Ana winds by running WindNinja [23] to calculate the wind field and perform post-processing to find Santa Ana winds. WindNinja is open-source software that computes spatially-varying wind fields. WindNinja reads the topography, vegetation, and weather station measurements to produce a vector field of wind speeds and directions over the input domain. The workflow queries the WIFIRE REST interface described in Section 4.1 to download the weather station measurements. Since WindNinja runs on domain sizes of up to 50 km by 50 km, the workflow partitions SD County into smaller tiles to calculate the winds over the entire county. For each tile, the workflow executes a separate WindNinja and all the WindNinja instances can run in parallel. The workflow uses Kepler's DDP framework [7] for executing WindNinja in parallel using either Hadoop or Spark. Note that this workflow currently is being updated to take into account the errors in the boundaries of the tiles by including a buffering technique for the overlaps.

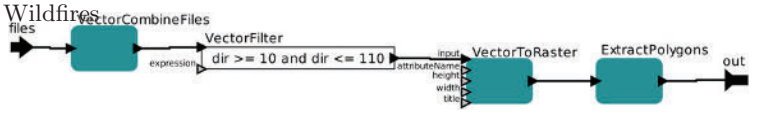
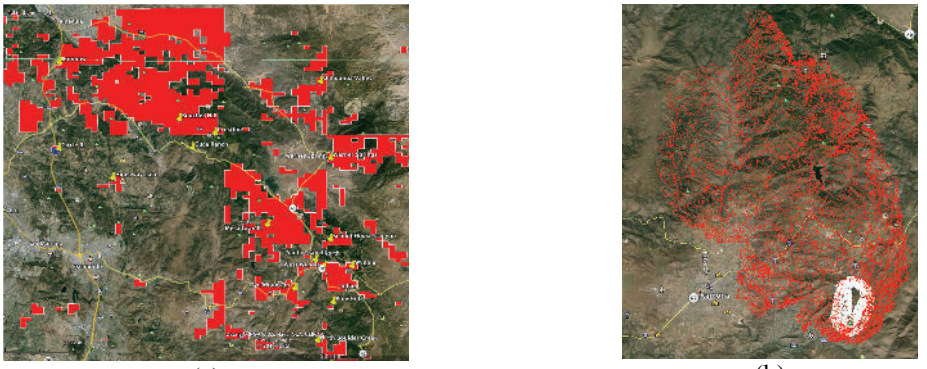


Figure 4. GIS post-processing workflow that filters and extracts the wind fields matching Santa Ana winds.

The output from WindNinja is a vector field of wind speed and direction over SD County. The workflow then performs post-processing on this field to find the areas with parameters defining Santa Ana wind conditions. We have implemented several *reusable* Kepler actors that perform these operations, some of which are shown in the post-processing sub-workflow in Figure 4.



(a) (b)

Figure 5. (a) Output from Santa Ana Workflow shows weather stations (placemarks) and regions (red polygons) experiencing Santa Ana winds; (b) FARSITE simulator output executed by a Kepler workflow.

Figure 5(a) shows the results for a small region in SD County: the green points are the weather stations and the red polygons are the areas experiencing Santa Ana winds. As can be seen from the figure, these polygons show many large areas surrounding the stations in addition to areas that do not surround the stations. In the next version of this workflow, we plan to run WRF for more accurate wind calculations and provide a comparison of WRF and WindNinja results.

4.2.2. Use Case 2: Run Fire Growth Model Workflows

The second use case simulates fires in SD County. For this application, we created a Kepler workflow to run FARSITE, an open-source fire growth simulator [10]. The inputs are the topography, fuels, weather conditions, and fire ignition site(s), and the outputs are fire perimeters and intensity, flame length, and spread rate. As in the previous use case, the workflow uses the REST interface to download the weather station measurements. Figure 5(b) shows the perimeters of two fires with the same ignition location. The fire with white perimeters had "normal" weather conditions, while the fire with red perimeters had wind speeds and relative humidity similar to Santa Ana Winds.

The first version of this workflow runs a single FARSITE instance. In a future version, we plan to run FARSITE in parallel for different starting ignition locations and weather conditions. We are also extending this workflow with data assimilation based on our new parameter and state estimation techniques as described in Section 3.2.

5 Conclusions

We presented our approach and early results from the multi-disciplinary dynamic data-driven WIFIRE project. Although some of this is work in progress, the reaction of the wildfire research community to the project architecture and early results has been very encouraging. We believe that

sharing these early results with the DDDAS community will enhance our understanding of the existing data assimilation efforts and similar work in other disciplines.

Acknowledgements. WIFIRE team is thankful for the valuable insights shared by our advisory board members and the collaboration of the wildfire modeling research community in guiding our data integration and modeling goals.

References

[1] Cedar Fire Website (2015) [Online]. Available: http://en.wikipedia.org/wiki/Cedar_Fire.
[2] "Cost of California wildfires is more than $1 billion," NPR, October 2010. [Online]. Available: http://www.npr.org/templates/story/story.php?storyId=15603441
[3]"http://www.npr.org/blogs/thetwo-way/2011/06/14/137170230/arizona-fire-about-to-be-states- biggest-ever," NPR, June, (2011).
[4] High performance wireless research and education network (2015) [Online] Available: http://hpwren.ucsd.edu
[5] B. Ludaescher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger-Frank, M. Jones, and E. Lee, and Y. Zhao, "Scientific workflow management and the Kepler system," Concurrency and Computation: Practice & Experience, Special Issue on Scientific Workflows, vol. 18, pp. 1039–1065, (2006).
[6] D. Barseghian, I. Altintas, M. B. Jones, D. Crawl, N. Potter, J. Gallagher, P. Cornillon, M. Schildhauer, E. T. Borer, E. W. Seabloom, and P. R. Hosseini, "Workflows and extensions to the Kepler scientific workflow system to support environmental sensor data access and analysis," Ecological Informatics, vol. 5, no. 1, pp. 42–50, 2010.
[7] J. Wang, D. Crawl, I. Altintas, and W. Li, "Big Data Applications using Workflows for Data Parallel Computing," IEEE Computing in Science & Engineering, 16(4), pages 11-22, July-Aug., (2014).
[8] R. Rothermel, "A mathematical model for predicting fire spread in wildland fuels," in INT-RP-115, USDA Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, 1972, p. 52.
[9] P. Andrews, C. Blevins, and R. Seli, "BehavePlus fire modeling system, v4.0: Users guide," US Dept. of Agriculture Forest Service, Rocky Mountain Res. Station, RMRS-GTR-106WWW, (2008).
[10] Finney, M.A.: FARSITE, Fire Area Simulator–model development and evaluation. RMRS-RP-4, Ogden, UT: U.S. Dept. of Agriculture, Forest Service, Rocky Mountain Research Station (1998).
[11] I. Nobel, G. Bary, and A. Gill, "McArthurs fire-danger meters expressed as equations," Australian J. Ecology, vol. 5, pp. 201–203, 1980.
[12] K. Hirsch, "Canadian forest fire behavior prediction (fbp) system: user's guide," Canadian Forest Service, Northwest Region, Northern Forestry Centre, Tech. Rep. 7, 1996.
[13] J. Mandel, L. Bennethum, J. Beezley, J. Coen, C. Douglas, K. M., and A. Vodacek, "A wildland fire model with data assimilation," Math. Comput. Simulat., vol. 79, pp. 584–606, (2008).
[14] M.C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, and A. Trouvé, "Towards predictive data-driven simulations of wildfire spread – Part I: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model for parameter estimation", Nat. Hazards Earth Syst. Sci., Vol. 14, pp. 2951-2973, (2014).
[15] H. Fang, P. Franks, and R.A. de Callafon, "Smoothed Estimation of Unknown Inputs and States in Dynamic Systems with Application to Oceanic Flow Field Reconstruction," International Journal of Adaptive Control and Signal Processing, DOI: 10.1002/acs.2529 (2014).
[16] C. Wunsch, "Discrete Inverse and State Estimation Problems", Cambridge University Press (2006).
[17] M.C. Rochoux, Emery, C. Ricci, S., Cuenot, B. and Trouvé, A. "Towards predictive data-driven simulations of wildfire spread – Part 2: Ensemble Kalman Filter for the state estimation of a front-tracking simulator of wildfire spread", Nat. Hazards Earth Syst. Sci., Vol. 2, pp. 3769-3820, (2014).
[18] H. Fang, R.A. de Callafon, J. Cortés, "Simultaneous input and state estimation for nonlinear systems with applications to flow field estimation," Automatica Vol. 49, pp. 2805-2812, (2013).
[19] I. Altintas, O. Barney, and E. Jaeger-Frank, "Provenance collection support in the Kepler scientific workflow system," in Provenance and Annotation of Data (IPAW 2006, Revised Selected Papers), ser. Lecture Notes in Computer Science, vol. 4145, 2006, pp. 118–132, (2006).
[20] Spark MLlib (2015) [Online]. Available: http://spark.apache.org/mllib.
[21] W.R. Sherman, M.A. Penick, S. Su, T.J. Brown, and F.C. Harris Jr, "VRFire: An immersive visualization experience for wildfire spread analysis," IEEE Virtual Reality Conf., 243-246, (2007).
[22] Open Scene Graph Doc. (2015) [Online]. Available: http://osgearth.org/wiki/Documentation.
[23] J.M. Forthofer, K. Shannon, B.W. Butler, "Simulating diurnally driven slope winds with WindNinja," In: 8th Symposium on Fire and Forest Meteorological Society (2009)