**Abstract**

תקציר של 02 שורות אשר יתאר את תוכנית העבודה של כל המשתתפים במחקר וכיצד הם משתלבים יחדיו.

יש להתייחס ל:

1. רקע מדעי.

0. מטרות המחקר.

0. מתודולוגיה.

0. חדשנות ויישומיות.

5. מידת שיתוף הפעולה בין חוקרים**Research Program: Improving Satellite Based estimation of PM2.5 and PM10 Exposure Modeling for Epidemiological Studies across Israel**

* 1. **Summary**

**1.2 Scientific Background**

Exposure to particulate matter (PM) air pollution has been associated with multiple adverse human health effects which have been extensively documented **(Atkinson 2013** Kloog et al. 2013; Zanobetti et al. 2009; Franklin et al., 2007; Dominici et al., 2006; Schwartz et al., 1996**)**…. . The World Health Organization has estimated that PM2.5 (PM less than 2.5 µm in diameter) contributes to approximately two million premature deaths per year, ranking it as the thirteenth leading cause of mortality worldwide (Lozano et al., 2013; WHO, 2002). PM2.5 has been associated with, *inter alia*, increased risk of myocardial infarction (D’Ippoliti et al., 2003; Hodas et al., 2013; Madrigano et al., 2012; Rich et al., 2013; Zanobetti and Schwartz, 2005), reduced birth weights (Bell et al., 2007; Kloog et al., 2012c; Zeka et al., 2006), cardiovascular disease (de Hartog et al., 2009; Dominici et al., 2006; Kloog et al., 2012b; Mann et al., 2002; Pope et al., 2011; Schwartz, 1997) and respiratory disease (Dominici et al., 2006; Kloog et al., 2013, 2012b; Schwartz, 1996). Hence, PM air pollution appears to be an important modifiable factor that affects public health on a global scale. Although these studies have included increasingly sophisticated analytical methods, wide gaps remain in our understanding of how human health is affected by PM. Typically, exposure to PM2.5 in such epidemiological studies is assessed using measurements from air quality monitoring stations. However, in many geographical areas ground PM2*.*5 monitoring networks are sparse and non-ideally deployed for environmental health studies. Satellite remote sensing can be used to assess PM2*.*5 in these areas (van Donkelaar et al., 2010; Gupta and Christopher, 2008; Gupta et al., 2006; Koelemeijer et al., 2006; Engel-Cox et al., 2004). A common satellite product for predicting PM2*.*5 concentrations is the aerosol optical depth (AOD). AOD measures light extinction at given wavelengths due to scattering and absorption along the measured atmospheric column. The Moderate Resolution Imaging Spectroradiometer (MODIS) land AOD products are calculated by two operational algorithms: the Dark Target (DT) algorithm and the Deep Blue (DB) algorithm. The DT algorithm uses as an input MODIS Level 1 data (L1B), including the integrated spectral reflectance of 7 bands in the 470-2110 nm range as well as the geo-location, and processes them by a set of rules to retrieve the AOD (Levy et al., 2009). However, it is well recognized that due to algorithmic limitations retrievals of AOD above dark surfaces (ocean, rainforest, etc.) are considerably more reliable than retrievals over bright and highly reflective surfaces, such as deserts, urban areas, snow, and coastal regions (Martin, 2008). To overcome this limitation, the DB algorithm processes spectral data mainly from the blue (412-470 nm) spectral channels (Hsu et al., 2013; 2006), since bright surfaces show high reflectivity in the red bands of the visible and the near infrared spectrum and low reflectivity in the blue spectral region. The DB algorithm has been found to outperform the DT algorithm over bright surfaces and to retrieve more accurate AOD values (Hsu et al., 2013; 2006; Sayer et al., 2013, Sorek-Hamer et al., 2015).

Furthermore, the recently developed Multi Angle Implementation of Atmospheric Correction (MAIAC) retrieval algorithm (1 km spatial resolution; Lyapustin et al., 2011a,b) revealed improved PM predictions and large spatiotemporal availability   
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
(Chudnovsky et al. 2013b; Kloog et al. 2014a). Similar results were obtained also when using MODIS/MAIAC retrievals within different regression models   
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
(Chang et al. 2014; Hu et al. 2014). Nonetheless, although MAIAC AOD was successfully used for estimating PM2.5 concentrations in North America, it has not been tested yet for estimating PM10 concentrations nor has it been evaluated in more complex geographic and climatic regions as Israel.

Another satellite based data source, launched in 2011, is the polar-orbiting mission Visible Infrared Imaging Radiometer Suite (VIIRS). VIIRS is supposed to substitute MODIS with a global coverage every two-days and a spatial resolution of 6km (Liu et al., 2014).

The broad spatial coverage enabled by satellites allows us to expand exposure data far beyond the range of conventional ground monitoring networks to penetrate rural and suburban areas which greatly enhances our ability to estimate subject-specific exposures.

**1.3 Relevancy of proposed research to MoST call**

A major aim in the MOST call is the use of earth observations from space. In this proposal, we make extensive use of novel satellite technology from multiple satellites and satellite based algorithms that are centered on earth observations:

Satellite AOD product quality is not uniform across the globe. This is due to multiple issues ranging from data availability, cloud cover and pixel accuracy to surface brightness and unhomogenity . By incorporating new statistical and methodological techniques we will improve upon the current satellite based AOD data quality which in turn will allow us to improve our ability to accurately estimate ground level PM2.5 and PM10 concentrations at high spatial and temporal resolutions.

**1.4 Research gaps and state of the art**

Epidemiology research tends to exploit a variety of methods to assign exposure, including nearest monitor assignment ,simple kriging methods, land use regressions (LUR) and more recently satellite based hybrid models. Exposure metrics associated with most of these above methods ignore intra-urban variation in exposure and may be missing data in rural areas. LUR is much more suitable for capturing long-term differences in exposure between locations. There are, however, a number of limitations when using land use terms and LUR: 1) they are generally not time varying and their temporal resolution tends to be limited and 2) they are based on sparse PM2.5 monitoring networks and/or require specific and costly PM2.5 measurement campaigns (Gryparis et al., 2009; Yanosky et al., 2009, 2008). Hence, they are only applicable to long-term exposure assessments, which are adequate for studies on chronic health effects but are useless regarding acute health effects. Moreover, the lack of monitors in semi-urban and rural locations means participants from such locations are either excluded or are assigned estimates from the land use regression that may have greater error. Satellite based data, and particularly AOD has allowed us in recent years to develop better estimation models (Chang et al., 2013; A. Chudnovsky et al., 2013; Cordero et al., 2013; Gupta et al., 2013; Kim et al., 2013). Traditionally, the health exposure studies have used the standard MODIS (Moderate Resolution Imaging Spectroradiometer ) AOD product of the “Dark Target” algorithm (Levy et al., 2007) which has a nadir resolution of 10x10 km2. Lately, AOD at significantly higher spatial resolution (1x1 km2) has been offered by the MAIAC algorithm. Chang and colleagues (Chang et al., 2013) used some novel ideas such as a statistical downscaling and data fusion techniques to predict PM2.5 concentrations at spatial point locations in the southeastern United States during the period 2003–2005. Their model showed relatively high cross-validated predictions (R2=0.78 and a root mean-squared error (RMSE) of 3.61 mg/m3, However, the developed model was restricted to retrieval days. Our group has also developed and published several studies using MAIAC based data (kloog 2014,under review). While we have successfully shown good model performance in the New England region (mean out-of-sample R2 =0.88, slope of predictions versus withheld observations = 0.99) model performance in other regions such as Mexico and Israel was moderate at best. For example in Israel model performance was R2=0.79 and 0.72 for PM10 and PM2.5 respectively, while in Mexico R2 of 0.72. These studies clearly show that in areas with very different geo-climate characteristics the model does not perform as well in large part due to the lower quality of the RAW AOD data product. This proposed study will help improve upon the current state of the art specifically in areas where currently AOD data isn’t preforming as well.

**1.5 Collaborations**

The principal investigators from Ben Gurion University and the Technion have successfully collaborated in projects before resulting in multiple publications. Prof Broday will provide the ground monitoring PM data and lead the conduct of the comparative analysis. He will provide

ground monitoring expertise throughout the study. Dr. Kloog will lead the work with the satellite data (e.g. MODIS, MAIAC) and provide his results for assessment of exposure and provide GIS, statistical and remote sensing expertise throughout the study.

**1.6 Benefits and expected significance**

Among the innovations inherent to the proposed study are: …

**1.7 Detailed description of the proposed research:**

The general goal of the proposed research is to improve the use of satellite based AOD data for developing state-of-the-art computationally efficient approaches that will combine novel physical satellite measurements to obtain daily validated and accurate high-resolution estimates for PM2.5 and PM10 . Specific goals are listed below.

**Aim 1— developing computationally efficient new modeling algorithms to improve upon the current state of the art modeling methodology**

Extending our initial model methodology in Israel (Kloog et al.,under review), We aim to develop and validate (using ten-fold cross validation) better methodologies to improve upon our initial estimation models with moderate predictive power to better estimate air temperature at a very high spatial resolutions (1×1 km and 200x200m) and temporal (daily) resolution across Israel for 2003-2015. We will make use of new methodological developments such as spatio-temporal moving windows to assess daily AOD quality, the use of very high spatial resolution meteorology predictors based on the Weather Research and Forecasting (WRF) and a new algorithm combining both aqua and terra MAIAC data.

**Aim 2- developing an approach to fuse multi-level satellite data ( MODIS, VIIRS and MAIAC) using advanced data fusion techniques (Kalman filtering) for enhancing satellite AOD data quality.**

Based on our previous work (Kloog et al., 2012d, 2011), we aim to improve upon AOD data quality in Israel by fusing multi-level satellite data (MAIAC, MODIS, VIIRS) allowing us to develop better PM estimation based models.

**Aim 3**- Characterizing the spatial and temporal features of global ground locations to better understand the statistical relationship between ground monitored particulate matter and satellite observations.

**Aim 4-** Reproducibility and Dissemination: Develop and share software tools for record linkage and implementation of epidemiological methods to achieve research reproducibility.

We briefly describe below the methods and models that will be used in our proposed research. Our preliminary results follow.

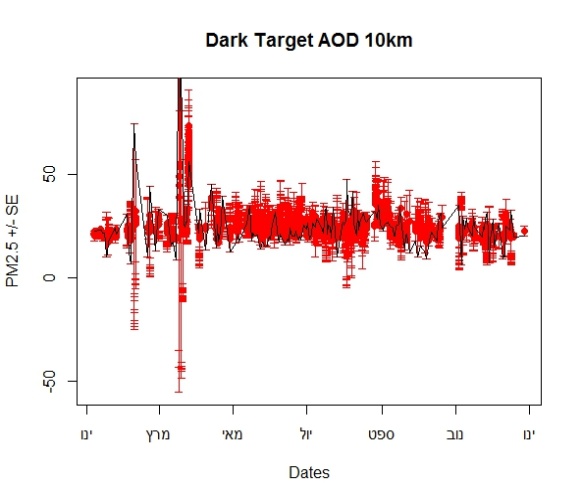
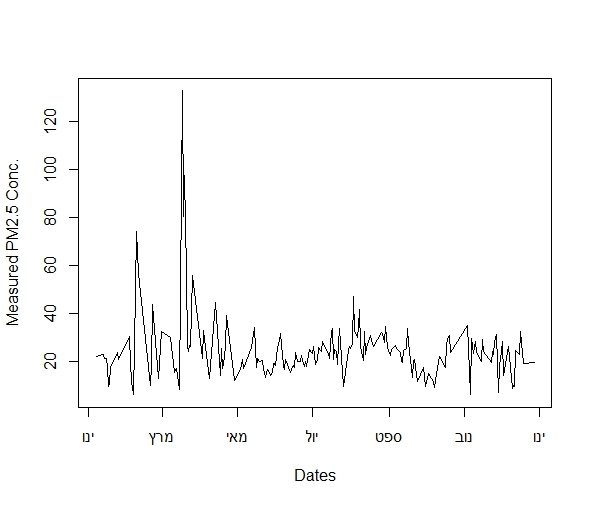
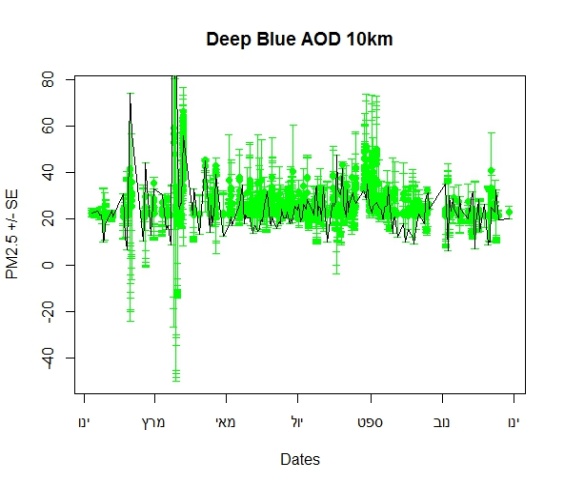
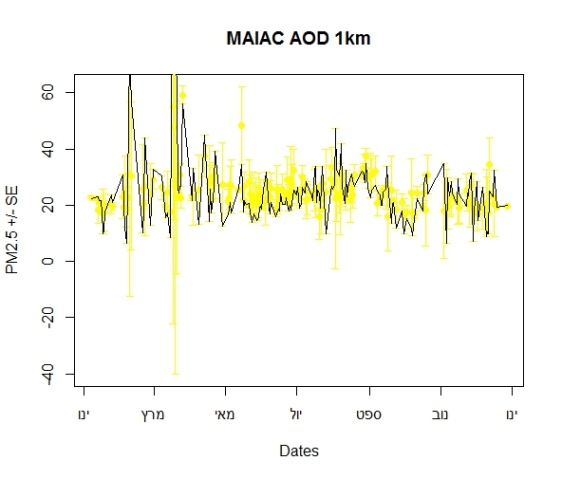
**Aim 1:**

**Aim 2**: There are different satellite-based sources for AOD retrievals which differ in their spatial resolution and retrieval algorithm: DT-AOD 10km, AOD 3km; DB- AOD 10km, MAIAC- AOD 1km, VIIRS- AOD 6km).

Deep Blue (DB) products have been valuable over bright areas. It has a 10km product on MODIS and a 6km product on VIIRS, soon a 3x3km product should be available for us.

As each of the datasets mentioned above are based on a different retrieval algorithm, the availability, and the AOD values change between datasets in time and space. In order to enhance data availability and estimation quality we propose to fuse the DT, DB and VIIRS data with MAIAC data through the use of Kalman filtering . The Kalman filter is over 50 years old but is still one of the most important and common data fusion algorithms in use today in a large variance of fields (Faragher, 2012) where the state space of the latent variables is continuous and where all latent and observed variables have a Gaussian distribution (often a multivariate Gaussian distribution). Theoretically, it has been called the linear least mean squares estimator (LLMSE) because it minimizes the mean-squared estimation error for a linear stochastic system using noisy linear sensors (Grewal and Andrews, 2015) although it works well also for non-linear systems. It has been used in several fields for estimating the state of dynamic systems and for performance analysis of estimation systems. Kalman filter is widely used for prediction of missing data (e.g. Seto and Arai, 1999) , but to our knowledge it hasn't been used for improving PM estimations based on satellite data.

The problem definition: Find an estimate x of state vector x from incomplete and noisy measurement vector y while x is represented by AOD retrieved from different sources, and y is ground PM measurements.

(d)

(c)

(b)

(a)

Figure 1. Collocated data time series for the year 2008, for station KMR (a) measured daily PM2.5 concentrations (b) daily PM conc. estimated by DT-AOD 10km (c) daily PM conc. estimated by DB-AOD 10km (d) daily PM conc. estimated by MAIAC 1km.

Figure 1 illustrates the estimated PM2.5 concentrations +/- SE based on linear regressions while the black line in all plots represents the measured concentrations. Our goal is to find the best combination of AOD datasets for each collocated pixel in order to minimize the PM estimation model based errors.

Methodology using Kalman Filter: Fuse the noisy signals from several sources with different temporal and spatial resolutions (e.g. MODIS: AOD 10km, AOD-DT 3km, AOD-DB 10km, MAIAC 1km; VIIRS: AOD 6km, MAIAC 1km) to improve the estimate of a certain variable (e.g. PM2.5 / PM10 concentrations).

The state space model is represented by the transition equation and the measurement equation. Let m be the dimension of the state variable, d be the dimension of the observations, and n the number of observations. The transition equation and the measurement equation are given by αt+1 = dt + Tt · αt + Ht · ηt yt = ct + Zt · αt + Gt · t, where ηt and t are iid N(0, Im) and iid N(0, Id), respectively, and αt denotes the state variable. The parameters admit the following dimensions: fkf 3 at ∈ Rm dt ∈ Rm ηt ∈ Rm Tt ∈ Rm×m Ht ∈ Rm×m yt ∈ Rd ct ∈ Rd t ∈ Rd . Zt ∈ Rd×m Gt ∈ Rd× (Luethi et al., 2015).

Kalman filtering will be performed at each site on each day to assimilate all products for an optimized estimation.

In this step, variance for process noise at monitoring site i is defined as , where Xi and Ui are monitoring results and estimation at site i respectively. Similarly, variance for observation noise is defined as , where Zi is AOD-derived PM2.5 estimation at site i.

We will later compare the assimilated PM estimation from the use of the Kalman filter with both AOD based predicted PM and monitored PM at each monitoring site.

**Aim  3:** A large variance in the Pearson correlations (0.06-0.26) between AOD-PM in Israel stations (e.g. Figure 2) has been seen with no clear trend in time nor in space. This section will be based on a large database obtained from different sources in Israel, and Europe over approx. ten years. In general, each ground PM station location has a table of spatial characteristics (e.g. geographical coordinates, height above sea level) and characteristics altering in time (e.g. brightness of surface, temperature, etc.).



Figure 2. Univariate correlation between PM10 and collocated MAIAC AOD 1km data on days not affected by dust.

Each location has the same table of characteristics but with different values per location per time point. These

In order to try and better understand if their are certain conditions characterizing a ground location that reveal a better correlation with satellite variables, we will be using clustering methods (unsupervised sparse clustering; Written & Tibshirani, 2013) in order to differentiate between locations based on spatiotemporal characteristics that clearly resolve a high/low correlation with the satellite products and moreover explain the whole range of results.

*Preliminary results*

In recent weeks we began to incorporate the novel 1 × 1 km MAIAC data into our New England models. Preliminary results were extremely encouraging, with cross validated R2 values of 0.91 in the first stage of the model. Figure 3 shows the resulting 1 × 1 km PM2.5 prediction surface for November 15, 2003, from these preliminary results. Figure 4 shows the difference of the estimated local pollution from the average PM2.5 concentrations (local PM) at very fine resolution (200 × 200 m)………….

**1.8 Available resources**

Some of the analyses and data processing can be performed on our standard Linux machines with an i5 core (8 threads) processor. For the data fusion and generation of estimates, an ultra-high end computer (16 cores, 256 GB of RAM) is needed, therefore we ask the MOST to fund an ultra-high end computer. The PI’s are well experienced with exposure modeling and work in laboratories that provides complementary infrastructure for the proposed research at both Ben-Gurion University and the Technion. The PI’s have vast experience in designing, building and running spatio-temporal estimation models, and a strong background in computer programming and GIS analysis, remote sensing and working and processing satellite data. We believe this experience will provide a firm basis for the proposed study. The labs currently provide ARCGIS, MATLAB and ERDAS IMAGINE licenses for the database design and data processing. SAS and R license are available for the statistical analysis.

**1.9 Bibliography**

**1.10 Relevant publications**

**1.11 suggested reviewers**

**2.1 Timetable of the Work Plan**



**Figures**