DYNAMIC DATA DRIVEN WILDFIRE MODELING

J. MANDEL[a], M. CHEN[a], J.L. COEN[b], C.C. DOUGLAS[c], L.P. FRANCA[a],
C. JOHNS[a] , R. KREMENS[d], A. PUHALSKII[a], A. VODACEK[d], W. ZHAO[e]

[a] *University of Colorado Denver, Denver, CO 80217-3364, USA*

[b] *National Center for Atmospheric Research, Boulder, CO 80307-3000, USA*

[c] *University of Kentucky, Lexington, KY 40506-0045, USA*

[d] *Rochester Institute of Technology, Rochester, NY 14623-5603, USA*

[e] *Texas A&M University, College Station , TX 77843-1112, USA*

Abstract. A proposed system for real-time modeling of wildfires is described. The system involves numerical weather and fire prediction, automated data acquisition from Internet sources, and input from aerial photographs and sensors. The system will be controlled by a non-Gaussian ensemble filter capable of assimilating out-of-order data. The computational model will run on remote supercomputers, with visualization on PDAs in the field connected to the Internet via a satellite.

1.   INTRODUCTION

Today, there exists virtually no capability to predict the spread of wildfires. It is of vital national interest that robust computer tools with known error and reliability analysis be developed now. In 2000 alone (the worst fire season in over 50 years), over 90,000 wildfires cost an estimated $10 billion in suppression and lost resources. Wildland fire is a timely application that presents many substantial challenges in real time communication, secure network transmission, mathematical modeling, and massively parallel computation.

Current field tools for diagnosing expected fire behavior are simple algorithms that can be run on simple pocket calculators. Researchers and fire managers alike envision a future when they can rely on complex simulations of the interactions of fire, weather, and fuel, driven by remote sensing data of fire location and land surface properties. *And* get the results as easy to understand animations on a small laptop or PDA, have the computer incorporate all information as soon as it becomes available, and assess the likelihood of possible scenarios. This is how computers work in the imagination of sci-fi movie directors. This project is to build a prototype as the first step to make it a reality.

Consider some typical questions that need constantly updated answers in a full scale wildland fire disaster: what is the best case scenario (in terms of forecasted meteorology, fire fighting strategies, and where a fire can be contained), what is the worst case scenario, and what is the most likely scenario? An example of the type of question a firefighter might ask is, "We know the lake will contain the fire on the east side, no matter what the weather is. To the north is a paved road. That should contain the fire, unless we get dry weather over the next few days and really strong

southerly winds. If we cannot contain it there, there are no fuel breaks and we may not be able to stop it in that unbroken forest for 3 miles. Meanwhile, we have to put out evacuation announcements: under the range of possible weather conditions, which communities will need to be evacuated?" Even without prediction, getting the existing information to the decision makers is not easy. The top problem reported by participants in a fire workshop [31] was "the need for timely data, as well as a means to track resources available to them."

2.   OVERVIEW OF THE PROPOSED SYSTEM

The objective of this project is to develop a prototype hardware and software system to integrate available information related to a wildfire in progress, and to provide a numerical prediction of the wildfire in a visual form, including tools to predict the outcome of firefighting strategies. The proposed system will have the following main components (Figure 1):

- Numerical coupled atmosphere/fire model
- Data acquisition (measurements)
  - From Internet: maps (GIS), aggregated fire information, weather
  - Field information: aerial photos, sensors
- Visualization and user interface
- Dynamic Data Assimilation control module
- Guaranteed secure communication infrastructure

The numerical model accepts data in a mesh format. The Dynamic Data Assimilation control module will call the numerical model to execute multiple simulations, extract data from the state of the numerical model to be compared with the measurements and modify the state of the numerical model to match the measurements. The visualization and user interface module will display the results of the simulation and support user input to control alternative firefighting scenarios.
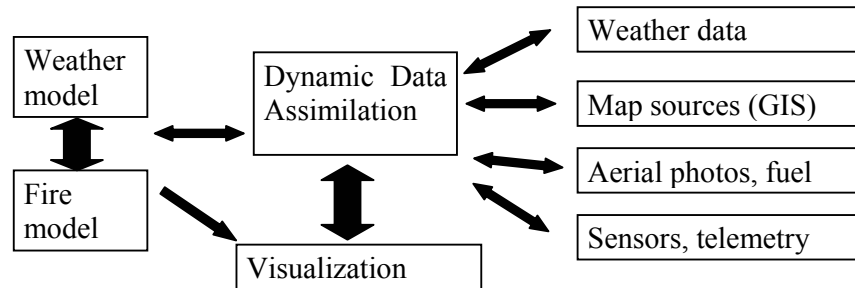

Figure 1. Wildfire model system architecture

The numerical model will run on one or more remote supercomputers, while the visualization and user interface will run on PDAs in the field. Software agents and search engines will provide internet data, while networked sensors and cameras in airplanes will provide the field data. The field internet connection will be

implemented using a local wireless network, bridged to the internet by a consumer-grade cheap broadband satellite connection and satellite data phones.

3.   WEATHER AND FIRE MODEL

3.1   *The NCAR coupled atmosphere-fire model*
NCAR's coupled atmosphere-fire model is described in detail in [9, 12, 13, 14]. An atmospheric prediction model [7, 8, 10, 11] has been coupled with an empirical fire spread model such that sensible and latent heat fluxes from the fire feed back to the atmosphere to produce fire winds, while the atmospheric winds drive the fire propagation.   This wildfire simulation model can thus represent the complex interactions between a fire and local winds (Figure 2).
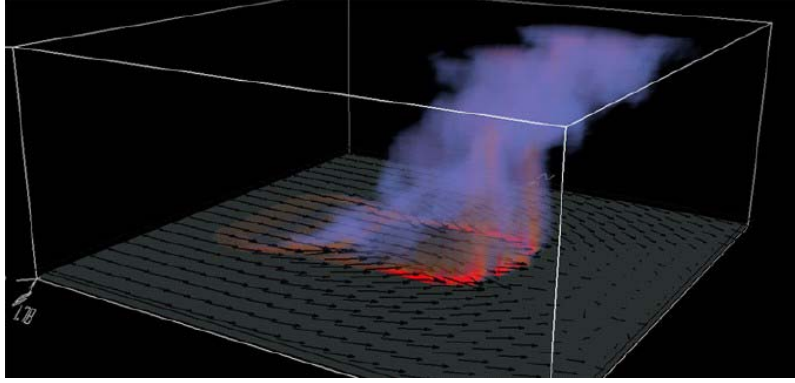

Figure 2. A characteristic fire shape in the NCAR model

The meteorological model is a three-dimensional non-hydrostatic numerical model based on the Navier-Stokes momentum, thermodynamic, and conservation of mass equations using the anelastic approximation. Vertically-stretched terrain-following coordinates allow the user to simulate in detail the airflow over mountainous topography. Currently, it can ingest forecasted changes in the larger-scale atmospheric environment by initializing the modeling domain and updating later boundary conditions with forecasts available at start-up. Its two-way interactive nested grids capture the outer forcing domain scale of the environmental mesoscale winds while allowing the user to telescope down to the meter-sized fine dynamic scales of vortices in the fireline through horizontal and vertical grid refinement. Cloud physics are approximated using a two-species (cloud droplets and rain) warm rain parameterization and a three-category (ice crystal, pristine snow, and graupel/hail) ice-phase parameterization.

Local fire spread rates depend on the modeled wind components through an application of the BEHAVE fire spread rate formula, based upon [29]. A BURNUP-type algorithm [1] characterizes how the fire consumes fuels of different sizes with time after ignition, distinguishing between rapidly consumed grasses that release

heat quickly and slowly consumed logs. Within each atmospheric grid cell, the land surface is further divided into fuel cells, with fuel characteristics corresponding to the 13 Anderson Fuel Models [2]. Four tracers, assigned to each fuel cell, identify burning areas of fuel cells and define the fire front. Fire spread rates are calculated locally along the fire line with wind vectors from the atmospheric model (which includes the effects of the fire), while a local contour advection scheme assures consistency along the fireline and avoids any ghosting effects [28]. The fire model has a simple formulation for canopy drying and ignition and a simple radiation treatment for distributing the sensible and latent heat into the lowest atmospheric grid levels.

3.2 Software reengineering of the NCAR model

The existing numerical model is a legacy FORTRAN code that took a quarter of century and substantial scientific expertise to develop. We are proceeding in these steps:

- Encapsulate execution of one time step as a subroutine
- Define and enforce software interfaces
- Upgrade the fire model from empirical to differential equations based
- Speed up the model using techniques such as OpenMP and multigrid

The overall system will run many instances of the model simultaneously; we expect that each instance of the model will run on a single CPU or an SMP node.

3.3 Modeling the fire by stochastic partial differential equations

The next step up from an empirical model of fire propagation is a model based on differential equations describing simplified fire physics. For some models based on physics, see [16, 4], and for a very detailed (and computationally expensive) coupled model of atmosphere and fire, see [22]. Our current model is a reaction-convection-diffusion equation, similar to [4], with additional stochastic terms, which also model spotting (secondary fires started by flying embers). The model variables are the temperature $T=T(t,x)$ and the fuel supply $S=S(t,x)$. The first equation is the balance of heat, given by

$$T(t) = T(t_0) + \int_{t_0}^{t} \left[-\nabla(k(T)\nabla T) + c_1 \cdot \nabla T - c_2(T - T_a) + c_3 \frac{\partial S}{\partial t} + \sigma \right] d\tau + E, \quad (1)$$

where $-\nabla(k(T)\nabla T)$ is diffusion of heat (e.g., by radiation or by turbulence), $c_1 \cdot \nabla T$ is convection (the coefficient c_1 incorporates the effect of the wind), $-c_2(T - T_a)$ is the heat loss to the atmosphere, $c_3 \partial S / \partial t$ is the heat generated by the burning fuel, σ is white noise, and

$$E(t,x) = \int g(t,x,y)(T(t,y) - T(t,x)) dy \quad (2)$$

is the heat transported by flying embers, with $g(t,x,y)$ a random process the values of which are sums of Dirac delta functions. The probability that $g(t,x,y) \neq 0$ drops off with distance and it depends also on the wind and on the fire temperature at the source location y of the embers. The reason for the equation (1) to be written in an integral form is that the secondary fires started by the embers show as a local jumps in the temperature. The effect of the term (2) is that at time t, the temperature at the location x is replaced by the temperature at a randomly selected location y, which models the randomly flying embers. The equation (1) is coupled with the balance of fuel equation

$$\frac{\partial S}{\partial t} = -f(T,T_i)S , \tag{3}$$

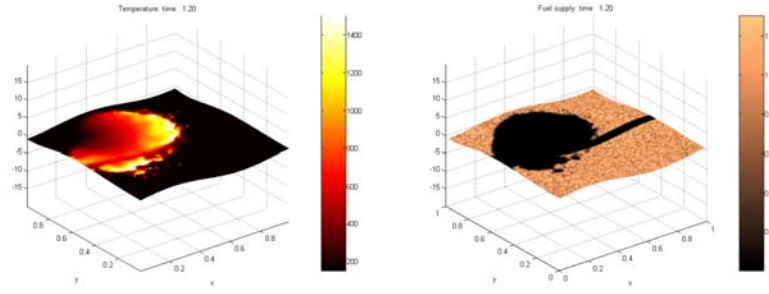where $f(T,T_i)$ is the intensity of burning and T_i is the ignition temperature.

Figure 3. Wind driven wildfire jumping a fire line, modeled by the stochastic reaction-convection-diffusion equation with spotting (1)-(3). Left: fire temperature, Right: fuel supply.

Anticipated developments of the fire model include several species of fuel to model different types of fire (grass, brush, crown fire), spotting terms with delay to account for the time it takes for the fire started by the flying embers to become visible on the computational grid scale, and of course coupling with the NCAR atmosphere model.

4.   DATA ACQUISITION

4.1   Internet data: geographical, weather, and fire information

Maps, sometimes including fuel information, are available from GIS files maintained by public agencies as well as private companies. We are working on translating the GIS format files into meshed data suitable for input into the model. Raw as well as assimilated weather data are readily available from numerous sources on the Internet, including NOAAPORT broadcast, MesoWest weather stations, and the Rapid Update Cycle (RUC) weather system by NOAA. Aggregated fire information is available from the GeoMAC project at USGS (Figure 4). The

challenge here is to develop intelligent converters of the information, which can deal with the constantly changing nature of Internet sites.
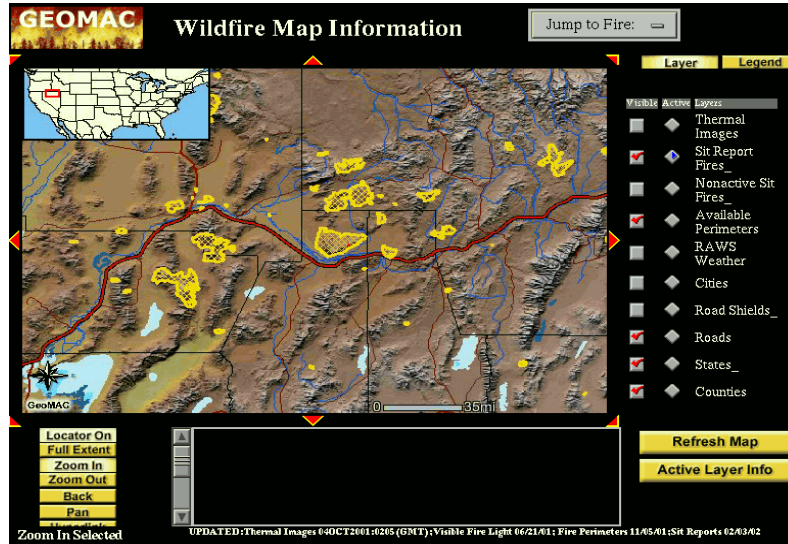

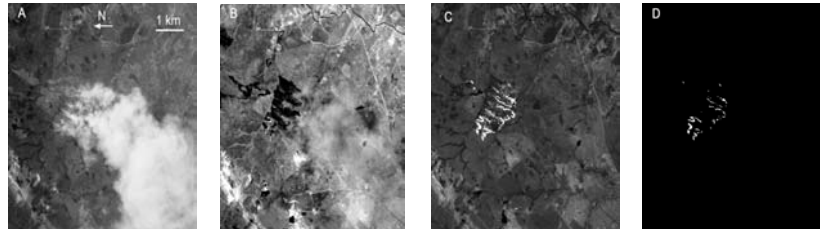*Figure 4. Example of fire boundary information from the GeoMAC project.*

Figure 5. AVIRIS images of Cuiaba, Brazil, at 589 nm (a), 770 nm (b), 1501 nm (c), and a ratio image (770 nm/779 nm) (d). The ratio image (d) isolates the atomic emission line of potassium observed in the 770 nm channel and has a threshold applied to show as white only those pixels that are six standard deviations or more above the mean ratio value for the entire image.. (Reproduced from [32] with permission.)

4.2   Image data

Thermal and near infrared images obtained from a manned or unmanned airborne platform is perhaps one of the best means of tracking the advance of wildland fires [27, 32]. First, even for cases where a fire location is completely obscured to the naked eye by smoke, infrared radiation will pass through smoke relatively unattenuated, thus providing a signal of the exact fire location (compare Figures 5a

and c). Second, the overhead view from an aircraft allows access to remote areas or areas of rough terrain. Finally, the full extent of most fires can be viewed rapidly from an aircraft. These types of images can be processed to identify fire perimeters, hotspots, and possibly areas of rapid fire advance (Figure 5d).   All of this information can feed into fire propagation models as input data or compared to model output for accuracy assessment. Either use of the data requires matching of the image data to the model grid.
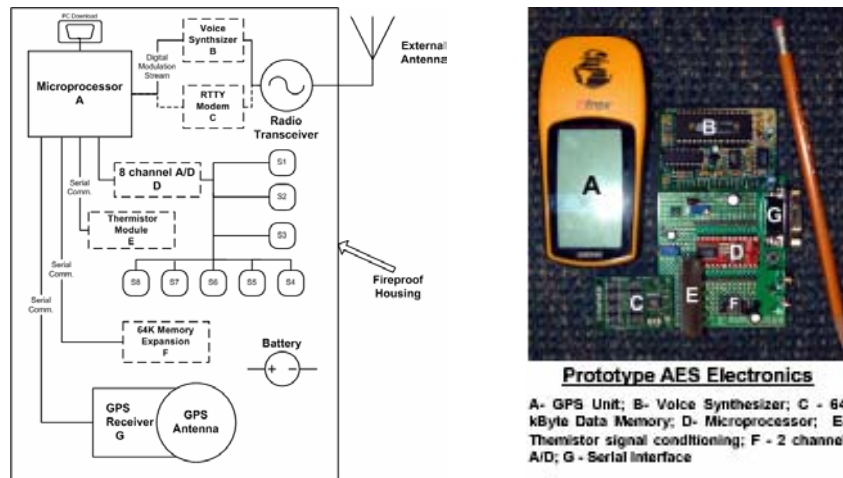

Figure 6. Left: Block diagram of the AFD (Reproduced from [21]); Right: an AFD prototype.

There are three steps to obtaining the geographic coordinates for the ground locations of individual pixels in an image that may correspond to the fire front or hotspot. First, the location and the direction that the camera was pointing at the time of image capture can be obtained by a combination of GPS readings for aircraft position and 3-axis inertial measurement unit data to determine the pointing direction of the camera. Second, the optical characteristics of the camera, such as distortion inherent in the lenses, can be assessed in a laboratory setting using test targets.   Third, there must be compensation for distortions caused by variable terrain.   One way to account for terrain induced distortion is to make parallax measurements on overlapping stereo images, but this imaging requirement adds complexity. However, if the terrain is known *a priori*, then the distortion in the images can be removed given knowledge of the camera pointing direction and altitude [24].   While topographic maps exist for much of the U.S., one very promising data set for the U.S. and much of the rest of the world is from the Shuttle Radar Topography Mission [26]. This data set has the distinct advantage of being created by a single measurement instrument over nearly all of the land mass of the world, greatly increasing the relative accuracy of the data. Further the horizontal (

30 m) and vertical (<10 m) resolutions of the data are in the range of expected model grid spacing.

4.3   Field data

Networked, autonomous environmental detectors (Figure 6) may be arbitrarily placed in the vicinity of a fire for point measurements of fire and weather information [21].   For geolocation, the units can either be built with a GPS capability or a less expensive approach is to plan for a GPS position to be taken at the exact location of the deployment using an external GPS unit and uploading the position data to the detector memory.

5.   DYNAMIC DATA ASSIMILATION

An important feature of DDDAS is that the model is running all the time and it incorporates new data as soon as it arrives. Also, in this application, uncertainty is dominant because important processes are not modeled, there are measurement and other errors in the data, the system is heavily nonlinear and ill-posed, and there are multiple possible outcomes. This type of problem is a natural candidate for *sequential Bayesian filtering* (sequential because the data is incorporated sequentially as it arrives rather than all at once). The state of the system at any time period is represented by a collection of physical variables and parameters of interest, usually at mesh points. To be able to inject data arriving out of sequence, we will work with the *time-state vector x,* which will contain snapshots of system states at different points in time. The knowledge of the time-state of the system is represented in the model as a *probability density function p(x).* The model will represent the probability distribution using an *ensemble of time-state vectors* $x_1, \dots, x_n$. Thus, the number of the system states maintained by the model will be equal to the number of time snapshots saved for updating out of order data multiplied by the ensemble size, a product that will easily reach the thousands. Each of these system states will be advanced in time via separate simulations and substantial parallel supercomputing power will be required.

Sequential filtering consists of successive updating of the model state using data via Bayes theorem. The current state of the model is called the *prior* or the *forecast* probability density $p^f(x)$. The arriving data consists of measurements *y* along with an information how the measured quantities are derived from the system state *x*, and information about the distribution of measurement errors. That is, the additional information provided by the data to the model is represented by the vector *y* and a conditional probability density function *p(y|x)*. From Bayes theorem, the update is incorporated into the model, resulting in the *posterior* or *analysis* probability density

$$p^a(x) = \frac{p(y|x)p^f(x)}{\int p(y|\xi)p^f(\xi)d\xi} , \qquad (4)$$

which then becomes the new state of the model. The model states continue advancing in time,

$$x_{t+\Delta t} = f(x_t),\tag{5}$$

with a corresponding advancement of the probability density, until new data arrives. The system states for times within the time stamps of the data are system estimation. The system states beyond the time stamp of the data constitute a prediction. Note that advancing the model in time and injecting the data into the model is decoupled.

Clearly, the crucial question is how the probability distributions are represented in the model. The simplest case is to assume normal distributions for the states x and the conditional distributions $p(y|x)$ and that the observations y depend linearly upon the states ,

$$p^f(x) \sim N(x^f, \Sigma), \;\; y = Hx + \varepsilon, \;\; \varepsilon \sim N(0, R).\tag{6}$$

These assumptions give the well-known Kalman filter formulas [20]. In particular, the posterior is also normal with the mean equal to the solution x^a of the least squares problem

$$\min_{x^a}\left\{(x^a - x^f)^T \Sigma^{-1}(x^a - x^f) + (y - Hx^a)^T R^{-1}(y - Hx^a)\right\}.\tag{7}$$

Ensemble Kalman filters [17, 18, 19, 30] represent a normal distribution by a sample, which avoids manipulation of covariance matrix Σ, replacing it with sample covariance estimate. Thus, the ensemble approach is particularly useful in problems where the state vector x or the observation vector y are large. However, existing methods do not support out of order data, because they operate on only one time snapshot, and extensions of ensemble filters to the time-state model described here have to be worked out. Further, if the time evolution function f in (5) is nonlinear or the observation matrix H in (7) is replaced by a nonlinear function, the probability distribution $p(x)$ is not expected to be normal. Particle filters [3, 15] can approximate any probability distribution, not only normal, by a sample of vectors $x_1, \ldots, x_n$, with corresponding weights $w_1, \ldots, w_n$; then approximately $x \sim D$, where the probability of any event ω is calculated by

$$D(\omega) = \sum_{x_i \in \omega} w_i.\tag{8}$$

Time-state models have been used in processing out-of-order data by particle filters [23]. But the approximation (8) is rather crude and, consequently, particle filters generally require large samples, which is not practical in our application. For this reason, we are investigating data assimilation by ensemble filters based on Gaussian mixtures [5, 6], where the probability distributions are approximated by a weighted mixture of normal distributions, such as

$$x \sim N(\mu_i, \Sigma_i) \text{ with probability } p_i. \tag{9}$$

To assimilate data that depends on system state in a highly nonlinear way, such as fire perimeters from aerial photographs, one of the options we are investigating is replacing the least squares problem (7) by the nonlinear optimization problem

$$\min_{x^a}\left\{(x^a - x^f)^T \Sigma^{-1}(x^a - x^f) + (y - h(x^a))^T R^{-1}(y - h(x^a))\right\}, \tag{10}$$

which may have multiple local minima. Alternatively, we consider updating the model state by minimizing some probabilities that the solution is off by more than a given amount, e.g.,

$$\min_{x^a}\left\{P(\neg \,|\, x^a - x^f \,|< \varepsilon_x \;\;\vee\;\; \neg\, |y - h(x^a)|< \varepsilon_y)\right\}, \tag{11}$$

where the inequalities are understood term by term, and ε_x, ε_y are vectors of tolerances. This type of threshold estimates, related to Bahadur efficiency [25], could be quite important in fire problems, because of the nature of fires. Unlike in the case when the distribution of x is normal, here a least squares estimate may say little about whether or not the ignition temperature in a given region has been reached.

The Dynamic Data Assimilation module needs also to steer the data acquisition. In terms of the Bayesian update (4), one method of steering would be to select the measurements to minimize the variance of the posterior distribution. In the linear Gaussian case (6), this becomes an optimization problem for the observation matrix H under the constraints of what measurements can be made.

Finally, guaranteed secure communication [33] delivers within a preset delay with a given probability close to one. This means that loss of data or loss of communication with ensemble members could be, in principle, handled as a part of the stochastic framework, and the Dynamic Data Assimilation module should take advantage of that.

REFERENCES

1. Albini, F. A. (1994).   PROGRAM BURNUP: A simulation model of the burningof large woody natural fuels. Final Report on Research Grant INT-92754-GR by U.S.F.S. to Montana State Univ., Mechanical Engineering Dept.
2. Anderson, H. (1982).   Aids to determining fuel models for estimating fire behavior. USDA Forest Service, Intermountain Forest and Range Experiment Station, Report INT-122.
3. Arulampalam, M., Maskell, S., Gordon, N. and Clapp, T. (2002).   A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Transactions on Signal Processing*, 50, pages 174-188.
4. Asensio, M. I. and Ferragut, L. (2002).   On a wildland fire model with radiation, *Int. J. Numer. Meth. Engrg*, 54, pages 137-157.
5. Bengtsson, T., Snyder, C. and Nychka, D. (2003).   A nonlinear filter that extends to high dimensional systems. *J. of Geophys. Res. - Atmosphere, in review.*
6. Chen , R. and Liu, J. S. (2000). Mixture Kalman filters, *J. of the Royal Statistical Society: Series B*, 62, pages 493-508.
7. Clark, T. L. (1977).   A small-scale dynamic model using a terrain-following coordinate transformation, *J. of Comp. Phys.*, 24, pages 186-215.
8. Clark, T. L. (1979). Numerical simulations with a three-dimensional cloud model: lateral boundary condition experiments and multi-cellular severe storm simulations, *J. Atmos. Sci.*, 36, pages 2191-2215.
9. Clark, T. L., Coen, J. and Latham, D. (2004).   Description of a coupled atmosphere-fire model, *Intl. J. Wildland Fire*, 13, in print.
10. Clark, T. L. and Hall, W. D. (1991).   Multi-domain simulations of the time dependent   Navier-Stokes equations: Benchmark error analysis of some nesting procedures, *J. of Comp. Phys.*, 92, pages 456-481.
11. Clark, T. L. and Hall, W. D. (1996).   On the design of smooth, conservative vertical grids for interactive grid nesting with stretching, *J. Appl. Meteor.*, 35, pages 1040-1046.
12. Clark, T. L., Jenkins, M. A., Coen, J. and Packham, D. (1996).   A coupled atmospheric-fire model: Convective feedback on fire line dynamics, *J. Appl. Meteor*, 35, pages 875-901.
13. Clark, T. L., Jenkins, M. A., Coen, J. and Packham, D. (1996).   A coupled atmospheric-fire model: Convective Froude number and dynamic fingering, *Intl. J. of Wildland Fire*, 6, pages 177-190.
14. Coen, J. L., Clark, T. L. and Latham, D. (2001).   Coupled atmosphere-fire model simulations in various fuel types in complex terrain, in 4th. Symp. *Fire and Forest Meteor. Amer. Meteor. Soc., Reno*, Nov. 13-15, pages 39-42.
15. Doucet, A., Freitas, N. de and Gordon, N. (eds), (2001).   *Sequential Monte Carlo in Practice, Springer.*
16. Dupuy, J. L. and Larini, M. (1999).   Fire spread through a porous forest fuel bed: A radiative and convective model including fire-induced flow effects, *Intl. J. of Wildland Fire*, 9, pages 155-172.
17. Evensen, G. (1994). Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res*.,99 (C5), pages 143-162.
18. Evensen, G. (2003).   The ensemble Kalman filter: Theoretical formulation and practical implementation. http://www.nersc.no/geir .
19. Houtekamer, P. and Mitchell, H. L. (1998).   Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review*, 126, pages 796-811.
20. Jazwinski, A. H. (1970). Stochastic processes and filtering theory, Academic Press, New York.
21. Kremens, R., Faulring, J., Gallagher, A.,Seema, A. and Vodacek, A. (2003). Autonomous field-deployable wildland fire sensors, *Intl. J. of Wildland Fire*, 12, pages 237-244.
22. Linn, R. Reisner, J.,Colman, J. and Winterkamp, J. (2002).   Studying wildfire behavior using FIRETEC, *Int. J. of Wildland Fire*, 11, pages 233-246.
23. Mallick, M., Kirubarajan, T. and Arulampalam, S. (2002).   Out-of-sequence measurement processing for tracking ground target using particle filters, *in Aerospace Conference Proceedings*, vol. 4, *IEEE*, pages 4-1809-4-1818.
24. Mostafa, M., Hutton, J. and Lithopoulos, E. (2000).   Covariance propagation in GPS/IMU - directly georeferenced frame imagery, in Proceedings of the Asian Conference on Remote Sensing

2000, Taipei, Center for Space; Remote Sensing Research, National Central University and Chinese Taipei Society of Photogrammetry and Remote Sensing.

25. Puhalskii, A. and Spokoiny, V. (1998). On large-deviation efficiency in statistical inference, *Bernoulli*, 4, pages 203-272.

26. Rabus, B. Eineder, M. Roth, A. and Bamler, R. (2003). The shuttle radar topography mission - a new class of digital elevation models acquired by spaceborne radar, *Photogrammetric Engineering and Remote Sensing*, 57, pages 241-262.

27. Radke, L. R., Clark, T.L., Coen, J. L., Walther, C., Lockwood, R. N., Riggin, P., Brass, J. and Higgans, R. (2000). The wildfire experiment (WiFE): Observations with airborne remote sensors, *Canadian J. Remote Sensing*, 26, pages 406-417.

28. Richards, G. D. (1994). The properties of elliptical wildfire growth for time dependent fuel and meteorological conditions, *Combust. Sci. and Tech*., 95, pages 357-383.

29. Rothermel, R. C. (1972). A mathematical model for predicting fire spread in wildland fires. USDA Forest Service Research Paper INT-115.

30. Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. (2003). Ensemble square root filters, *Monthly Weather Review*, 131, pages 1485-1490.

31. Vodacek, A. (2001). Wildland fire needs assessment workshop report. http://www.cis.rit.edu/research/dirs/research/fires.html .

32. Vodacek, A., Kremens, R. L., Fordham, A. J., VanGorden, S. C., Luisi, D. and Schott, J. R. (2002). Remote optical detection of biomass burning using a potassium emission signature, *Intl. J. of Remote Sensing*, 13, pages 2721-2726.

33. Zhao, W. (2001). Challenges in design and implementation of middlewares for real-time systems, *J. of Real-Time Systems*, 20, pages 1-2.