Spatiotemporal machine learning modelling with season-trend decomposition to capture and quantify extreme particulate matter air pollution (PM2.5) attributable to bushfire, dust and woodheater inversion layer events for health research

Ivan Hanigan PhD[1,2] and “CAR/CSA” Bushfire smoke exposure team

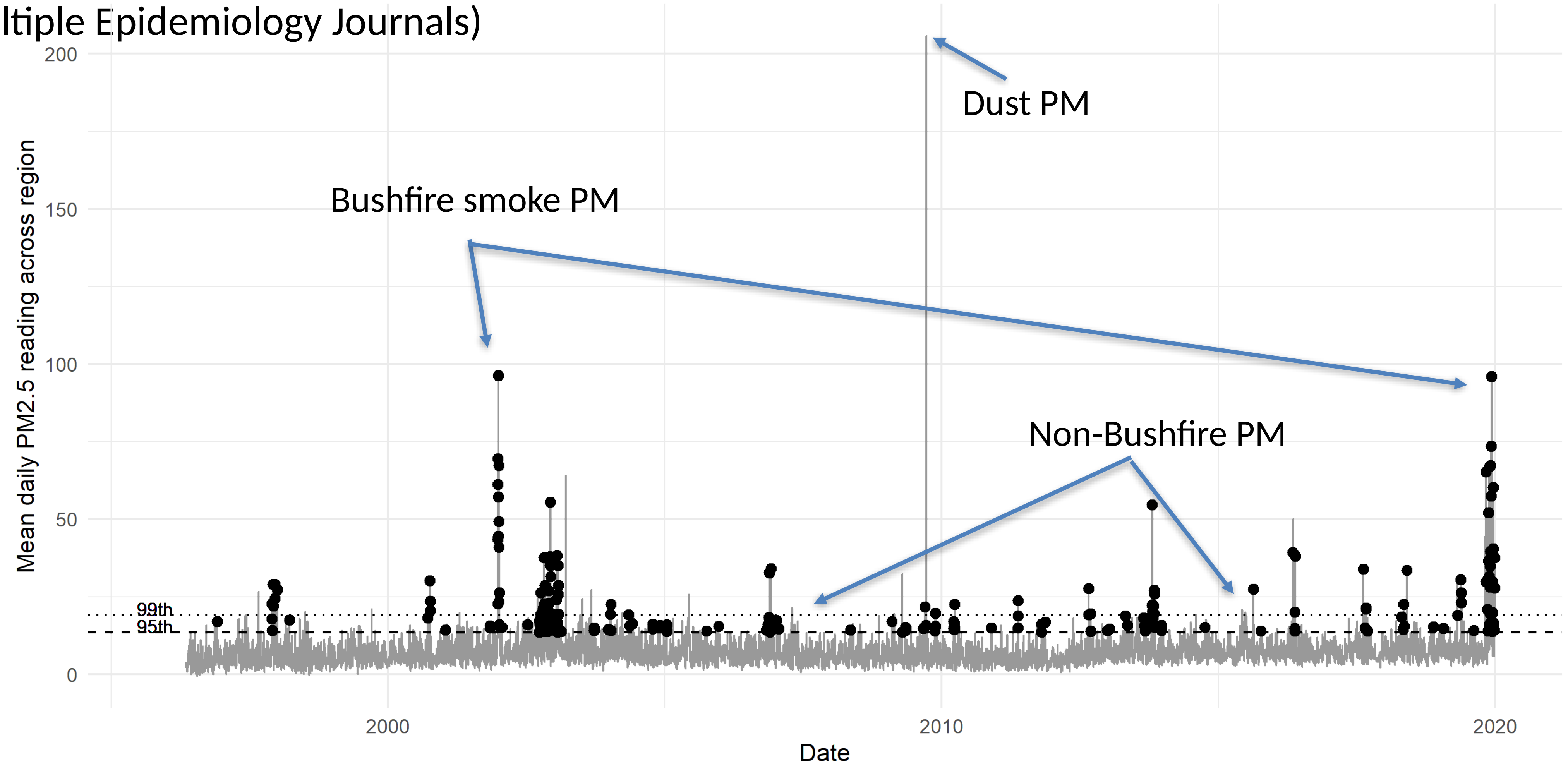
[1] WHO Collaborating Centre for Climate Change & HIA Curtin University:
https://research.curtin.edu.au/whocc-cchia/

[2] Healthy Environments And Lives (HEAL) National Research Network, Australia

Not all PM2.5 is bushfire events= relevant to public health

- E.g. Valid Fire Events of Airborne Particles (PM2.5) used in health studies for Sydney (Published in multiple Epidemiology Journals)

Sydney
NSW



Hanigan, Morgan, Williamson et al (2018). Extensible Database of Validated Biomass Smoke Events for Health Research. MDPI Fire, 1, 50; http://doi.org/10.3390/fire1030050

Statistical and machine learning methods
GIS Air Quality Modelling

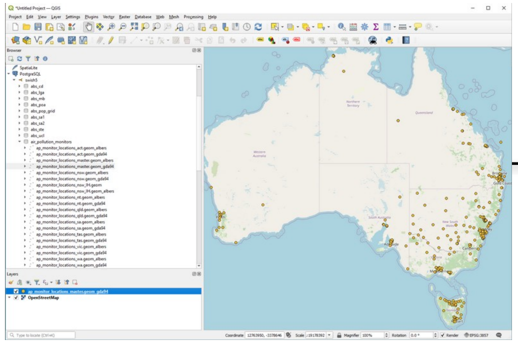
Daily $PM_{2.5}$ Observations

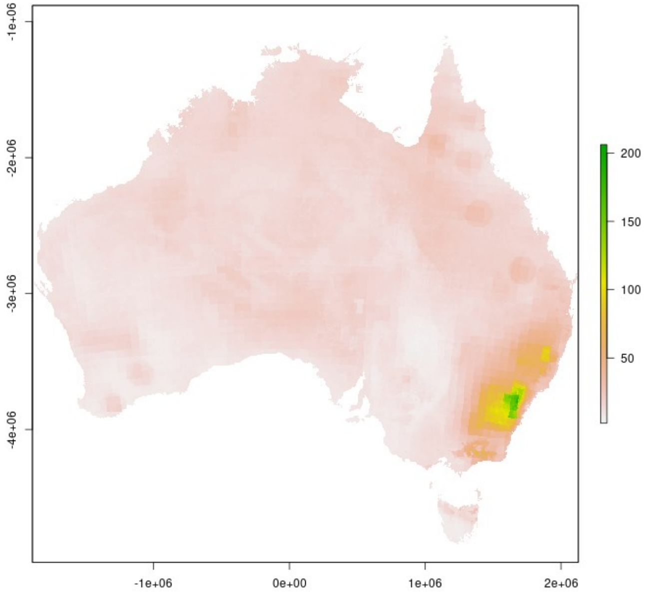
Other covariates (e.g. Sat, Land, Road)

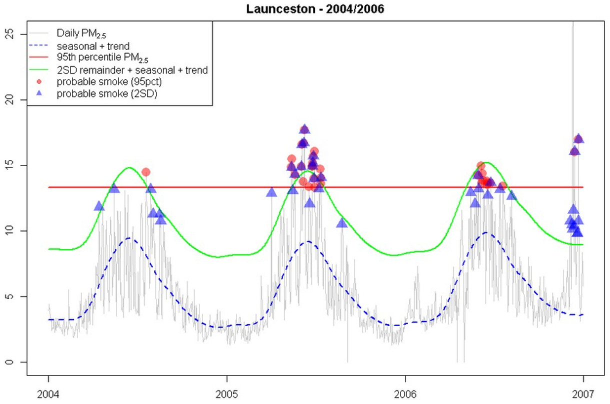
Predicted Daily $PM_{2.5}$ 0.05deg/5km

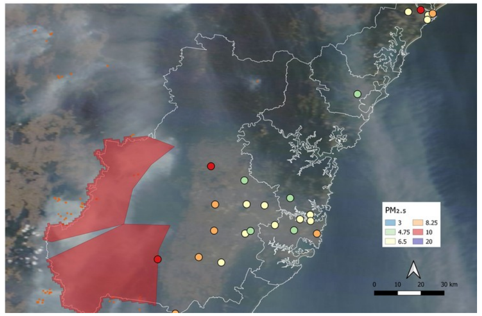
STL Decomposition

Validation "ground truthing"







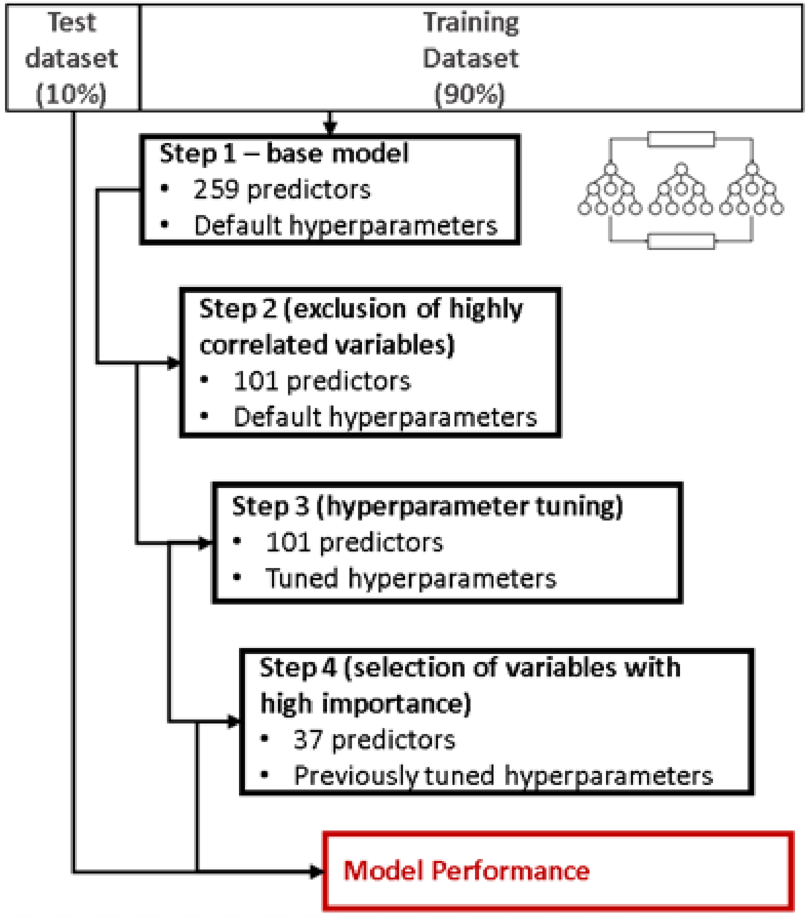


Statistical and machine learning methods
$PM_{2.5}$ (Random Forest Machine Learning)

1) Geospatial predictors collected on CARDAT's data platform:
   E.g. temperature, rainfall, drought, fires, burned, AOT, elevation, landcover

Clean Air Research
Data and Analysis Technology
(CARDAT)

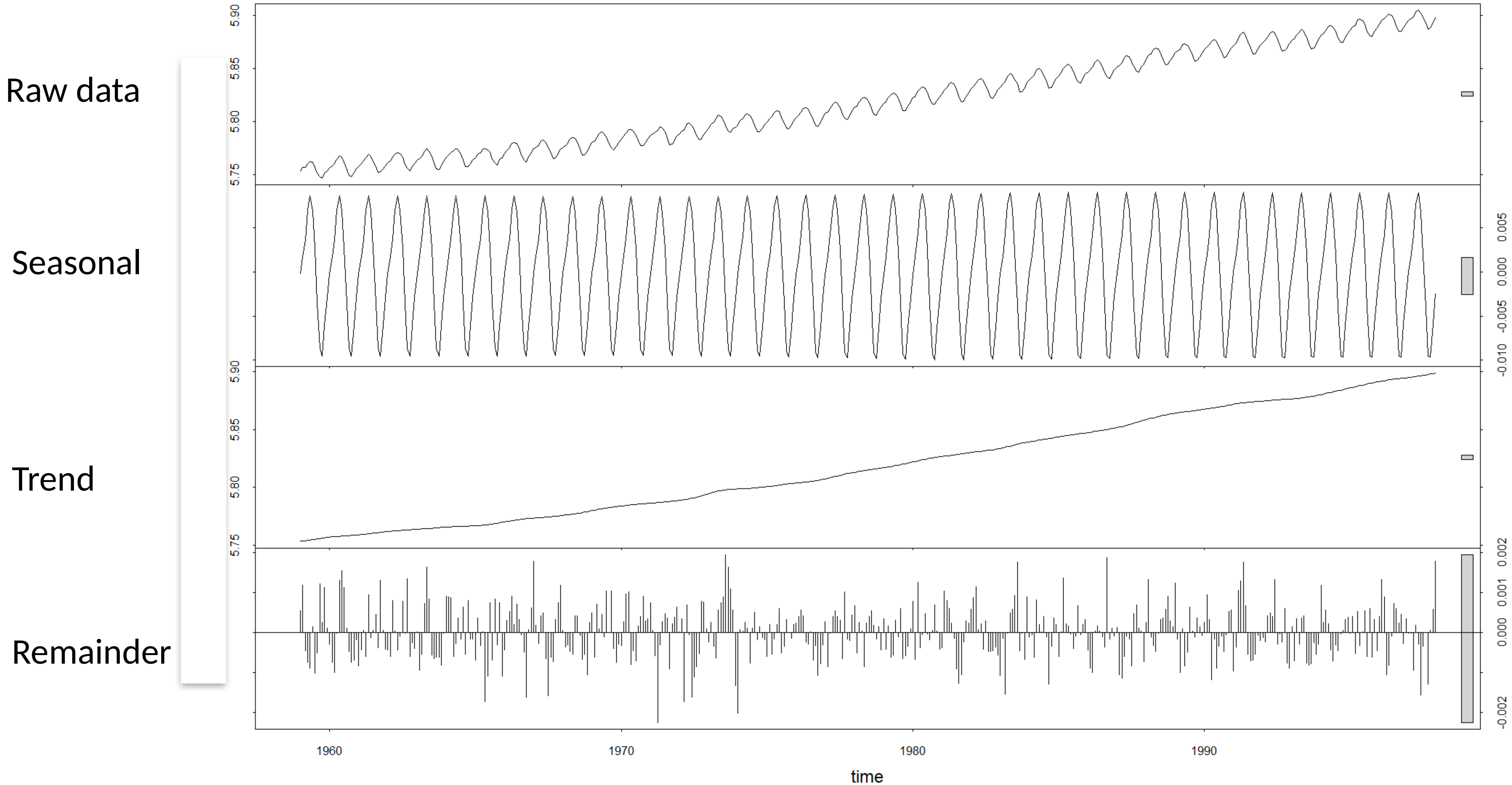
2) $PM_{2.5}$ modelling



Model	Step 1 (base model)	Step 4 (final model)
# variables	259	37
RMSE (oob)	4.82	4.42
R-squared (oob)	66.7%	72.0%
RMSE (test)	4.40	4.07
R-squared (test)	64.3%	69.7%
RMSE (train)	2.42	2.07
R-squared (train)	93.8%	95.2%

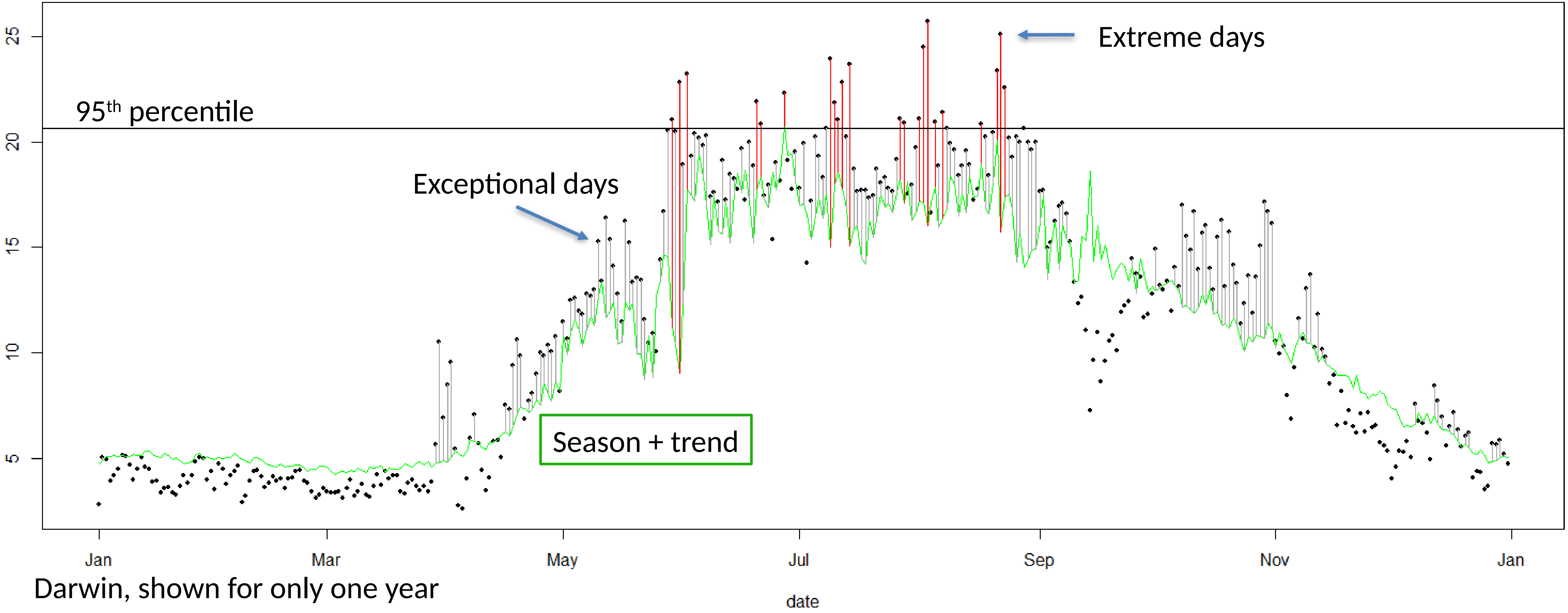
Seasonal Trend Lowess (STL) Decomposition

Robust time-series modelling technique: extract season, trend and remainder



STL analysis on dailydata

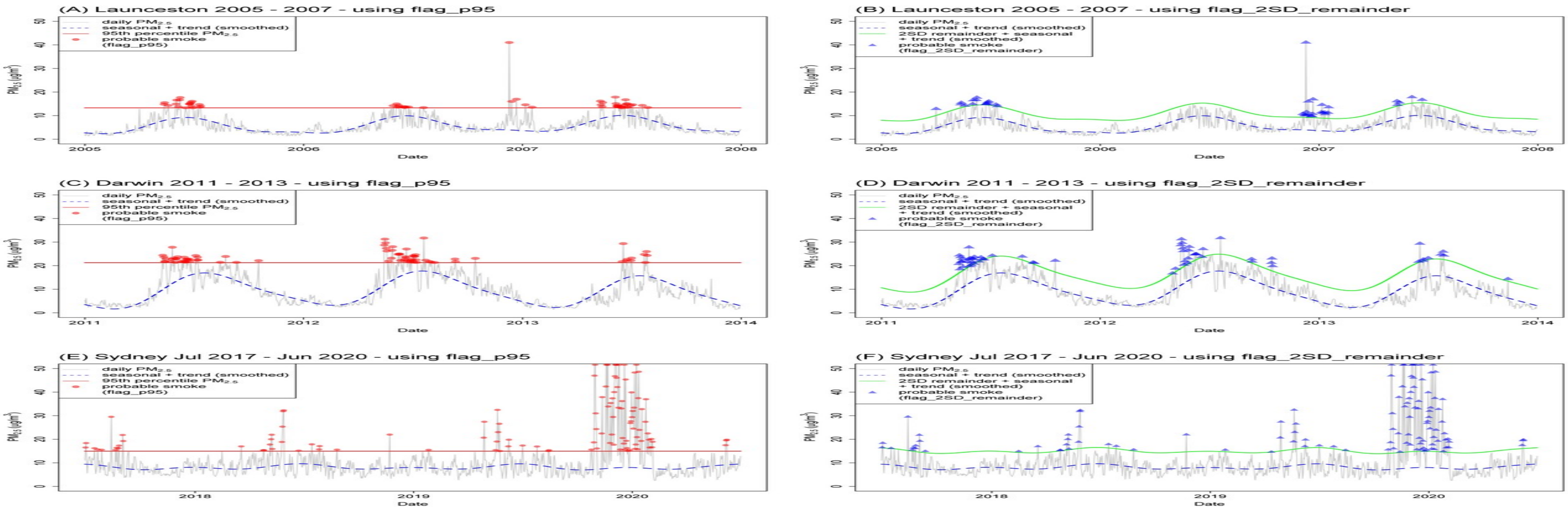
Results look “noisy”. The 95th percentile can be used to identify extreme and exceptional events



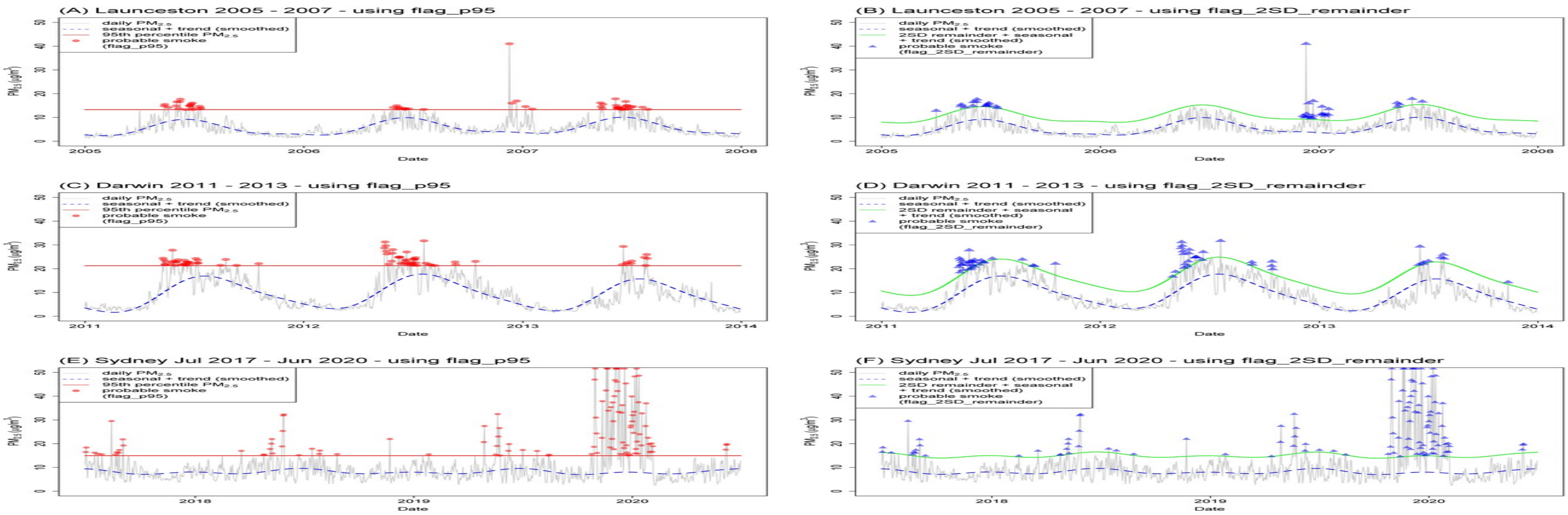
Darwin, shown for only one year

Flags in high winter woodsmoke city

Frequently used criterion is 95th percentile, or 2 standard deviations of the error term



Flags in low seasonality location



Validated events database

- 16 sites, manual review
- Additional flags
  - i) active fire hotspots
  - ii) CTM dust (CAMS)
  - iii) heating deg days
- Classification and Regression Tree CART model to predict validated events DB
- USYD team TODO

Hanigan, Morgan, Williamson et al (2018). Extensible Database of Validated Biomass Smoke Events for Health Research. MDPI Fire, 1, 50; http://doi.org/10.3390/fire1030050

Validated events database: current gold standard, validated by manual process

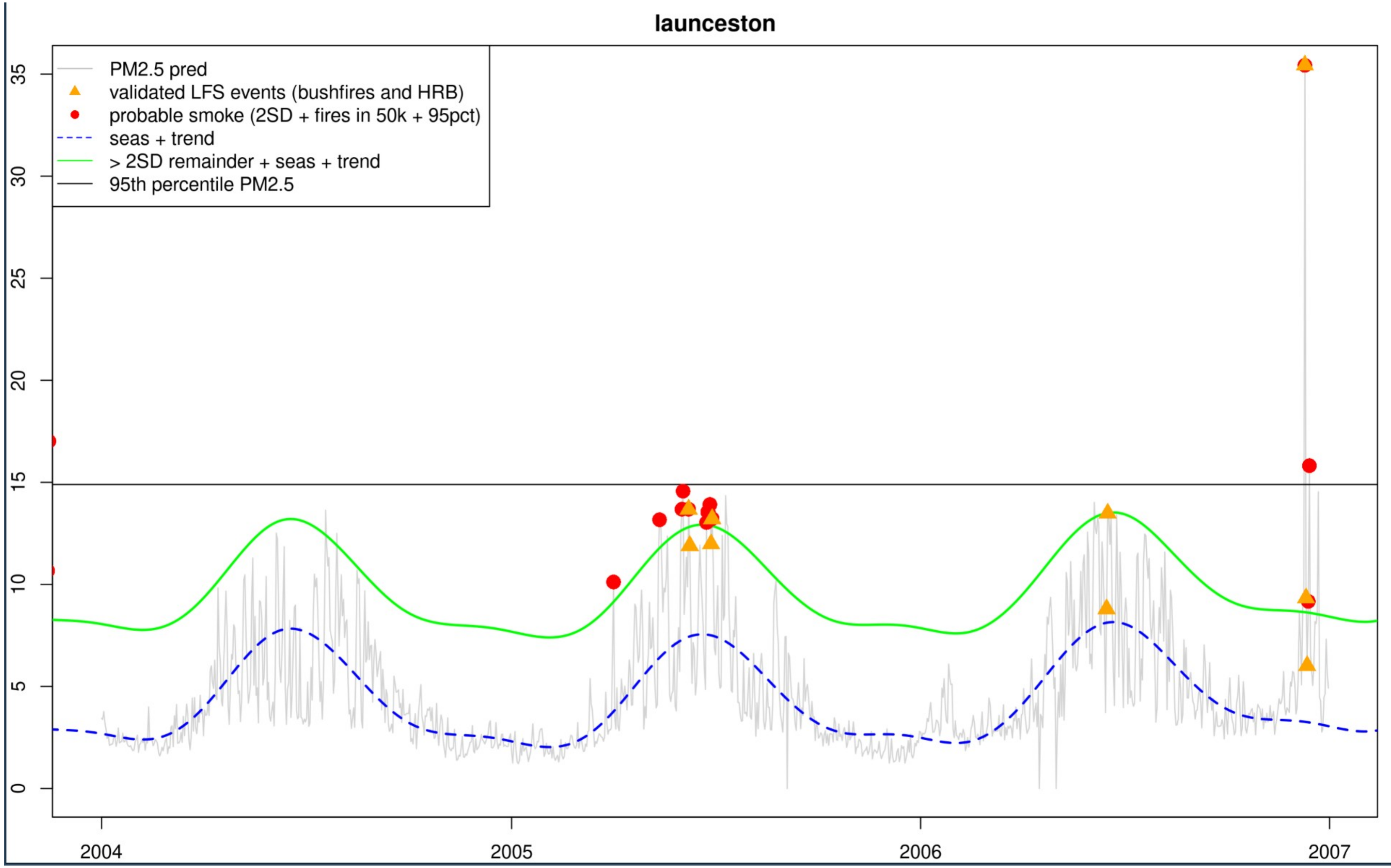
Obvious 20/10/2013

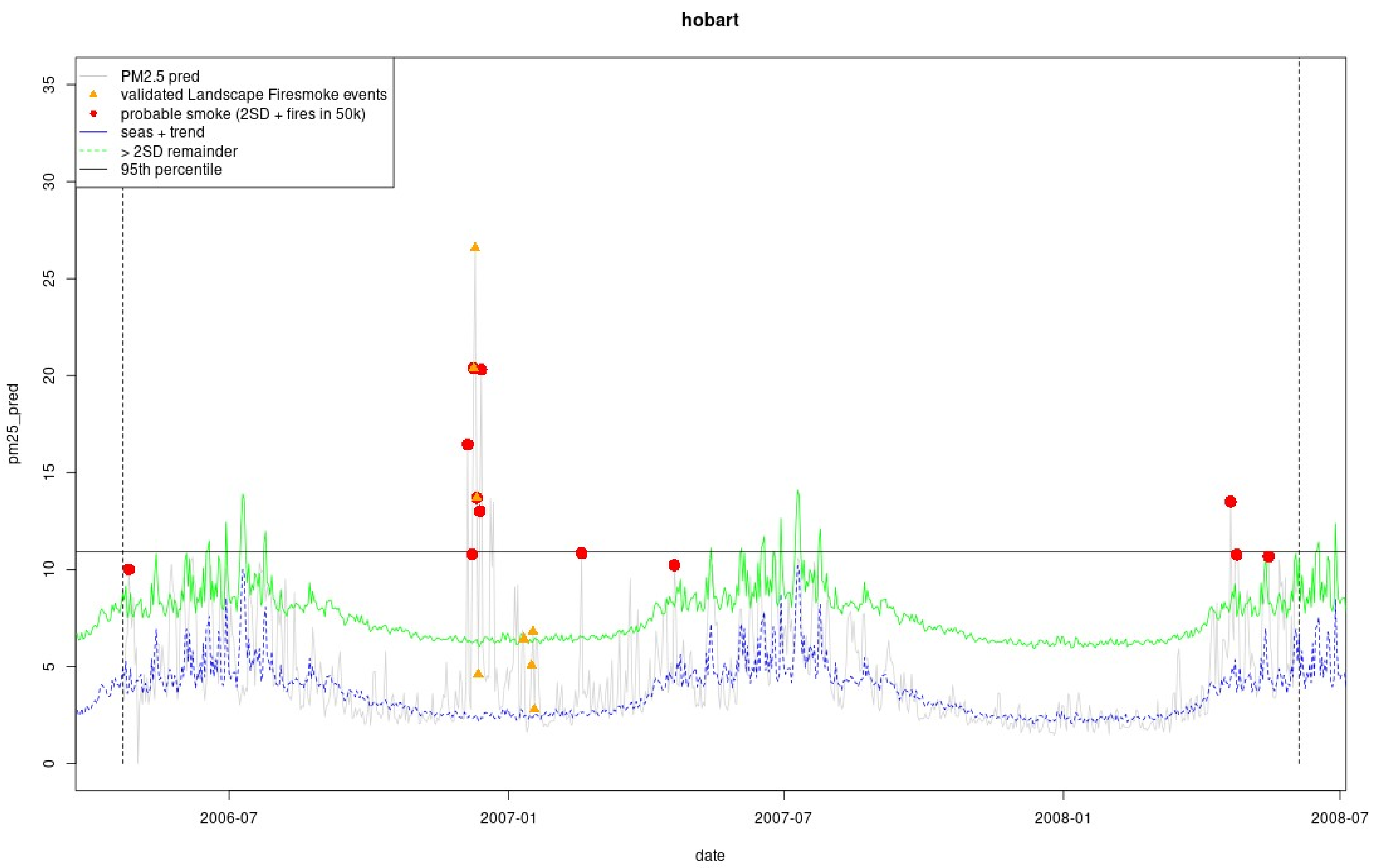
Ambiguous 04/11/2016

Using validated events and the flags we can find coinciding events or false positives, false negatives

Launceston, TAS

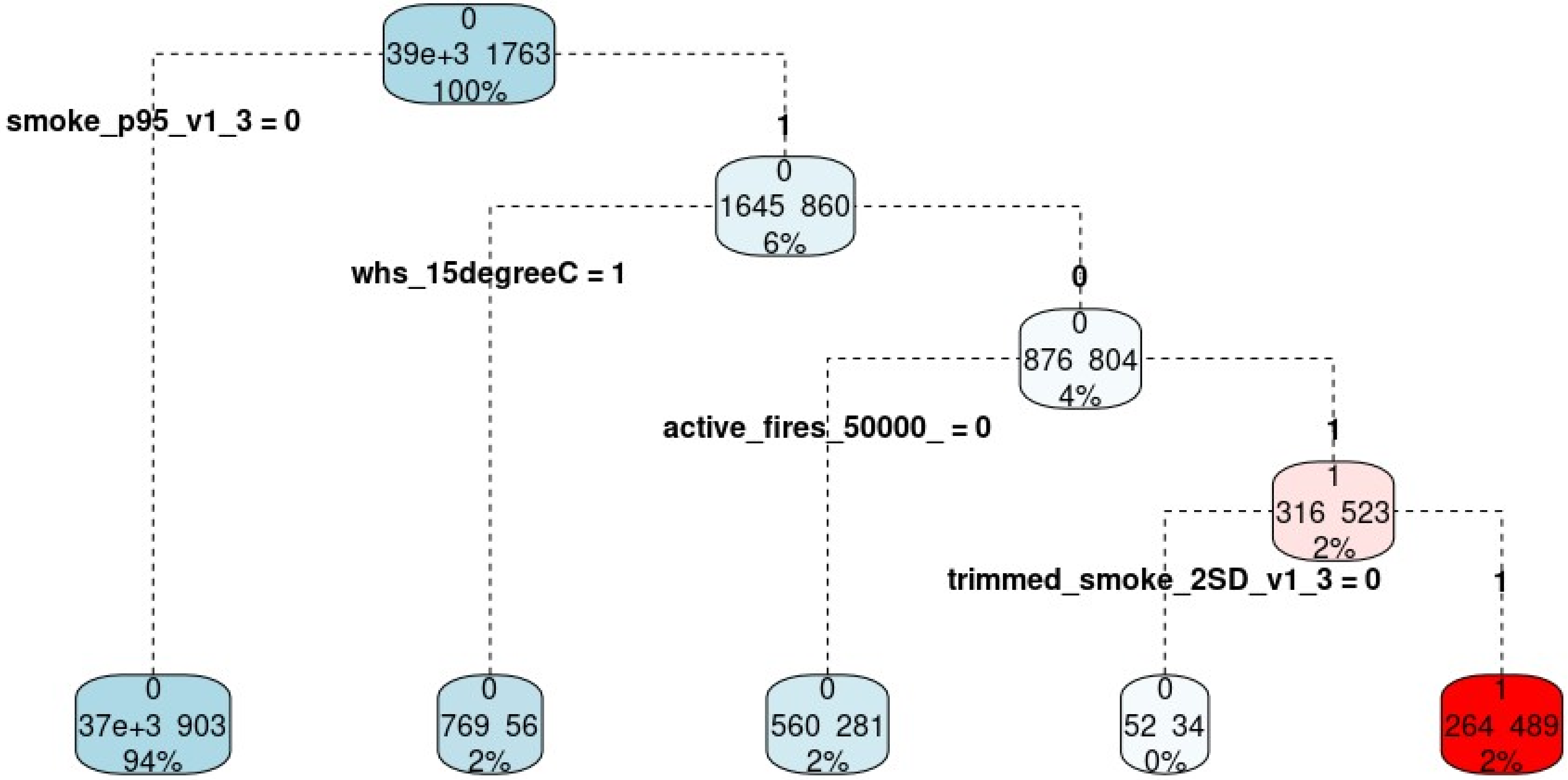
PM2.5




hobart
PM2.5 pred
validated Landscape Firesmoke events
probable smoke (2SD + fires in 50k)
seas + trend
> 2SD remainder
95th percentile
pm25_pred
0
5
10
15
20
25
30
35
2006-07
2007-01
2007-07
2008-01
2008-07
date

Classification tree machine learning model

Preliminary results



IF > 95th pctl AND heating-degree-days-less-than-15deg AND active fires within 50K AND remainder > 2SD from season-trend THEN probably a true positive.

Acknowledgements

- Nicolas Borchers-Arriagada1,2,3, Geoffrey Morgan3,4,5, Joseph Van Buskirk4,6, Karthik Gopi3,4,5, Cassandra Yuen3,4, Fay H Johnston1,3, Yuming Guo3,7, Martin Cope3,8
- 1. Menzies Institute for Medical Research, University of Tasmania, Tasmania, Australia
- 2. Sohn Hearts and Minds Research Fellow
- 3. Centre for Safe Air, NHMRC Centre for Research Excellence, 17 Liverpool Street, Hobart, Tasmania, Australia (https://ror.org/04ccf0j10r).
- 4. Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Australia
- 5. Health Environments and Lives (HEAL) National Research Network, Australia (Grant No. 2008937)
- 6. Public Health Unit, Sydney Local Health District, Sydney, Australia
- 7. Climate, Air Quality Research Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
- 8. CSIRO Oceans and Atmosphere, Aspendale, Victoria, Australia

Funding

- This research was undertaken with the assistance of resources from the Clean Air Research Data and Analysis Tools platform (CARDAT) which is supported by funds from The Centre for Safe Air (CSA; https://safeair.org.au/) which is funded by the National Health and Medical Research Council (2015584),
- the Curtin WHO Collaborating Centre for Climate Change and Health Impact Assessment, and
- the Australian Research Data Commons (ARDC) AirHealth Data Bridges project (https://doi.org/10.47486/PS022), as well as the ARDC Bushfire Data Challenges project (https://ardc.edu.au/project/assessing-the-impact-of-bushfire-smoke-on-health/) and
- the Australian National Health and Medical Research Council (APP2004514) Ideas Grant - Bushfire smoke exposure during pregnancy and epigenetic changes in offspring.