International Journal of Remote Sensing
Vol. 30, No. 1, 10 January 2009, 57–83

Image construction using multitemporal observations and Dynamic Detection Models

A. KOLTUNOV*†‡, E. BEN-DOR‡ and S. L. USTIN†
†Center for Spatial Technologies and Remote Sensing (CSTARS), University of California Davis, The Barn, One Shields Avenue, Davis, CA 95616, USA
‡Department of Geography and the Human Environment, Tel Aviv University, P.O.B. 39040, Ramat-Aviv, Tel-Aviv, 69978, Israel

(*Received 11 October 2006; in final form 12 August 2007*)

This paper systematically derives and analyses the generic phenomenon of space-invariant predictability of spatio-temporal observation fields using past multitemporal observations. We focus on thermal infrared remote sensing as a non-trivial example illustrating the predictability concept. The phenomenon and the systematic analysis thereof are experimentally demonstrated to be productive for developing effective automated anomaly detection and classification methods operating under the assumption of *dynamic* environment and sensor response. Using a simple preliminary experiment involving uncalibrated tower-based high-resolution thermal infrared surveillance, we test the conceptual validity of the space-invariant multitemporal prediction and exemplify its potential applications. In addition, we use a MODIS thermal image sequence and the task of hot anomaly detection to demonstrate the applicability of the approach for monitoring the status of large territories from space-borne platforms.

1. Introduction

The increased quantity, quality, and informational content of multitemporal remote sensing datasets, as well as shorter revisit time of modern space-borne platforms, allow new and complex thematic tasks to be defined, and hence, require that the scientific community develop new and effective methodologies for data analysis. Multitemporal technique development has focused mainly on the reflective spectral domain, while the thermal infrared (TIR) domain has received much less attention in the remote sensing literature. This paper discusses the concept of combining past uncalibrated thermal images of a scene to predict the pixel intensities of a current image.

Most applications of multitemporal analysis involve (directly or indirectly) a comparison of measurements from objects across time. However, under unsteady environmental conditions (e.g. gusty wind or partial cloudiness), the pixel intensities of thermal images may change considerably within minutes. Even when the weather is relatively stable (weak wind, few clouds), intervals as short as 15–20 minutes between collecting the previous and the current image data may adversely affect classification or change detection, unless the target and background classes are well-separated.

Instead of comparing pixel intensities, it is reasonable to retrieve and compare time-invariant quantities. In thermal infrared imagery, this entails estimating temperature

---

*Corresponding author. Email: akoltunov@ucdavis.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2009 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160802220193

and emissivity from calibrated thermal data. This is a complex problem that has not yet been completely resolved (Dash *et al.* 2002). Even if it had been resolved, much of the thermal information separating targets from background lies exactly in the temperatures *which are not useful*, because they themselves depend on *dynamic* weather conditions. Hence, the absolute value of the change in object temperature cannot always be used directly for detection or classification.

Jensen (1983) and more recently Schaum and Stocker (1997) have observed that a multispectral image of a scene can be predicted by a linear operator of a previous multispectral image (or their sequence) of that scene, provided all the images are accurately registered, i.e. the pixel intensities can be transformed *jointly*. They applied this observation to detect changes in a fixed scene and reported considerable improvements over the simple subtraction of images. Likewise, Carlotto (2000) has experimentally compared linear vs. non-linear predictors and applied them for change detection in Synthetic Aperture Radar (SAR) images, optical images, and for cross-sensor prediction. These methods of joint prediction share the underlying ideas and assumptions with the image normalization approach that has been proposed by Schott *et al.* (1988) and extended by Furby and Campbell (2001) and by Du *et al.* (2002). In this approach, pre-selected pseudoinvariant targets, such as roof tops, roads, parking lots, and other manmade objects, or natural targets, are used to determine the coefficients of a linear transformation between two images of the same area. This approach requires the linearity of the observation process model, the linearity of on-ground preprocessing operations, spatial invariance of the illumination and atmosphere conditions (in particular, no partial shadowing), temporal invariance of reflectance of a given target, image registration across time, and the negligibility of the bi-directional reflectance function (BRDF) effects due to nearly Lambertian surfaces or little temporal variability of the illumination source position. All this is often nearly true with remote sensing in a reflective domain. This approach proved effective for between-date calibration of Landsat TM reflective data of the same area.

1.1 *Problem statement and scope of the paper*

To date, the potential of a multitemporal prediction approach has not been fully realized, with uncertainties about its theoretical basis, general properties, and practical value beyond bi-date radiometric normalization of reflective images. General questions to be addressed, which are critical for constructing working methods, include:

- How many past images should be used for prediction?
- When should the past images be acquired?
- Is the band plurality critical to prediction?
- How can the underlying assumptions be weakened?

A problem where the usefulness of the approach is far from obvious, is monitoring the status of large territories using satellite thermal imagery. Indeed, the prediction quality would be reduced by various factors, including but not limited to:

- both temporally and spatially variable viewing geometry, due to wide-angle surveillance;
- misregistration and swath edge effects;
- spatially variable influence of weather conditions (e.g. dynamic partial cloud cover, local precipitation, etc.);

- many pixels with missing values;
- small undetected convective clouds;
- thin undetected clouds typically bordering large, easily detectable clouds;
- spatially and temporally variable sensor failures of non-extreme amplitude (e.g. stripes).

Thus, the applicability of multitemporal prediction for space-borne datasets needs experimental demonstration.

This paper discusses the theoretical basis, presents algorithms, describes experimental validation, and illustrates applications of the multitemporal prediction approach. The theoretical part (§2) consists of mathematically and physically based analyses of the approach, and we derive the space-invariant prediction model. Using these systematic analyses, we address (§3) the general questions posed above. We further present two methods based on multitemporal prediction: an anomaly (change) detection method and a classification method. Both methods use past uncalibrated imagery for constructing their decision rules under dynamic environmental conditions, and they will be generically referred to as Dynamic Detection Model(s) (DDM).

When applied for classification, the DDMs are attractive, because they enable a recognition system to significantly separate in time the stages of training and classification. As a result, the time-consuming labeling and statistical analysis of training data to build the classification rule can be done before the acquisition of inspection images. This leads to the previously unreported possibility of near-real-time application of powerful classification algorithms capable of discriminating between classes that have low separability. As an initial step in development, we introduce a DDM based on the space-invariant prediction of Gaussian mixture (GM) distributions of classes (§2.4). At the training stage, the DDM requires multitemporal surveying of the scene and determining the class parameters for each training survey. At the detection stage, the inspection image is used to transform the class GM parameters so that they become appropriate for the conditions of the detection survey. Finally, the pixels of the detection image are classified based on the transformed parameters.

The experimental work presented in this paper (§4) has several objectives. The first objective is to demonstrate the conceptual validity of the multitemporal prediction approach, which is done using a preliminary experiment with high-resolution thermal infrared tower-based remote sensing. Second, we apply the approach to detect anomalous temporal changes in a scene and compare the performance with that of simpler techniques. Additionally, we show the potential of the GM-based DDM for thematic classification of anomaly regions. Finally, we demonstrate the applicability of multitemporal prediction for space-borne datasets by using a MODIS (MODerate Resolution Imaging Spectroradiometer) image sequence. In this example, subpixel hot anomalies are detected at the 1-km ground resolution over a scene occupying more than 150 000 km^2.

2. Predictability and recognition under dynamic environmental conditions

We begin with a formal derivation of the abstract law (model) governing the predictability of a spatio-temporal observation field: the space-invariant prediction model. An example of an observation field is a remotely sensed image of a scene; and we will use the terms 'observation field' and 'image' interchangeably. The derivation does not rely on linearity or knowledge of the observation process; the model will be derived using a set of simplifying assumptions that are necessary to emphasize the principle. In §2.2, we explain why the assumptions under which the model is applicable, are much broader than those used in deriving the model.

We consider an observation process whose characteristics can be split into two disjoint groups:

- characteristics depending only on spatial location (e.g. geometric distortion);
- characteristics depending only on time (e.g. radiometric calibration coefficients).

Regarded in this way, the observation process can be multitemporal sensing under the same viewing conditions, possibly followed by a non-uniformity correction (e.g. destriping, correction for straylight, etc.).

2.1 *Space-invariant model for prediction in time*

Let $W=\{w_i\}_{i=1}^{N}$ denote a spatio-temporal field of physical observations, obtained via an observation process $\mathcal{F}$ (either known or not). For a given transform $\mathcal{F}$, an element (e.g. pixel intensity) w_i being observed at spatial location s at time t, depends on two types of factors: the internal properties (i.p.) of the observed objects, denoted by $\boldsymbol{\alpha}$, and the external influence factors (e.f.): $\mathcal{X}$. In real physical processes, both i.p. and e.f. depend on time t and spatial location s, and thus:

$$W=\mathcal{F}[\boldsymbol{\alpha}(s,t);\ \mathcal{X}(s,t)]. \tag{1}$$

For the sake of brevity of the derivation, we introduce several constraints into the above general model (1):

(a) internal properties $\boldsymbol{\alpha}$ do not change with time, i.e. $\boldsymbol{\alpha}(s, t)=\boldsymbol{\alpha}(s)$;
(b) external factors $\mathcal{X}(s, t)$ are the same for each element of W, i.e. $\mathcal{X}(s,t)=\mathcal{X}(t)$;
(c) each element w_i is autonomous, meaning that its i.p. are independent of the i.p. of other w_j;
(d) any w_i has a *finite* number m_i of internal properties;
(e) the number of observation field elements is finite.

If constrained by (a)–(e), model (1) can be written in the separate-variable form as follows:

$$W=\mathcal{F}[\boldsymbol{\alpha}(s);\ \mathcal{X}(t)]. \tag{2}$$

Measure the observation field at P time moments, *at which the external factors are essentially different*, thus obtaining a system of $N\cdot P$ equations

$$W_{t_j}\overset{\text{def}}{=}W(t_j)=\mathcal{F}[\boldsymbol{\alpha};\ \mathcal{X}(t_j)],\qquad \text{where}\quad j=1,\ \dots,P;\quad P\geqslant\max_i\{m_i\}, \tag{3}$$

wherein each element w_i appears P times. Therefore, if operator $\mathcal{F}$ is invertible for the chosen set $t_1, \dots, t_P$, then the vector of i.p. can be uniquely found from (3) as

$$\boldsymbol{\alpha}=\mathcal{G}[W_{t_1},\ \dots,W_{t_P};\ \mathcal{X}_{t_1},\ \dots,\mathcal{X}_{t_P}],\qquad \text{where}\quad \mathcal{X}_{t_j}\overset{\text{def}}{=}\mathcal{X}(t_j). \tag{4}$$

Since (2) *always* holds, we can introduce (4) into (2) to obtain *at any new time* $t=t_v$

(more precisely, for any new external conditions):

$$W_{t_v} \overset{\text{def}}{=} W(t_v) = \mathcal{H}_{t_v}[W_{t_1}, \ldots, W_{t_P}]. \tag{5}$$

Operator $\mathcal{H}_{t_v}$ depends only on e.f. $\mathcal{X}(t_v), \mathcal{X}_{t_1}, \ldots, \mathcal{X}_{t_P}$. Note that these factors should not be considered as only instantaneous influence factors. Some of the e.f. can be long-lasting, for example, weather history. However, the prediction operator does not depend on internal properties that are different for different elements w_i of the observation field. Hence, this operator transforms 'old' fields into a 'new' field *jointly*, element into element, according to the same law for all elements. The fields $W_{t_1}, \ldots, W_{t_P}$ will be called *basis* fields.

Thus, we have derived a statement that can be conveniently worded as follows: Any physical field currently observed can be arbitrarily accurately approximated by a space-invariant operator of a finite number of fields previously observed.

A practical, feasible process of prediction is as follows: given the basis fields and the new field currently observed, we represent $\mathcal{H}$ parametrically, using a presumed form $\widetilde{\mathcal{H}}$ of the model, which is

$$W_{t_v} \approx \widetilde{\mathcal{H}}[W_{t_1}, \ldots, W_{t_P}; \gamma(t_v)]. \tag{6}$$

The parameter vector γ is estimated, for example, in the least-squares sense using the field elements whose correspondence across time has been established (see §3.3). These elements will be termed *indicators* of prediction. Next, the estimates are used to predict the remaining elements of the observation field. It is reasonable to start searching for a good parametric model $\widetilde{\mathcal{H}}$ from a linear model.

2.2 *Prediction under less restrictive assumptions*

Model (5) has been derived from the observation process model (2) with separate dependencies: the internal properties of the objects depend only on spatial location, whereas the external influence depends only on time. In a general situation this may be not the case: both i.p. and e.f. may depend on space and time (see equation (1)). However, both the internal and external factors can always be parameterized approximately:

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}(s, t). \tag{7}$$

Denote by $\boldsymbol{\beta}(s) = \{\boldsymbol{\beta}_1(s), \boldsymbol{\beta}_2(s)\}$ the vector of space-dependent parameters of both groups ($\boldsymbol{\alpha}$ and $\mathcal{X}$), and by $\mathcal{Y}(t) = \{\boldsymbol{y}_1(t), \boldsymbol{y}_2(t)\}$ the vector of time-dependent parameters of groups $\boldsymbol{\alpha}$ and $\mathcal{X}$. Then the general model (1) becomes

$$W(s, t) = f[\boldsymbol{\beta}(s); \mathcal{Y}(t)], \tag{8}$$

which is, again, of the form (2). The group of parameters $\boldsymbol{\beta}(s)$ will be called *temporal invariants*, and the group $\mathcal{Y}(t)$ – *spatial invariants*. The analogue of system (3) is the following system:

$$W_{t_j} = f\left[\boldsymbol{\beta}; \mathcal{Y}(t_j)\right], \qquad j = 1, \ldots, P, \tag{9}$$

from which the space-invariant prediction model (5) is immediately derived following the way described in §2.1. The assumption that the i.p. values of a spatial location are independent of the i.p. values of other spatial locations can be removed analogously.

The above mathematics suggests that the prediction model is actually based on the general physical hypothesis that separable aspects (i.e. the invariants) play a dominant role in forming many real datasets. Consequently, the error introduced due to replacing observation process (1) with (8), can be sufficiently small for many thematic tasks. The actual range of circumstances in which the prediction model (5) is productive should be determined experimentally. Theoretically, we have shown the reasons why this range is broader than (a)–(e) of §2.1. To this end, it is rather interesting that model (5) with 20 basis images succeeded in the experiment (§4.1, not yet discussed), in which *none* of the conditions (a)–(d) of §2.1 were met. Furthermore, in §4.2 we will consider a real satellite dataset with many features *severely* violating the initial, simplifying assumptions (a)–(d). Despite these features, the separated-form approximation (8) of the real process (1) results in a useful detection method.

2.3 *Physical explanation of thermal image prediction*

In the previous sections we presented a general derivation of the joint prediction law for an unspecified spatio-temporal field of observations. Therefore, below we give a brief thermo-physical explanation of the predictability phenomenon.

Without assuming diurnally periodic weather conditions, we can represent the surface temperature T at time t in the following integral form (10) that is equivalent to equation (A9) derived in Appendix A:

$$T(t) = T(t_0) + \sum_{k=1}^{5} \alpha_k \int_{t_0}^{t} \frac{\mathcal{G}_k(\tau, T(\tau))}{\sqrt{t-\tau}} \mathrm{d}\tau, \tag{10}$$

where $\mathcal{G}_k(t)$ denotes the heat fluxes into or from the surface, as prescribed mostly by the surface energy balance; α_k denotes functions of thermo-physical and emissive properties; and t_0 is the initial time moment. As shown in (10), some of the fluxes on the right-hand side may also depend on the surface temperature, i.e. both on spatial location and time. In contrast, the iterative solutions of (10) (at any required accuracy) are of the form (8), which in turn implies a separable-form representation for radiances. Applying the procedure described in §2.1 to this representation results in a space-invariant predictor expressed in terms of physically meaningful parameters.

Appendix A provides the derivation details and the explicit formulas of the prediction operator under the simplest circumstances that can be briefly described as follows: a narrow-channel, radiometrically calibrated and flat-fielded long-wave TIR sensor surveys a planar scene from a fixed position at low altitude under arid conditions with no near-surface wind; and the observations are performed during a period of a stable anticyclone.

2.4 *Dynamic Detection Models*

Based on the space-invariant prediction, we can implement the DDMs as follows. At the recognition time the inspection image W_{t_v} is first registered toward the basis image sequence. To detect anomalies in the scene, the prediction coefficients $\gamma(t_v)$ corresponding to the inspection image are determined, along with the r.m.s. error of prediction, σ_{t_v}. Next, the coefficients are introduced into (6) and the predicted value of pixel intensity, $\widehat{W(s, t_v)}$, is computed by evaluating the right-hand side of (6).

Finally, the following test is applied:

$$\frac{\left| W(s, t_v) - \widehat{W(s, t_v)} \right|}{\sigma_{t_v}} > z. \tag{11}$$

The threshold z controls the sensitivity of the detector.

For a classification task an inspection image pixel with an intensity w_{t_v} is assigned to thematic classes by Bayes' formula. With the GM-based learning, both class-conditional and unconditional distributions of the data are Gaussian mixtures (McLachlan 1992, McLachlan and Peel 2001). At the training stage these distributions are estimated for each basis time moment $t_1, \ldots, t_P$, e.g. by using the procedure proposed by Koltunov and Ben-Dor (2004). At the recognition time, we estimate the parameter vector $\gamma(t_v)$ of the parametric prediction model (6). If we denote $\boldsymbol{w}_\mathrm{b} = (w_{t_1}, \ldots, w_{t_P})^\mathrm{T}$ (superscript 'T' standing for 'transpose') and exploit the normal distribution properties, then up to a linear approximation, the expectation $\mathbb{E}w_{t_v}$ and variance $\mathrm{Var}(w_{t_v})$ are determined by

$$\mathbb{E}w_{t_v} = \widetilde{\mathcal{H}}[\mathbb{E}w_{t_1}, \ldots, \mathbb{E}w_{t_P}; \gamma(t_v)], \qquad \text{and} \qquad \mathrm{Var}(w_{t_v}) = \nabla\widetilde{\mathcal{H}}^\mathrm{T} \cdot \mathbf{C}(\boldsymbol{w}_\mathrm{b}) \cdot \nabla\widetilde{\mathcal{H}}, \tag{12}$$

where $\nabla\widetilde{H}$ is the gradient of $\widetilde{H}$ with respect to $w_{t_1}, \ldots, w_{t_P}$ and $\mathbf{C}$ represents the covariance matrix of vector $\boldsymbol{w}_\mathrm{b}$. Formulas (12) are applied to each Gaussian component *separately* to obtain the class parameters corresponding to the recognition time. Then, we apply Bayes' formula, as usual, to compute the posterior probabilities of class membership for pixels, and assign pixels to classes by the Maximum A Posteriori Probability rule. For the case of the multi-channel feature vector the covariances of the i-th and j-th bands at time t_v are given by

$$\sigma\left(w^i_{t_v}, w^j_{t_v}\right) = \nabla\widetilde{\mathcal{H}}^\mathrm{T} \cdot \mathbf{C}^{ij}(\boldsymbol{w}_\mathrm{b}) \cdot \nabla\widetilde{\mathcal{H}}, \tag{13}$$

where $\mathbf{C}^{ji}(\boldsymbol{w}_\mathrm{b})$ is the $P \times P$ matrix formed by the cross-time covariances $\mathrm{cov}\left(w^i_{t_k}, w^j_{t_n}\right)$; $k, n = 1, \ldots, P$.

If in the course of the training stage the prediction operator $\widetilde{H}$ is chosen to be a linear operator of past image intensities $\boldsymbol{w}_\mathrm{b}$, then equalities (12)–(13) are exact. Also, these equalities hold exactly after an obvious modification in the case of non-linear $\widetilde{H}$, such that

$$\widetilde{\mathcal{H}}[\boldsymbol{w}_\mathrm{b}; \gamma(t_v)] = \gamma_1(t_v) g_1(\boldsymbol{w}_\mathrm{b}) + \cdots + \gamma_K(t_v) g_K(\boldsymbol{w}_\mathrm{b}),$$

where g_k can be arbitrarily complicated functions of vector $\boldsymbol{w}_\mathrm{b}$.

3. Practical aspects of multitemporal prediction and training

This section discusses several issues that are important for utilizing the image predictability phenomenon and DDMs in practice. Among these issues are cross-sensor prediction, the role of plurality of bands, selecting basis images or times, and selecting indicators of prediction.

3.1 *On mutually predictable fields and band plurality*

Suppose there are several observation systems that differ in the technical parameters of the sensors, for instance, band wavelengths or viewing angle. Then a reasonable question is when can the data of one system be used to predict the data of another system. Empirically, partial answers to this question for the cases of hyperspectral data and SAR–electro-optical data have been given in earlier work (Carlotto 2000, Schaum and Stocker 1997). This section adds more details to this issue.

Let $W^{(1)}$ and $W^{(2)}$ denote two observation fields corresponding to two different observation processes, ϕ and ψ:

$$W^{(1)}(s,t)=\phi\left[\boldsymbol{\beta}^{(1)}(s);\ \mathcal{Y}^{(1)}(t)\right],$$

$$W^{(2)}(s,t)=\psi\left[\boldsymbol{\beta}^{(2)}(s);\ \mathcal{Y}^{(2)}(t)\right].$$

Then, analogously to (5), one can obtain the prediction model (for $k=1$, $\ell=2$, or vice versa):

$$W_t^{(k)}=\mathcal{H}^{k\ell}\left[W_{t_1}^{(\ell)},\ \ldots,\ W_{t_P}^{(\ell)}\right],\qquad \text{if}\quad \boldsymbol{\beta}^{(1)}=\boldsymbol{\beta}^{(2)}, \tag{14}$$

where operator $\mathcal{H}^{k\ell}$ is a space-invariant operator. In other words, two fields with the same temporal invariants are mutually predictable in time. Mutual predictability in time (14) holds if the temporal invariants are equal, because $\boldsymbol{\beta}^{(\ell)}$, found by multitemporally measuring $W^{(\ell)}$, should replace $\boldsymbol{\beta}^{(k)}$ in the observation process model, as is shown in §2.1.

An example of mutually predictable fields $W^{(1)}$ and $W^{(2)}$ is when they are thermal images acquired in different wavelength intervals Λ_1 and Λ_2, respectively. When both bands Λ_1 and Λ_2 are sufficiently narrow and close to each other, the object's emissivity $\varepsilon(\lambda)$ is nearly the same for both bands, and for any temperature T the Planck's function (Price 1989) denoted by $\mathcal{B}$, satisfies the following approximate equalities:

$$\mathcal{B}(T,\lambda_1)\approx\mathcal{B}(T,\lambda_2)\qquad\text{and}\qquad\frac{\partial}{\partial T}\mathcal{B}(T,\lambda_1)\approx\frac{\partial}{\partial T}\mathcal{B}(T,\lambda_2), \tag{15}$$

for any wavelengths: $\lambda_1\in\Lambda_1$ and $\lambda_2\in\Lambda_2$. Furthermore, assume that $\varepsilon(\lambda)$ and the objects' thermo-physical characteristics prescribing the temporal variation of the temperature can be different for different pixels, but they do not significantly change with time, i.e. they are temporal invariants. Under these conditions, thermal images acquired in sufficiently close and narrow intervals of wavelength are mutually predictable.

Two observation fields obtained under different camera viewing angles θ_1 and θ_2 are generally not mutually predictable. In fact, the temporal changes in the camera viewing angle may influence the measurements in a spatially non-uniform way, according to the objects' BRDFs. Hence, the temporal invariants that are learnable by multitemporal remote sensing may be *different* for the two observation fields if the scene contains objects that emit different amounts of energy in θ_1 and θ_2. Such objects do not comply with the space-invariant prediction concept and a prediction error is inevitable. The problem can be reduced by correcting the data first. The correction accuracy depends on knowledge of the viewing geometries and the objects' BRDFs.

From (14) and the process of deriving (5), it can be inferred that using multispectral data for predicting images *instead of the previously collected multitemporal data* is incorrect, without additional assumptions. Formally, introducing an additional band does add one more equation per basis time moment into system

(9), from which the temporal invariants should be resolved. However, some invariants (e.g. functions of thermal inertia or its temporal evolution characteristics) cannot be learned from multispectral data, whereas they *can* be learned from multitemporal data. Therefore, without assuming a low-parametric model of emissivity the additional equations can only increase the accuracy of the measurement data in the multitemporal system (9). However, they do not influence the system's invertibility. The band plurality, however, has been shown by Schaum and Stocker (2003) to greatly improve the robustness of prediction with respect to misregistration of reflective domain images across time. Yet another way of using multi- and hyperspectral sensing to increase image prediction accuracy is important. One can explore the spatial distribution of the external factors within the scene. For instance, using a three-channel ratioing technique (Gao *et al.* 1993) permits segmenting the scene into several regions, each of which has its own spatial model of water vapor concentration in the atmosphere. Then the prediction operator can be found separately for each of these regions. Developing this idea should be the subject of future work.

3.2 *Selection of basis images*

The basis times are determined in advance, at the stage of training the detection system. This sequence is split into *Selection* subsequence (SEL) and a *Test* subsequence (TST). The simplest feasible approach to selecting the vector of basis times is sketched below:

(i) choose an initial number of basis times P;
(ii) for a chosen P, select basis time points from SEL to minimize a prediction error function over TST adjusted for the number of parameters in the model;
(iii) if the prediction error does not decrease, then exit search;
(iv) increment P and goto step (ii).

The computation can be simplified if we take into account certain physical reasoning to choose an initial number of basis times, as exemplified below.

3.2.1 Number of basis images. Using an asymptotic representation of surface temperature in the form (10), it can be shown (Appendix A) that if the diversity of the thermo-physical properties of the surfaces in the scene is not limited, then

- any thermal image can be represented by a linear combination of at least eight basis images.

The number of *linearly independent* terms, m, in the right-hand side of (10), or equivalently (A9), is the number of the principal independent factors forming the surface temperature:

- downward solar irradiation;
- radiative heat transfer between the surface and the atmosphere;
- sensible heat flow;
- latent heat transfer;
- conductive heat flow from/to the bottom of diurnally active layer of soil.

Thus, in a general case, when all these factors are significant, $m=5$. Iterating equation (10), as exemplified in Appendix A, we obtain an approximate solution given by

$$\Delta T(s,t)=\sum_{k=1}^{M}\beta_k(s)\mathcal{Y}_k(t), \tag{16}$$

where the terms $\beta_k(s)$ and $\mathcal{Y}_k(t)$ are temporal and spatial invariants, respectively. With one–two iterations, the number of invariants M may be 2–3 times greater than m, leading to a lower bound: $M \geqslant 10 \sim 15$.

Furthermore, the intensity value of a TIR channel image pixel can be described by the following formula:

$$W_T=\int_{\Lambda}\varpi(\lambda)\varepsilon_\lambda\mathcal{B}(T,\lambda)\mathrm{d}\lambda+\int_{\Lambda}\varpi(\lambda)(1-\varepsilon_\lambda)\mathcal{B}(T_e,\lambda)\mathrm{d}\lambda, \tag{17}$$

where ϖ (λ) is the camera filter function; ε_λ is the emissivity; T_e is the effective temperature of the environment (in particular, of the sky), the radiation of which is reflected by the scene's objects; and Λ denotes the channel. Planck's function is nearly linear in temperature when $|\Delta T| \leqslant 20\,\mathrm{K}$. Therefore, linearizing Planck's function to convert the temperature model (16) to an analogous linear model for radiance results in the number of basis times M increasing by two, when ε_λ negligibly changes with time; or by three, when small temporal changes of emissivity are possible. Furthermore, if we allow a scene to have a relief, then at least two terms should be added to (10). Indeed, the downward solar irradiance is multiplied by the factor Φ that depends on the solar zenith and azimuth angles, Z and φ_{sol}; and on the local topographic slope and aspect, S and φ, for example as suggested by Teillet *et al.* (1982):

$$\Phi=\cos Z(t)\cos S+\sin Z(t)\sin S\cos(\varphi_{\mathrm{sol}}(t)-\varphi). \tag{18}$$

Thus, *under linear prediction conditions* we can bound from below the number of basis images: $M \geqslant 15 \sim 20$. This initial estimate should be refined as suggested above in §3.2.

3.2.2 Selecting basis time moments. Some brief practical recommendations for choosing the basis times $\boldsymbol{\tau}=[\tau_1, \ldots, \tau_P]$ are as follows. First, the basis time moments are chosen within the SEL to be the moments corresponding to the maximum and minimum of each one of the heating/cooling factors listed in the previous section. The rest of the basis time points are placed on the time axis equally spaced. Next, the basis time vector $\boldsymbol{\tau}_0$ selected in this way is used as the starting value for an iterative algorithm to minimize the prediction performance during the Test period. In Appendix B, we derive a simple selection algorithm. This algorithm is more suitable for the applications in which at the training stage the training scene can be surveyed *frequently*, e.g. every 5–10 minutes.

When the scene visits are spaced by hours or more, as under polar orbiting satellite sensing, one may choose to use additive stepwise regression. At each step of this method, basis times are added to or removed from the prediction model to end up with a maximal number of statistically significant terms in the model. The coefficients whose significance values are less than a threshold are zeroed. The problem of optimal selection of basis images in a general case needs further study.

3.3 *Indicators of prediction*

The principal sub-problems in our approach include

- choosing the form $\widetilde{\mathcal{H}}$ of the parametric model (6);

- selecting the basis time vector $t_1, \ldots, t_P$; and
- determining the indicators of prediction.

These tasks depend on each other. Therefore, they should be solved simultaneously in the course of a recursive procedure. We leave this as a challenge for future fundamental and experimental work. Presently, the issue of finding indicators is better understood for the application 'monitoring the status of a fixed, multitemporally surveyable area'. For this application the basis images and the detection time image can be registered or georeferenced. Therefore, the set of indicators can be the entire or randomized set of image pixel locations, except for known or detected anomalies, i.e. outliers with respect to the prediction model. Another approach is to search for pseudo-invariant objects in the scene (Furby and Campbell 2001, Schott *et al.* 1988).

The unknown true operator of prediction is not linear in the general case. Nevertheless, if we insist on a linear operator, then we actually make an approximation of the true operator at the expense of adding more basis times. Then the indicators of prediction that must be identified in the inspection image may become too numerous. To this end, non-linear models with fewer free parameters may be of greater practical value.

4. Experimental validation and examples of applications

The experimental work described below has several objectives stated in §1.

4.1 *Proof-of-concept experiment using a fixed camera*

If the basis images and their number are selected properly, then the prediction law (5) should work well *for all* possible objects or materials. Therefore, a test site should include a variety of surfaces that are substantially different in their thermo-physical properties. Our site was constructed to include the following:

- a natural landscape with mixed grasses and shrubs;
- several types of bare soil (sea-sand, clay soil, organic matter);
- several types of stones (natural, man-made, matte, polished);
- a water reservoir;
- objects made of various metals (e.g. steel, aluminium), wood, paper and paperboard, asphalt and concrete, glass, foamed plastic, and synthetic polyamide materials;
- other two-dimensional and three-dimensional objects of various shapes.

The measurements were conducted in October 2002 in the vicinity of Tel-Aviv, Israel, with a thermal camera (Agema) equipped with five filters in $3 \leqslant \lambda \leqslant 5\,\mu\text{m}$. The camera was not radiometrically calibrated but was maintained in a fixed position at a height≈4m and a viewing angle≈40°.

Two oblique image sequences were acquired with an eight-day interval between them. The first sequence was used for selecting basis images and the second one for testing the model. In each sequence, the images were taken every 10 min for 50 hours. During the eight-day interval between the sequences, some objects were moved from one place to another. The experiment was carried out under relatively stable weather conditions, with the air temperature varying between 22 and 33°C. However, partial cloudiness (0–4 oktas) and gusty wind ($0\text{–}12\,\text{m}\,\text{s}^{-1}$) caused frequent and rapid changes in the object surface temperatures. The temperature differences for different pixels (for a given frame), reached 40°C (also for the images of the test sequence). The temporal differences for the same pixels reached 40–50°C. Since the camera had uncontrollable *spatial non-uniformity* of the response function of the detectors, the images were corrected, approximately each half-hour, using two uniformly heated blackbody plates with different temperatures. In this experiment, the search for basis images was conducted only among those images that were corrected for spatial non-uniformity (87 images). The prediction was done by using a linear operator of 20 basis images, determined by the algorithm relying on the high frequency of surveying (see §3.2.2 and Appendix B).

4.1.1 Goodness of fit. Figure 1 shows as mosaics the examples of inspection images, the corresponding predicted reference images, and the prediction residual images at two different times of the day. The relatively bright areas in the residual images are pixel-outliers with respect to the prediction model (5). They represent those objects that have moved or changed during the interval between the acquisition of basis images and the inspection image (such as the lid of the water container or leaves of shrubs or grass), or the pixels that do not conform to the prediction model. Figure 2 plots the relative prediction error (i.e. the absolute error divided by the range of the data values) for the two images. The background r.m.s. prediction error amounted to 1.7% for the night-time image and 2.3% for the daytime image. With temperature differences of 40–50°C, this percentage is readily calibrated to the brightness temperature prediction accuracy of 0.7–1.5°C.

4.1.2 Anomaly detection. Using basis images that do not contain anomalies, one can construct the reference image *that corresponds to the environment and the survey*

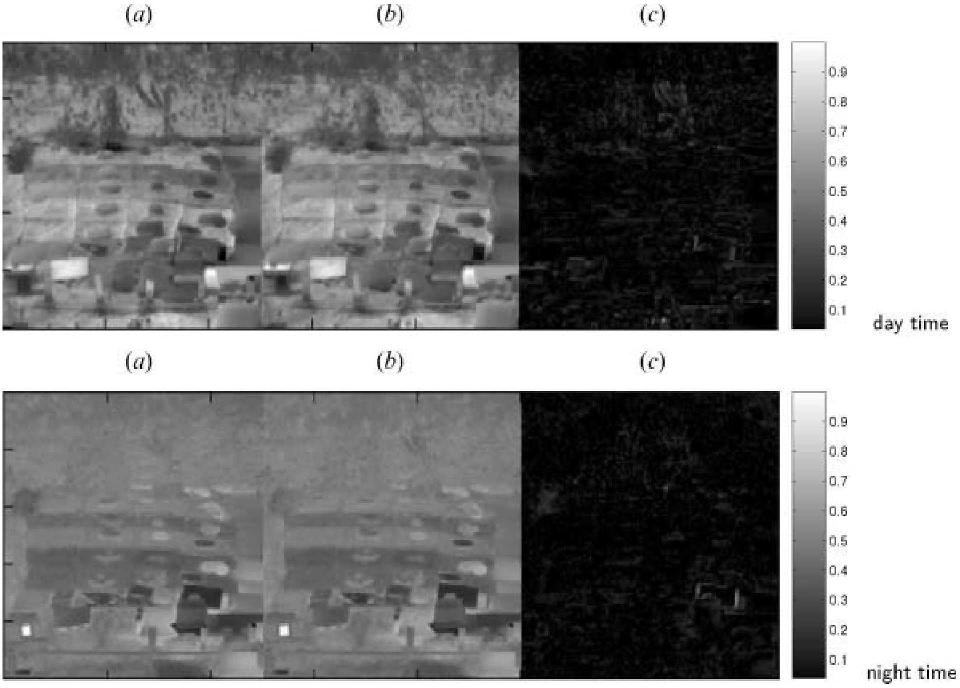

Figure 1. (*a*) Observed TIR intensity images; (*b*) predicted reference images; (*c*) prediction residual images. Prediction range – 8 days. Intensity values are linearly scaled to the range [0, 1].

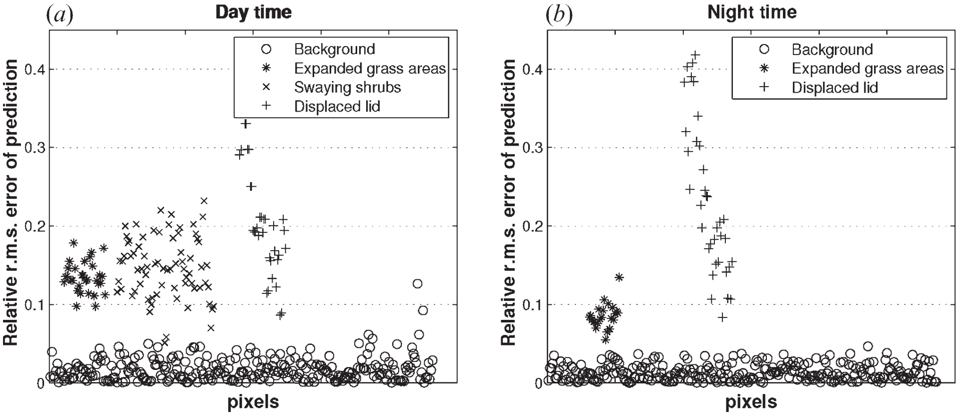

Figure 2. Scatter plot of the relative prediction residual for the unchanged (background) and moved/changed pixels at day time (*a*) and night time (*b*).

conditions of the inspection image currently observed. Suppose, during acquisition of the test sequence a human enters the site. Figure 3(*a*) shows a day-time inspection image with a barely distinguishable human partly occluded by the surrounding background objects. However, after subtracting the reference image predicted by

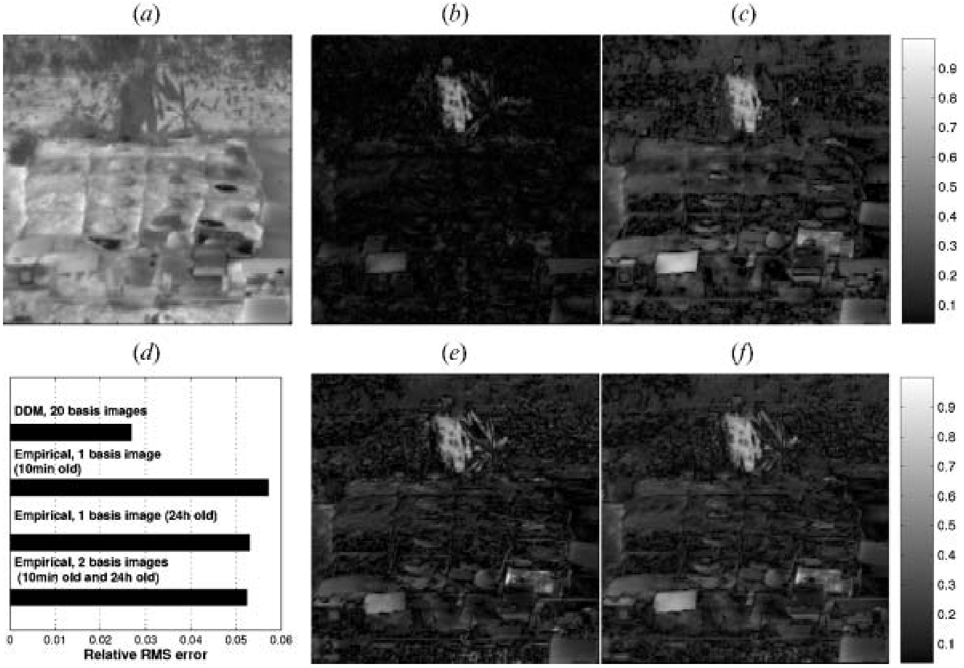

Figure 3. Anomaly detection. (*a*) IR image W_{t_v} with a man indistinguishable from the background objects. (*b*) DDM-based prediction residual $\left|W_{t_v}-\widetilde{\mathcal{H}}[W_{t_1}, \ldots, W_{t_{20}}; \gamma]\right|$. (*c*) Empirical prediction residual $\left|W_{t_v}-\left(aW_{(t_v-10\mathrm{min})}+b\right)\right|$. (*e*) Empirical prediction residual $\left|W_{t_v}-\left(aW_{(t_v-24h)}+b\right)\right|$. (*f*) Empirical prediction residual $\left|W_{t_v}-\left(aW_{(t_v-10\mathrm{min})}+bW_{(t_v-24h)}+c\right)\right|$. (*d*) Relative r.m.s. error of background prediction. Intensity values are linearly scaled to the range [0, 1].

(5), the scene anomalies (the human and the moving leaves) can be clearly seen (figure 3*b*).

A simpler alternative to our method could be a bi-date normalization, which is equivalent to using as a basis image an image $W_{t_{\text{old}}}$ acquired under apparently 'the same' weather conditions, e.g. 10 minutes or exactly 24 hours before t_v (i.e. the time the inspection image was taken). Then the background is predicted by a simple linear relation

$$W_{t_v} = \mathrm{a} W_{t_{\text{old}}} + \mathrm{b}. \tag{19}$$

The resulting residual images are shown in Figure 3(*c*,*e*). As this figure suggests, not only is the background prediction less accurate in terms of the r.m.s. error (see figure 3*d*), but it also leads to *higher spatial clutter* in the residual images. Not much of an improvement was achieved when we used both previous images as the basis images. Figure 3(*f*) displays the corresponding residual images.

4.1.3 Intruder detection by DDM. Detecting a scene anomaly is not equivalent to identifying a class 'human-intruder'. In deciding whether an anomaly detected is or is not a human, we used the multispectral intensity vectors. Using a database of human images taken at the basis time weather conditions, we estimated the Gaussian mixture parameters of class 'human' by the method proposed by Koltunov and Ben-Dor (2004). The parameters of class background were computed using the basis images. Furthermore, using the estimated prediction operator, we recovered the dynamics of these parameters in the way described in §2.4, thus obtaining the current weather versions of the parameters. Next, only the anomalous pixels of the inspection frames (figure 4*A*) were classified. This was done by maximization of the class posterior probabilities. In this way, the classifier can ignore irrelevant anomalies due to a natural background motion (swaying grass or bush leaves, in our case). The resulting classification image examples are shown in Figure 4(*c*).

The complexity of the recognition problem in our example can be explained not only by a significant background motion violating the method assumption. In addition, at the time (more precisely, under the specific weather conditions) of acquiring the inspection frames, the class 'human' *was objectively inseparable* from many background classes, e.g. vegetation, in multispectral short-wave TIR data. Therefore, even if we used the inspection image for training, as is conventionally done (and not the past images), we would obtain false alarms all over the image. This can be seen in Figure 4(*b*), which displays in dark tones all those pixels that simply must be classified as 'human' in the scene, because of the class distribution overlap at the recognition time.

This example shows that object classification by identifying an anomaly already detected is more reliable and effective than classification using multispectral features only. It can be especially useful for controlling the status of a known, previously accessible territory.

4.2 *Sub-pixel hot anomaly detection in MODIS imagery*

In this section we demonstrate the applicability of the multitemporal prediction approach presented in this paper to space-borne multitemporal datasets at a coarse resolution. The thematic example-task is detection of subpixel-scale thermal anomalies, such as fires.

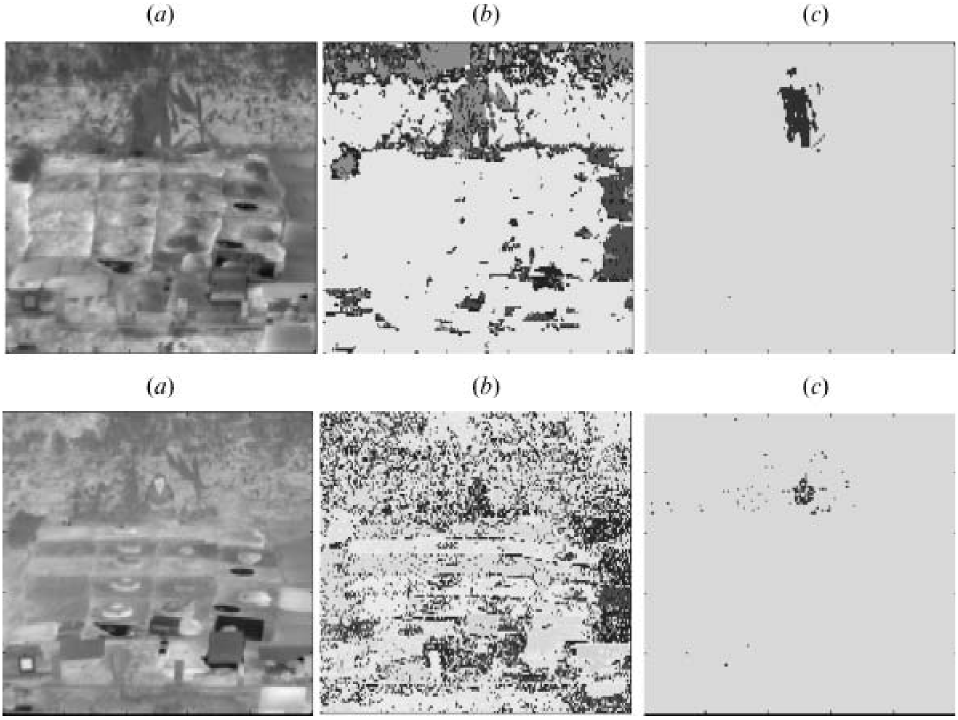

Figure 4. Examples of intruder detection by DDM. Top row: detection of standing intruder. Bottom row: detection of sitting intruder. Columns: (*a*) Inspection images with intruders; (*b*) Class 'human' (dark tones), identified using pixels' multispectral TIR features *only*. Numerous false positives. (*c*) Scene anomaly identified as 'human', using DDM.

The scene $\mathbb{S}$ used in this experiment is located in Northern California and occupies over 150 000 km^2. Figure 5 displays the natural colour image of the scene. The scene represents a typical western U.S. landscape, with a mosaic of forests, shrubs, grasslands, urban structures, and water bodies. The MODIS instruments on the Terra

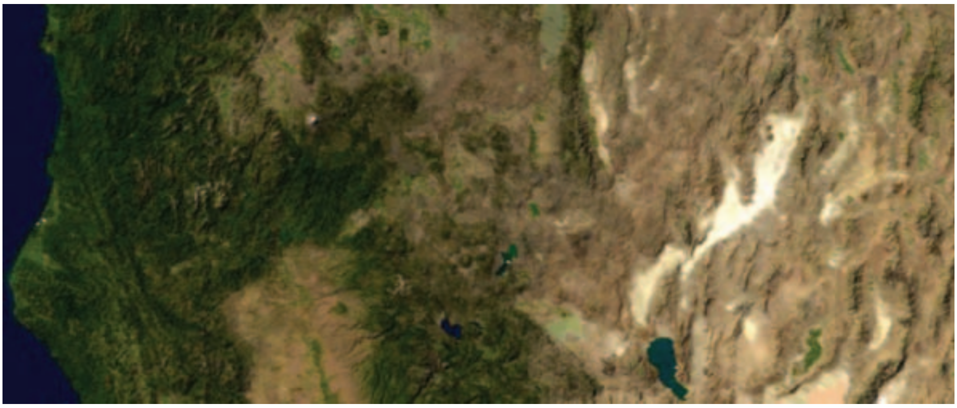

Figure 5. Natural colour image of the test scene located in Northern California. The scene is confined within a bounding rectangle with the corner coordinates: $-124°36'$ W, $42°2'24''$ N (upper-left) and $-117°12'$ W, $39°36'$ N (lower-right). The image size is 250×600 pixels at the ground resolution of 1 km. The MODIS bands used for [Red, Green, Blue] composition: [1, 4, 3].

and Aqua satellites overpass scene $\mathbb{S}$ three to seven times each 24-hour period, collecting the data under different solar-view geometries and a nominal ground resolution of 1 km at nadir. Two uncalibrated thermal bands: band 22 (~4 μm) and band 31 (~11 μm) were used in our experiment. We denote these bands by R_4 and R_{11}, respectively, and the corresponding brightness temperatures – by T_4 and T_{11}.

The multitemporal sequence consisted of 360 image frames acquired for eight months (March 1 – November 1) of year 2004. The first 278 images were used for training the anomaly detection algorithm. To assess the detection performance, we selected 63 test image frames acquired 4 to 52 days *after* the images that we used for training. These were 32 evening images and 31 morning ones. All images in the sequence were georeferenced to a common projected coordinate system with nearest-neighbour interpolation. The resulting ground pixel size was 1 km. Cloud-contaminated observations were determined using the MODIS cloud mask data (MOD35 and MYD35 products). The pixels flagged 'not-determined', 'confident cloud', or 'probably cloud' in these products, were marked as having missing values and excluded from processing.

The incremental selection algorithm (§3.2) with stepwise regression resulted in a basis time vector containing 32 basis times. They were selected to minimize the prediction error for band R_4 only. First 259 images formed the *Selection* subsequence and the next 19 images – the *Test* subsequence.

As can be seen from (5), only those pixels can be predicted that were available at all basis times. However, primarily due to cloud cover, many pixels have missed observations in one or more basis images. To overcome this problem, a straightforward procedure was implemented in this paper. Given a basis time vector $\boldsymbol{\tau}_{\mathrm{b}}$ of length P, at the training stage we constructed P additional vectors of length $P-1$ each, by excluding one basis time from $\boldsymbol{\tau}_{\mathrm{b}}$. At the detection time, the prediction coefficients were computed for each vector by stepwise regression, using as indicators at most $2 \cdot 10^4$ pixel locations. These locations were chosen at random out of those that were not missing in the current basis images or in the inspection image. Finally, for each pixel a predictor with the smallest r.m.s. error (adjusted for the number of parameters in the model) was chosen among the predictors for which the pixel did not have missing values at the basis times.

4.2.1 Comparative evaluation of DDM performance.

To compare the DDM to other detection methods, we simulated 10 groups of 100 idealized fires each. All fires were assumed to have the same temperature – 600 K and placed in the centres of random pixels that were located on land and did not have missing observations. The area of each fire in the k-th group was set to $k \cdot 100\,\mathrm{m}^2$. The emitted radiance (under the blackbody assumption) was computed by Plank's law for each fire to alter the actual pixel background radiance, according to the anomaly area proportion. Finally, the radiances were converted back to the original digital counts. The rest of the pixels, excluding known actual fires or pixels having missing values, were considered background pixels. In general, using the *simulated* fire data may bias the *absolute* characteristics of the method performance. However, the use of simulated data is necessary and acceptable for *comparing* different methods or their components. Furthermore, when actual anomalies occur in the image with low probability, modeling is the only practical way to obtain a large and representative test set over the limited time range and scene size.

Two other techniques compared to the DDM are described as follows:

MOD: The contextual method implemented in the standard MODIS fire product (MOD14) logically combines a series of contextual and absolute-threshold tests, preceded and followed by additional algorithms to filter out false positives. The contextual tests are applied to the swath data and use the valid neighbouring pixels in windows of variable size to estimate two background statistics: the mean and the mean absolute deviation. Details can be found in Giglio *et al.* (2003). In our comparison we used the combination of two contextual tests comprising the central part of the method:

$$T_4 > \mu_{T_4} + v_4 \delta_4 \tag{20}$$

$$\Delta T > \mu_{\Delta T} + v \delta_{\Delta T}, \tag{21}$$

where $\Delta T \overset{\mathrm{def}}{=} T_4 - T_{11}$; μ_{T_4}, $\mu_{\Delta T}$ denote the mean values of T_4 and ΔT, respectively; and δ_4, $\delta_{\Delta T}$ denote their mean absolute deviations. The thresholds v and v_4 that can be calibrated to detection confidence control the test sensitivity. The optimized selection of valid neighbours and the window size was implemented as described by Giglio *et al.* (2003).

BiDate: A two-date method that predicts the detection image background by linearly transforming a previous image taken at a time t_{prev} as close as possible to the same time of the day as the detection image, e.g.: $\widehat{R_4(t)} = aR_4(t_{\mathrm{prev}}) + b$, which is identical to (19) and was used by Schott *et al.* (1988) and by Furby and Campbell (2001).

All methods but BiDate were applied to 63 detection times. To avoid inventing a way to deal with missing observations that would be appropriate for BiDate, we applied this method to 10 cloud-free evening images. Our method, DDM, was applied to R_4 and $\Delta R \overset{\mathrm{def}}{=} R_4 - R_{11}$, which provided the equivalent for the brightness temperature tests, (20) and (21), by Giglio *et al.* (2003). Finally, for all methods we logically combined the two tests to obtain the higher performance fire detectors. Note that all methods combined the single-band tests in the identical way: 'a pixel is flagged fire if and only if both tests return true'. Therefore, the discrepancy between the detection results were solely due to differences between methods.

Figure 6 plots the false positive rates given the fixed true detection rates. As is evident from this figure, the performance of DDM is similar to that of the state-of-the-art contextual method, MOD, during evening hours. In morning hours, DDM significantly outperformed MOD. Averaged over the test period false alarm rates are shown in figure 7. Overall, DDM committed about 13–20% fewer false positives than MOD in the evening and 1.5 to 3.5 times fewer false positives than MOD in morning hours. The simple bi-date technique showed considerably lower performance than the other two methods.

5. Discussion and conclusion

The abstractness and mathematical formality of presenting the joint prediction approach, in our opinion, emphasize its generality and potential applicability to various physical phenomena and processes. For instance, besides thermal anomaly detection and near-real-time classification, the space-invariant prediction approach could be applied to predicting geomagnetic field variations, forecasting weather, or taking into account motion and shape variability of 3D objects to be recognized. Focusing on thermal infrared, we have also illustrated the predictability phenomenon from thermo-physical perspective (§2.3 and Appendix A) by deriving the simplest thermal imagery predictor expressed in terms of physically meaningful quantities.

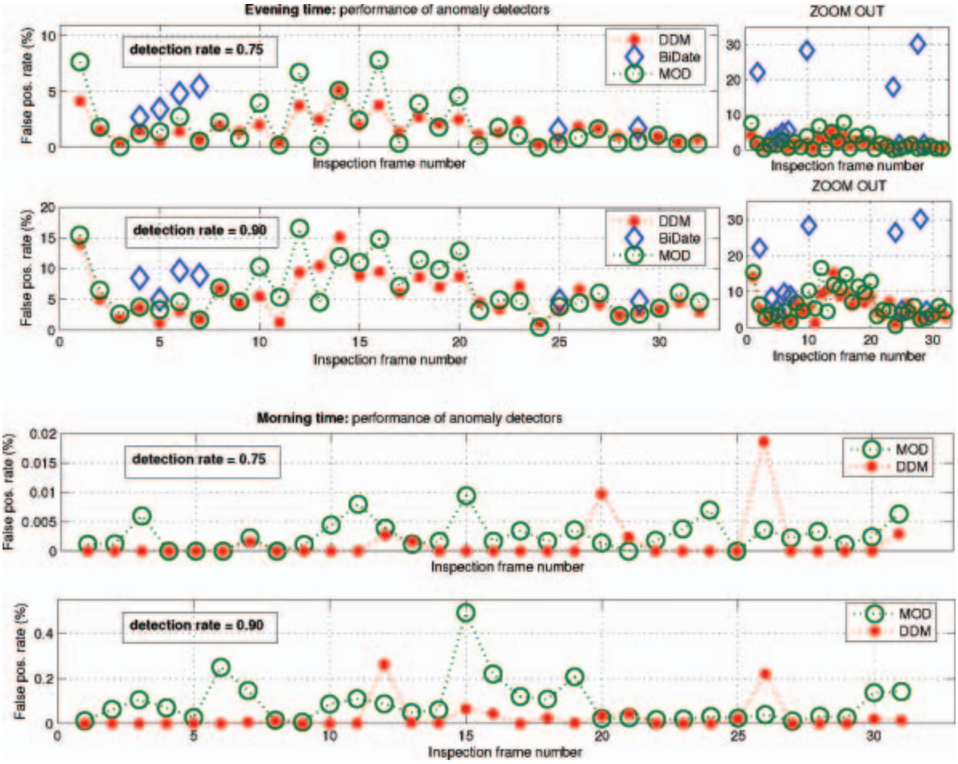

Figure 6. MODIS dataset over the Northern California scene. False positive rates for the anomaly detection algorithms (see §4.2.1) combining the R_4-test with the ΔR-test.

Both the general form of the prediction law (§2.1) and the explicit TIR example-model (A15) have been derived for a single-channel case. Thus, in contrast to previous work (Schaum and Stocker 1997), our prediction approach applies equally under hyperspectral, multispectral, *or even single-channel* surveillance.

The model of the observation process does not have to be known or assumed linear. As mentioned in §2, the key assumption is that the observation process can be represented in the form, in which any parameter does not simultaneously depend on time and on spatial location. Consequently, as long as the sensor response function is spatially uniform it can be viewed as an additional space-invariant factor of external influence. That is why a correction for spatial non-uniformity (e.g. destriping) improves the prediction accuracy, whereas radiometric calibration or any other space-invariant preprocessing manipulations with the raw data, whether they are linear or not, are absolutely unnecessary for using our approach. If the camera viewing angle varies with time, the observation process models corresponding to different viewing angles are different. Therefore, the accuracy of image prediction and classification may considerably deteriorate for highly non-Lambertian objects. Overcoming this problem without additional prior knowledge would require that basis images be taken from multiple directions. The MODIS images used as the basis images in our example were actually taken from different angles, which was helpful for prediction.

Theoretically, the absolute range of prediction is not an issue because Δt does not appear in the final formulas anywhere. What does matter, however, is how long the

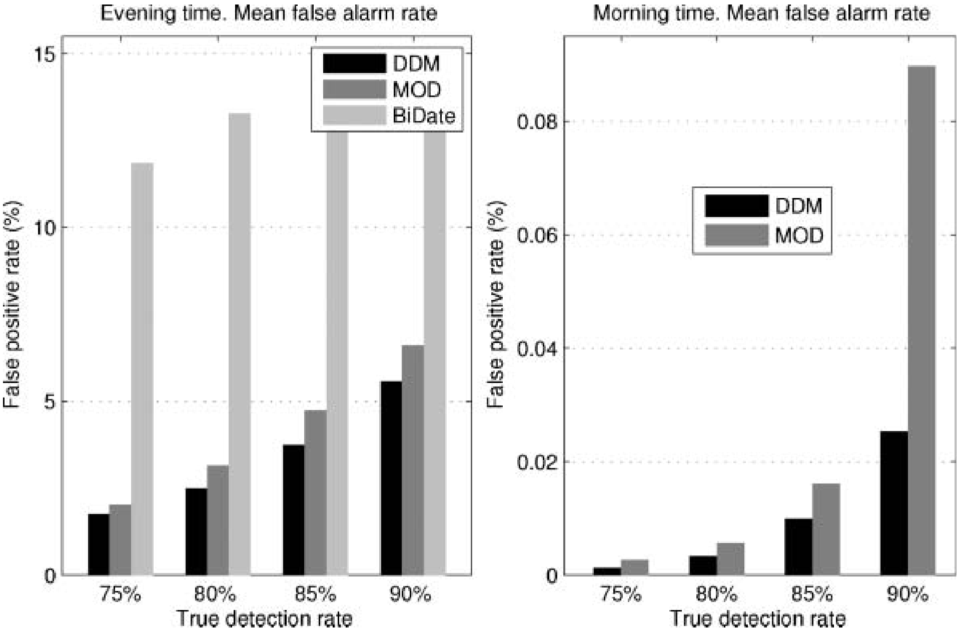

Figure 7. MODIS dataset over the Northern California scene. False positive rates averaged over 48 days (Sept. 4–Nov. 1, 2004) for the anomaly detection algorithms combining the R_4-test with the ΔR-test.

temporal invariants that have been learned from a given set of basis images will be the only spatially non-uniform factors influencing the measurements. For the case of TIR surveillance, this may depend on the variability of local weather conditions within a given period. Besides that, under real conditions after extended surveillance periods the object properties at most pixel locations may be found to have changed since the basis times. Therefore the same set of basis images will not remain good indefinitely long for all pixels.

The implicit parameterization mechanism explained in §2.2 leads to an important conclusion. The restrictions on the spatial behaviour of the external factors and the temporal behaviour of the objects' properties are removed at the expense of introducing additional basis fields into the prediction model. Conversely, if prediction is to be made for a limited variability (spatial or temporal) of external conditions or for a limited set of objects, then the number of basis fields that is sufficient for prediction can be reduced. Therefore, simple empirical approaches, e.g. bi-date image normalization, can work well for *some* objects and under *some* external conditions. To this end, in our preliminary experiment it was important that the DDM was tested for *diverse* objects and materials in a scene with complex three-dimensional structure and shading effects. Likewise, the MODIS data were characterized by a high spatial and temporal variability of external conditions. The prediction models used in these experiment were not the best possible: not all temporal invariants were accurately resolved, owing to the technical limitations in the selection of basis images and perhaps, due to the linearity of the model. As a result, some objects were predicted less accurately. In the tower-based experiment, these included a polished stone and some objects' boundaries. Also, in the MODIS data experiment, we observed spatially contiguous regions of low prediction accuracy that could be attributed to local changes (relative to the basis times) in vegetation conditions, or to local sunlight or cloud influences. Nevertheless, the underfitted prediction models provided considerable improvements over their counterparts.

The accuracy of the background prediction may be lower when the scene is severely clouded at the detection time, and therefore a sufficiently complex prediction model cannot be built for the lack of indicators.

5.1 *The prospect of applying DDM for exploration of new territories*

The most challenging potential application of the DDM is classification of a new territory. For this application, the multitemporally surveyed training scene and the recognition scene (i.e. the new territory) are geographically different areas. In this case the DDM requires that the recognition scene contain only those classes that are represented in the training scene, which may at first seem to be always unrealistic due to enormous diversity of the Earth's surface. The assumption becomes indeed unrealistic if the notion of the training scene is *mistaken* as a spatially contiguous, geographically compact area whose size is so small that the external conditions remain the same for the time of surveying the area. The actual meaning of the notion 'training scene' used by the DDM is substantially wider, as specified below:

- Training scene does not have to occupy a single, geographically compact area: it can and should comprise a plurality of geographically different sites, further termed *sub-scenes*; the distance between any two sub-scenes is limited only by the Earth's equator length, e.g. they may be situated in different continents, along the satellite overpass or aircraft flight lines.
- Though each sub-scene comprising the training scene must be surveyed multitemporally, they are not required to be surveyed simultaneously or under the same external factors.
- The only objects that must be present in each sub-site and during each survey are the prediction indicators. This requirement is met easily if the sub-scenes are partly overlapping.

Multitemporal data from different sub-scenes can be merged by sequentially appending the data of a new training sub-scene $\mathbb{S}_{\text{new}}$ to the data of an already existing training sub-scene $\mathbb{S}_{\text{old}}$. The information about thematic categories of $\mathbb{S}_{\text{new}}$ can be extracted prior to the multitemporal survey, by the traditional means of training data acquisition.

Overall, it can be concluded from the above discussion that applicability of Dynamic Detection Models under the most difficult circumstances is a wide and interesting topic for future studies and experiments. In many circumstances, it is better or sufficient to apply the traditional supervised statistical classification with the training data sampled from the inspection image. Below are examples of such circumstances:

- past images are too few or unavailable;
- prior information about the recognition site is enough to form the correct list of classes and obtain training data that are representative of the thematic classes;
- near-real-time performance is not required.

In addition, only if the prediction operator is determined with a high accuracy, can the past training data provide the equivalent for the training data selected from the current image.

5.2 *Summary and conclusion*

The contributions of the present paper include the following developments:

(i) a generic law of predictability of physical observations is derived;
(ii) underlying assumptions and several practical aspects of multitemporal prediction are discussed for the first time, including selection and the number of past images, indicators of prediction, underlying assumptions;
(iii) lower bounds on the number of past thermal infrared images are obtained;
(iv) algorithms for automatic selection of the basis images and number thereof are described;
(v) a DDM using a Gaussian mixture based classifier is proposed;
(vi) the proof-of-concept tests are performed;
(vii) potential applications are suggested and exemplified.

In conclusion, we emphasize that the multitemporal predictability phenomenon is a general fact rather than representing occasional empirical successes and can be productive for solving a range of thematic tasks. Yet, at present, many of the issues raised in this paper need further research. Furthermore, a more complete list of theoretical and practical problems, as well as classification and detection tasks that can be solved using the space-invariant prediction model, should be formulated. We hope that future studies will determine how to best utilize this objectively existing phenomenon to make good use of data from repeatable remote sensing missions.

Acknowledgments

This research was supported by the German-Israeli Scientific Foundation (GIF), under grant I-547-177.02/97, and by NASA, under subcontract on grant #NNG04GK34G. We thank the anonymous reviewers whose comments substantially improved the manuscript. We also thank George Scheer and Lawrence Ross for systems administration and other computation support that enabled us to analyse MODIS data.

References

CARLOTTO, M.J., 2000, Nonlinear background estimation and change detection for wide-area search. *Optical Engineering*, **39**, pp. 1223–1229.

DASH, P., GÖTTSCHE, F.-M., OLESEN, F.-S. and FISCHER, H., 2002, Land surface temperature and emissivity estimation from passive data: theory and practice – current trends. *International Journal of Remote Sensing*, **23**, pp. 2563–2594.

DU, Y., TEILLET, P.M. and CIHLAR, J., 2002, Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. *Remote Sensing of Environment*, **82**, pp. 123–134.

FURBY, S.L. and CAMPBELL, N.A., 2001, Calibrating images from different dates to 'like-value' digital counts. *Remote Sensing of Environment*, **77**, pp. 186–196.

GAO, B.-C., HEIDEBRECHT, K.B. and GOETZ, F.H., 1993, Derivation of scaled surface reflectances from AVIRIS data. *Remote Sensing of Environment*, **44**, pp. 165–178.

GIGLIO, L., DESCLOITRES, J., JUSTICE, C.O. and KAUFMAN, Y.J., 2003, An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*, **87**, pp. 273–282.

Jensen, J.R., 1983, Urban/suburban land use analysis. In *Manual of Remote Sensing*, R.N. Colwell (Ed.) volume 2 (Falls Church: American Society of Photogrammetry), pp. 1571–1666.

Koltunov, A. and Ben-Dor, E., 2004, Mixture density separation as a tool for high-quality interpretation of multi-source remote sensing data and related issues. *International Journal of Remote Sensing*, **25**, pp. 3275–3299.

McLachlan, G.J., 1992, *Discriminant Analysis and Statistical Pattern Recognition* (New York: John Wiley).

McLachlan, G.J. and Peel, D., 2001, *Finite Mixture Models* (New York: John Wiley).

Milman, A.S., 1999, *Mathematical Principles of Remote Sensing: Making Inferences from Noisy Data* (Chelsea, MI: Sleeping Bear Press).

Price, J.C., 1989, Quantitative aspects of remote sensing in the thermal infrared. In *Theory and Applications of Optical Remote Sensing*, G. Asrar (Ed.) (New York: Wiley), pp. 578–627.

Schaum, A. and Stocker, A., 1997, Subclutter target detection using sequences of thermal infrared multispectral imagery. In *Algorithms for Multispectral and Hyperspectral Imagery*, S.S. Shen and M.R. Descour (Eds), volume 3071 of Proceedings of SPIE, pp. 12–22.

Schaum, A. and Stocker, A., 2003, Hyperspectral multipass mapping for target detection. In *Algorithms for Multispectral and Hyperspectral Imagery*, S.S. Shen and P.E. Lewis (Eds), volume 5093 of *Proceedings of SPIE*, pp. 1–8.

Schott, J.R., Salvaggio, C. and Volchok, W.J., 1988, Radiometric scene normalization using pseudoinvariant features. *Remote Sensing of Environment*, **26**, pp. 1–16.

Susskind, J., Rosenfield, J. and Reuter, D., 1983, An accurate radiative transfer model for use in the direct physical inversion of HIRS2 and MSU temperature sounding data. *Journal of Geophysical Research*, **88**, pp. 8550–8568.

Teillet, P.M., Guindon, B. and Goodenough, D.G., 1982, On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, **8**, pp. 1537–1540.

Tikhonov, A.N. and Samarsky, A.A., 1963, *Equations of Mathematical Physics*, Chapter III (New York: Pergamon Press).

Appendix A: Thermo-physical insight into the image predictability phenomenon

This appendix illustrates the material from §2 from a thermo-physical perspective by considering a specific case of prediction: predicting a radiometrically calibrated thermal infrared image. This is done under the conditions, in which the explicit formulas for the prediction operator will be least complicated (see also §2.3). The scope of the notation and the symbols we use (see table 1) is limited to the present appendix only.

Consider the following circumstances:

1. The scene is planar and horizontal.
2. All scene objects have a large interior area, and thus the heat flow at the interior points is nearly vertical.
3. The observation field W consists of *only* the interior points of the objects.
4. Thermo-physical parameters, albedo, and the emissivity of different objects vary but are time-invariant.
5. The external influence (from the sun and the atmosphere) is space-invariant.
6. Narrow-band long wave TIR surveying: each band Λ has a filter that is uniform in wavelength λ; and the bandwidth $\delta\lambda$ is small enough to assume the emissivities $\varepsilon(\lambda)$ and the Planck function $\mathcal{B}(\lambda, T)$ independent of λ within the band. The scene is surveyed from a low altitude. The sensor is radiometrically calibrated and flat-fielded.

Table 1. Symbols used in Appendix A.

Symbol	**Description**	**Units**
x	depth	m
t	time	s
T	temperature	K
R	net radiation flux density	$\mathrm{J\,m^{-2}\,s^{-1}}$
H	sensible heat flux density	$\mathrm{J\,m^{-2}\,s^{-1}}$
L	latent heat flux density	$\mathrm{J\,m^{-2}\,s^{-1}}$
Q	depth flux density	$\mathrm{J\,m^{-2}\,s^{-1}}$
a^2	thermal diffusivity	$\mathrm{m^{2}\,s^{-1}}$
k	thermal conductivity	$\mathrm{J\,m^{-1}\,K^{-1}\,s^{-1}}$
R_{sun}	downward solar irradiance	$\mathrm{J\,m^{-2}\,s^{-1}}$
R_{atm}	downward atmospheric irradiance	$\mathrm{J\,m^{-2}\,s^{-1}}$
$\mathcal{A}$	surface albedo integrated over $0.4 \leqslant \lambda \leqslant 3\,\mu\mathrm{m}$	dimensionless
ε	surface emissivity integrated over $3 \leqslant \lambda \leqslant 30\,\mu\mathrm{m}$	dimensionless
σ	Stefan–Boltzmann constant	$5.7 \cdot 10^{-8}\,\mathrm{J\,m^{-2}\,K^{-4}\,s^{-1}}$
$\wp$	thermal inertia	$\mathrm{J\,m^{-2}\,K^{-1}\,s^{-1/2}}$
λ	wavelength	µm
$\delta\lambda$	narrow band width	µm
ε_{λ}	surface emissivity at wavelength λ	dimensionless
$\mathcal{B}(T, \lambda)$	Planck's function (spectral radiance)	$\mathrm{J\,m^{-2}\,\mu m^{-1}\,s^{-1}}$
ρ	density	$\mathrm{kg\,m^{-3}}$

(7) Weather/surface conditions: stable anticyclone with no near-surface wind; an arid environment. Thus, the sensible and latent heat transfer components of the surface energy balance are considered negligible; the only heating/cooling factors are solar irradiance, the radiative heat exchange between the scene objects and the atmosphere, and the conductive heat flux at the bottom of the diurnally active layer of soil, referred to as the *depth flux*. The depth flux is assumed constant within the observation period. The atmospheric transmittance approaches one. The weather conditions are not periodic.

Under the assumptions (1)–(5), the depth distribution of temperature $T(x, t)$ for all points of the objects can be described by a one-dimensional heat equation

$$\frac{\partial T(x, t)}{\partial t} = a^2 \frac{\partial^2 T(x, t)}{\partial x^2}, \tag{A1}$$

with two boundary conditions and an initial condition:

$$-k \frac{\partial T(0, t)}{\partial t} = G(t); \quad \text{(on the surface)} \tag{A2}$$

$$k \frac{\partial T(\infty, t)}{\partial t} = Q; \quad \text{(at the bottom of the diurnally active layer)} \tag{A3}$$

$$T(x, t_0) = \phi_0(x); \quad \text{(the initial depth distribution of the temperature)} \tag{A4}$$

where $G(t) = R(t) + H(t) + L(t)$; and R is the net flux of radiation, H is the sensible heat flux, L is the latent heat flux, and Q is the depth flux. The net flux on the surface can be conventionally expressed (Price 1989) as the difference between the absorbed and emitted radiation described by the Stefan–Boltzmann law.

$$R = (1 - \mathcal{A}) R_{\mathrm{sun}} + \varepsilon R_{\mathrm{atm}} - \varepsilon \sigma T^4. \tag{A5}$$

The solution of problem (A1)–(A4) is given (Tikhonov and Samarsky 1963) by the following function:

$$T(x,t)=\frac{Q}{k}x+\frac{1}{\wp\sqrt{\pi}}\int_{t_0}^{t}\frac{Q+G(\eta)}{\sqrt{t-\eta}}\mathrm{d}\eta+\frac{1}{2\sqrt{\pi a^2\Delta t}}\int_0^{\infty}\left[e^{-\frac{(x-\xi)^2}{4a^2\Delta t}}-e^{-\frac{(x+\xi)^2}{4a^2\Delta t}}\right]\left[\phi_0(\xi)-\frac{Q}{k}\xi\right]\mathrm{d}\xi, \tag{A6}$$

where $\wp=k\big/\sqrt{a^2}$ denotes the thermal inertia and $\Delta t=t-t_0$. On the surface ($x=0$), equation (A6) reduces to

$$T(t)=\frac{1}{\wp\sqrt{\pi}}\int_{t_0}^{t}\frac{Q+G(\eta)}{\sqrt{t-\eta}}\mathrm{d}\eta+\frac{1}{\sqrt{\pi a^2\Delta t}}\int_0^{\infty}e^{-\frac{\xi^2}{4a^2\Delta t}}\left[\phi_0(\xi)-\frac{Q}{k}\xi\right]\mathrm{d}\xi. \tag{A7}$$

Here the second term reflects the influence of $\phi_0(x)$, the initial depth-distribution of temperature. It is reasonable to assume that this influence vanishes with time, which allows us to impose the limitation of the initial distribution:

$$\phi_0(x)=T_0+\frac{Q}{k}x, \tag{A8}$$

where T_0 is some constant. The difference between the real, unknown $\phi(x)$ and (A8) soon becomes unimportant for determining the surface temperature by (A7), if we choose t_0 such that $G(t_0)=0$, as in Price (1989), and set $T_0=T(t_0)$. Hence we arrive at the asymptotic representation:

$$T(t)=T_0+\frac{1}{\wp\sqrt{\pi}}\int_{t_0}^{t}\frac{1}{\sqrt{t-\eta}}\left[Q+(1-\mathcal{A})R_{\mathrm{sun}}(\eta)+\varepsilon\left\{R_{\mathrm{atm}}(\eta)-\sigma T^4(\eta)\right\}+H(\eta)+L(\eta)\right]\mathrm{d}\eta. \tag{A9}$$

Under the above assumptions (5)–(7), the spatial-temporal field $W(s,t)$, measured at channel Λ, depends on time through the objects' temperatures T, and at a given location s it can be written:[1]

$$W(T,\lambda)=\varepsilon_\lambda\mathcal{B}(T,\lambda)\delta\lambda+(1-\varepsilon_\lambda)\mathcal{B}(T_e,\lambda)\delta\lambda, \tag{A10}$$

where T_e is the effective temperature of the environment (in particular, of the sky), the radiation of which is reflected by the scene's objects.

Furthermore, with reasonable variability of the temperature $|T-T_0|\leqslant 20\,\mathrm{K}$, we can write

$$\max_t\left|\frac{\Delta T}{T_0}\right|\lessapprox 0.06,\quad \text{where}\quad \Delta T\overset{\mathrm{def}}{=}T-T_0,$$

and therefore approximate the Planck's function by

$$\mathcal{B}(T,\lambda)\approx\mathcal{B}(T_0,\lambda)+\Delta T\cdot\frac{\partial}{\partial T}\mathcal{B}(T_0,\lambda). \tag{A11}$$

In view of assumption (7), the on-surface representation (A9) reduces to the following integral equation:

$$\Delta T=\frac{2Q\sqrt{\Delta t}}{\wp\sqrt{\pi}}+\frac{1-\mathcal{A}}{\wp\sqrt{\pi}}\int_{t_0}^{t}\frac{R_{\mathrm{sun}}(\eta)}{\sqrt{t-\eta}}\mathrm{d}\eta+\frac{\varepsilon}{\wp\sqrt{\pi}}\int_{t_0}^{t}\frac{R_{\mathrm{atm}}(\eta)}{\sqrt{t-\eta}}\mathrm{d}\eta-\frac{\varepsilon}{\wp\sqrt{\pi}}\int_{t_0}^{t}\frac{\sigma T^4(\eta)}{\sqrt{t-\eta}}\mathrm{d}\eta, \tag{A12}$$

---

[1] The channel downwelling atmospheric radiance is approximated by the Planck's function. Clearly, more accurate representations are also possible (Susskind *et al.* 1983).

Note that the last integral in (A12) depends on the object's temperature, and therefore it is intrinsically variable in space. That is why we cannot *directly* derive from this equation a space-invariant prediction model. To overcome the problem, we recognize that (A12) reduces to a Fredholm integral equation *of the second kind* (Milman 1999). Consequently, it can always be solved iteratively. Introducing the 97%-accurate linearization $T^4 = T_0^4 + 4T_0^3 \Delta T$ into the right-hand side of (A12) and performing one iteration of the successive approximation method of solving integral equations, we obtain an approximate solution given by

$$\Delta T = \sum_{i=1}^{6} \mu_i(s) z_i(t), \tag{A13}$$

where

$$\mu_1 = \frac{1-\mathcal{A}}{\wp\sqrt{\pi}}; \quad \mu_2 = \frac{\varepsilon}{\wp\sqrt{\pi}}; \quad \mu_3 = \frac{1}{\wp\sqrt{\pi}} \cdot \left(\frac{Q}{\sigma} - \varepsilon T_0^4\right);$$

$$\mu_4 = \frac{4(1-\mathcal{A})\sigma\varepsilon T_0^3}{\pi\wp^2}; \quad \mu_5 = -\frac{4\sigma\varepsilon^2 T_0^3}{\pi\wp^2}; \quad \mu_6 = -\mu_5 T_0^4;$$

$$z_1 = \int_{t_0}^{t} \frac{R_{\mathrm{sun}}(\eta)\mathrm{d}\eta}{\sqrt{t-\eta}}; \quad z_2 = \int_{t_0}^{t} \frac{R_{\mathrm{atm}}(\eta)\mathrm{d}\eta}{\sqrt{t-\eta}}; \quad z_3 = 2\sigma\sqrt{\Delta t};$$

$$z_4 = \int_{t_0}^{t} \frac{z_1(\eta)\mathrm{d}\eta}{\sqrt{t-\eta}}; \quad z_5 = \int_{t_0}^{t} \frac{z_2(\eta)\mathrm{d}\eta}{\sqrt{t-\eta}}; \quad z_6 = 2\sigma \int_{t_0}^{t} \sqrt{\frac{\eta - t_0}{t-\eta}}\mathrm{d}\eta.$$

Finally, we introduce (A13) into (A11), and then (A11) into (A10), resulting in

$$W(s, t) = \sum_{k=0}^{7} \alpha_k(s) X_k(t), \tag{A14}$$

where $X_k = z_k$, $\alpha_k = \mu_k \cdot \varepsilon_\lambda \dfrac{\partial}{\partial T}\mathcal{B}(T_0, \lambda)\delta\lambda$ (for $1 \le k \le 6$), $X_0 = 1$, $X_7 = \mathcal{B}(T_e, \lambda)$, $\alpha_0 = \varepsilon_\lambda \mathcal{B}(T_0, \lambda)\delta\lambda$, and $\alpha_7 = (1-\varepsilon_\lambda)\delta\lambda$.

Equation (A14) inherits its separated-dependency form from (A13). It is a linear case of (2) and can be used to make predictions in time or in space. To predict W at a new time t_v one should

(a) perform $N \geqslant 8$ surveys at the properly chosen basis time moments $t_1, \ldots, t_P$ to obtain $\boldsymbol{W}_{\mathrm{b}} \overset{\mathrm{def}}{=} (W_{t_1}, \ldots, W_{t_N})$;
(b) determine vector $\boldsymbol{\alpha}$ from the system: $\mathbb{X}^T \boldsymbol{W}_{\mathrm{b}} = \mathbb{X}^T \mathbb{X} \boldsymbol{\alpha}$, where $\mathbb{X}$ is the $8 \times N$ matrix formed by $X_k(t_j)$; and
(c) introduce the solution $\boldsymbol{\alpha} = \left(\mathbb{X}^T \mathbb{X}\right)^{-1} \mathbb{X}^T \boldsymbol{W}_{\mathrm{b}}$ into (A14) with $t = t_v$, thus obtaining:

$$W_{t_v} = \boldsymbol{X}(t_v)^T \left(\mathbb{X}^T \mathbb{X}\right)^{-1} \mathbb{X}^T \cdot \boldsymbol{W}_{\mathrm{b}}. \tag{A15}$$

Now, observe that expression (A15) is an explicit form of (5) with a *linear* space-invariant operator. Thus, we have shown:

> Under the set of assumptions (1)–(7), any thermal image can be represented by a <u>linear</u> combination of <u>at least</u> eight properly selected basis images. The coefficients of the linear combination depend only on the survey time.

The lower bound for the number of basis images needed for linear prediction is obtained under the set of first-order approximations. Clearly, higher-order approximations do not affect the bound. Similarly, if we assume more complex circumstances, for example, when the sensible and latent heat are not negligible, then again, the number of basis images that are necessary for a minimally accurate prediction of thermal images will only increase. For example, we can use in (A9) a common expression (Price 1989) for the convection flux:

$$H(t,\, T(t)) = \rho_{\mathrm{air}}(t) c_p(t) \frac{T(t) - T_{\mathrm{air}}(t)}{\ln^2(h_1/h_0)} \kappa^2 v(t), \tag{A16}$$

where the variables are ρ_{air}, the density of air; c_p, heat capacity of air; T_{air}, the air temperature; h_0, the surface roughness length; v, the wind speed; a standard height h_1 (=2 m); and κ, the von Karman constant. Then iteratively solving (A9), as shown above, will introduce at least two additional independent terms into (A13).

Appendix B: An example of a basis time search algorithm

As before, we let $W(s, t)$ denote the *observed value* at spatial location s and at time t. Consider $\widehat{W}(s, t, \boldsymbol{\tau})$ (i.e. the predicted value of $W(s, t)$) as a function of the basis time vector $\boldsymbol{\tau}$ of length P. Let K denote the number of prediction indicators used to determine $\widehat{W}(s, t, \boldsymbol{\tau})$. Let Δt denote the time interval between the images in the Selection period, and $\boldsymbol{h} \overset{\mathrm{def}}{=} \boldsymbol{\tau} - \boldsymbol{\tau}_0$.

To search for the best displacement vector $\boldsymbol{h}$, when Δt is small, we can use a locally optimal algorithm that exploits the image intensity gradient with respect to $\boldsymbol{\tau}$ and iteratively updates the displacement. We use the prediction error function $E(\boldsymbol{\tau})$, which is the sum of square differences over the *Test* subsequence (TST) images that is adjusted for the degrees of freedom:

$$E(\boldsymbol{\tau}) = \frac{K}{K-P} \sum_{t \in \mathrm{TST}} \sum_{s} \left[W(s,t) - \widehat{W}(s,t,\boldsymbol{\tau}) \right]^2, \tag{B1}$$

where P is the number of elements in vector $\boldsymbol{\tau}$. Linearizing $\widehat{W}(s, t, \boldsymbol{\tau})$ near $\boldsymbol{\tau_0}$, we obtain the following minimization problem with respect to $\boldsymbol{h}$:

$$E(\boldsymbol{h}) = \frac{K}{K-P} \sum_{t \in \mathrm{TST}} \sum_{s} \left[W(s,t) - \widehat{W}(s,t,\boldsymbol{\tau}_0) - \frac{\partial}{\partial \boldsymbol{\tau}} \widehat{W}(s,t,\boldsymbol{\tau}_0) \cdot \boldsymbol{h} \right]^2 \to \min_{\boldsymbol{h}}. \tag{B2}$$

Then the unknown vector $\boldsymbol{h}$ is found as a solution of the system:

$$\frac{\partial E(h)}{\partial h_j} = \sum_{t \in \mathrm{TST}} \sum_{s} \left[W(s,t) - \widehat{W}(s,t,\tau_0) - \frac{\partial}{\partial \tau} \widehat{W}(s,t,\tau_0) \cdot h \right] \cdot \frac{\partial}{\partial \tau_j} \widehat{W}(s,t,\tau_0) = 0 \tag{B3}$$

$j = 1, \ldots P,$

where the partial derivative $\frac{\partial}{\partial \tau_j} \widehat{W}(s, t, \boldsymbol{\tau}_0)$ is approximated by

$$\frac{\partial}{\partial \tau_j} \widehat{W}(s, t, \boldsymbol{\tau}_0) = \frac{\widehat{W}\big(s, t, \tau_{0,1}, \ \ldots, \tau_{0,j-1}, \tau_{0,j} + \Delta t, \tau_{0,j+1}, \ \ldots, \tau_{0,P}\big) - \widehat{W}(s, t, \boldsymbol{\tau}_0)}{\Delta t}.$$

This algorithm (or its weighted version) can be combined with multi-scale smoothing of the Selection period sequence across time.