CNRS
Centre National de la Recherche Scientifique

INFN
Istituto Nazionale di Fisica Nucleare

PROSPECTOR: An Automated Systematic Tool for Noise Hunting

VIR-0361A-15

Kathryn Plant[1*], Irene Fiori[2], and Maddalena Mantovani[2]

[2] *University of California, 1156 High St, Santa Cruz, CA 95064, United States*
[2] *EGO - European Gravitational Observatory*

Date: July 31, 2015

[*] *corresponding author:* kaplant@ucsc.edu

VIRGO* A joint CNRS-INFN Project
Via E. Amaldi, I-56021 S. Stefano a Macerata - Cascina (Pisa)
Secretariat: Telephone (39) 050 752 521 * FAX (39) 050 752 550 * Email: ego@ego-gw.it

Contents

Abstract

This report characterizes a systematic statistical noise hunting tool, PROSPECTOR by describing the working principle and discussing several applications. This statistical approach returns a ranked list of channels according to their prospective correspondence to a set of query events. Channels are ranked by comparing cumulative distributions of the channel state at the event times to the channel state at all times. Among other applications, PROSPECTOR successfully finds a mode cleaner chiller as the cause of brief mirror misalignments [1]. Investigating unlocks and mis-alignments in the mode cleaner is a useful way to study problems that generalize to the entire interferometer.

1 Introduction

The European gravitational observatory is currently in the process of commissioning Advanced Virgo. Part of this work involves Noise Hunting, the process of tracing events in the optical cavities to their causes in the environment surrounding the interferometer. Virgo collects and processes many channels of timeseries data, measuring not only the power in the resonant cavities but also many channels of mirror alignment measurements, actuator corrections, and environmental noise monitors (seismometers, magnetometers, thermometers, etc.). In the current stage of commissioning, the 3-km arms of the interferometer itself are not yet ready to lock (locking an optical cavity is the process of setting in its operating state, on resonance), but the mode cleaner optical cavity does lock (see section 1.1 for detailed introduction to the mode cleaner), and many noise hunting questions involve searching for connections between environmental channels (e.g. seismic events) and problem events in the mode cleaner (e.g. sudden unlocks or mirror misalignments). Investigating unlocks and mis-alignments in the mode cleaner is a useful way to study problems that generalize to the larger interferometer.

This paper characterizes a statistically rigorous code (named PROSPECTOR) that systematically searches for correlations between events in many channels. First, I discuss several preliminary investigations that motivated the development of a multipurpose approach. The second part introduces the most general version of the code and discusses several applications.

1.1 Control of the Input Mode Cleaner

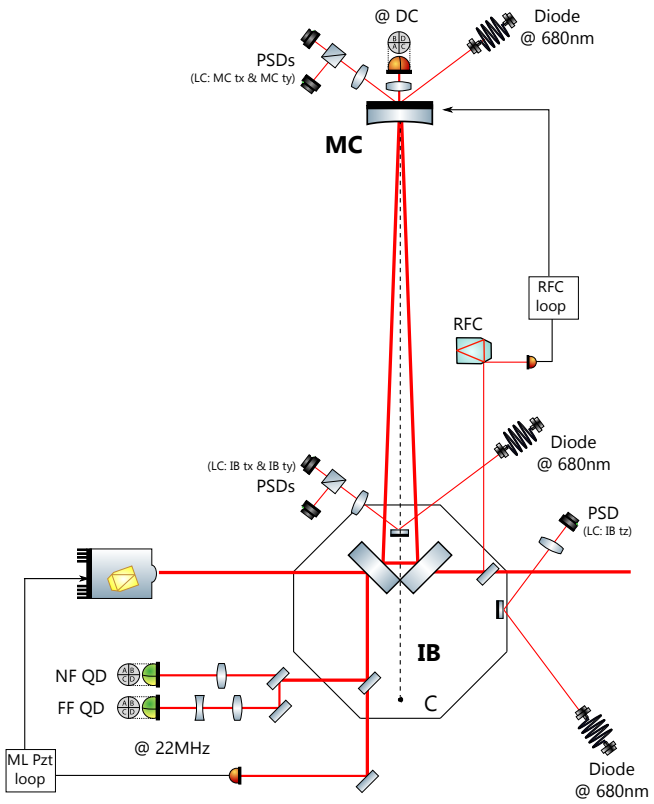

Figure 1: **Diagram of the Input Mode Cleaner control loops.** *Longitudinal loops:* Master Laser PZT loop (ML PZT loop) and Reference Cavity loop (RFC). *Angular control loops for local control:* Several Position-Sensing Devices (PSDs) are placed both close to the MC tower and close to the IB tower; the ones on the end of the mode cleaner (MC) controls MC pitch and yaw (LC: MC tx & MC ty); one of the three on the injection bench (IB) controls IB pitch and yaw (IB tx & IB ty); the other controls IB roll (LC: IB tz) *angular loops for automatic alignment:*The Near field quadrant diode (NF QD demodulated at 22MHz), The Far field quadrant diode (FF QD demodulated at 22MHz) and the end mirror quadrant diode (DC) together control MC pitch and yaw and IB pitch, yaw, and roll.

The input mode cleaner (IMC) is a triangular resonant cavity 144 m long whose purpose is to remove higher order modes from the infrared 1064 μm beam before it enters the central interferometer [2]. Control of the cavity is divided into longitudinal loops (controlling the length of the cavity) and angular loops (controlling the alignment of the mirrors). In the longitudinal loops, both the distance to the mode cleaner end mirror and the frequency of the laser must be controlled in order to keep the IMC locked. Two feedback control loops control these two parameters (see Figure 1). The PZT correction loop, which error signal operates at 22 MHz, tunes the frequency of the master laser with a very high control bandwidth (hundreds of kHz) to keep the mode cleaner locked. This loop alone, however, could not prevent the frequency from drifting following drifts of the length of the IMC cavity. Thus, part of the beam is sent through a rigid reference cavity (RFC) which only resonates at a fixed frequency (since the length of the cavity is fixed). The RFC loop, which error signal operates at 6 MHz, adjusts the length of the Mode Cleaner until the RFC is locked (hence indirectly keeping the laser frequency stable as well, with an accuracy of one Hz of total rms fluctuation).

A brief note on coordinate systems is useful here. At Virgo, the direction along the beam is set as the z-axis. The vertical direction is the y-axis, and the horizontal direction perpendicular to the beam is the x-axis. Thus,

for a mirror whose normal vector is aligned with the z-axis, the possible angular motions of the mirror are as follows. Pitch corresponds to rotation about the x-axis. Yaw is a rotation about the y-axis. Roll is a rotation about the z-axis. In the terms used in the channels that monitor the control loops, misalignments in pitch, yaw, and roll are referred to as theta-x, theta-y, and theta-z.

Maintaining mirror alignment uses two separate sets of angular control loops at different times: automatic alignment (AA when the IMC is already locked) and local control (LC during lock acquisition).

Automatic alignment uses three wavefront-sensing quadrant diodes, i.e. 6 error signals, to control five degrees of freedom: pitch, yaw, and roll of the injection bench and pitch and yaw of the mode cleaner end mirror. The roll of the end mirror does not need to be controlled here, because the beam is of course symmetric about the z-axis. Misalignments of the mirrors correspond to a combination of shift and tilt of the beam on the diodes. The diode at the end mirror fixes the position with respect to ground (pointing control). The pair of wavefront-sensing diodes on the injection bench are a near field diode and a far field diode. The near field diode sees the wavefront as a plane, and so it can control beam tilt but is insensitive to shift. The far field diode sees the curve of the wavefront, and so it is insensitive to tilt but can control shift.
The combination of all three diodes controls the five rotational degree of freedom.

Local control uses optical levers, in which a red diode reflects (not at normal incidence) off a mirror on the optical bench and impinges on a position-sensing device (PSD). The position of the light incident on then PSD corresponds to the shift or tilt of the diode beam, which determines the alignment of the mirror. The IMC has three optical levers: two on the injection bench and one on the end mirror (see Figure 1). One optical lever on the injection bench controls pitch and yaw; the other controls roll. The optical lever at the end mirror controls pitch and yaw of the end mirror, and also measures longitudinal position (position along the z-axis, but not roll about the z-axis).

2 Preliminary Investigations

Problem events in the control loops, such as misalignments of the mirrors cause power drops in the IMC or even complete unlocks. A variety of mechanisms can cause these events and so it is useful to be able to search for connections between events in the IMC and many noise-monitoring channels.

Several preliminary investigations motivated the development of the automated systematic tool. For several months, sudden fast (below 1 millisecond) unlocks of the IMC had been a re-occuring problem. These unlocks were easily recognizable in trend data and the speed of the unlocks indicated an electronic origin. The unlocks only occurred during working hours, and so one suggested potential cause of the unlocks was the electronic relay in the doors of the central building, which could produce noise when a person swipes their badge to open a door. Another suggested cause was saturation of the Master Laser Piezo correction (ML PZT in short). Investigating these two possible causes illustrates two slightly different types of question for an automated search to answer. In the case of the door badges, we compare one set of discrete events to another. In the case of the ML PZT, we compare discrete events to continuous data.

2.1 Door Badging

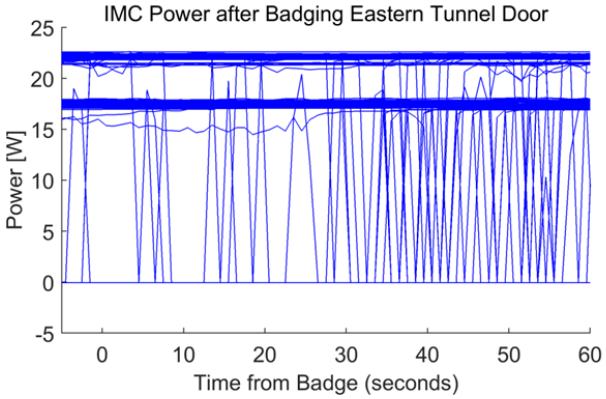

Figure 2: This plot shows the IMC power during the 60 seconds after each badge swipe at the "Eastern Accanto tunnel" door. The graphs for all badge swipes are overlaid. Note that the number of unlocks increases 30 seconds after the badge time.

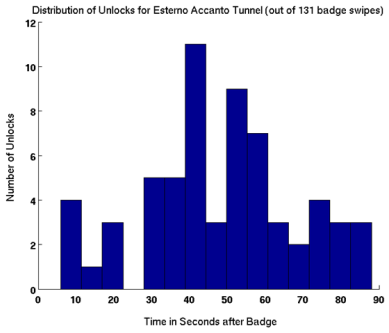

Figure 3: Histogram of the number of unlocks for each time interval after door-badging. Note the number of unlocks increases 30 seconds after the badge time.

If door badging or door relay switch on and off were correlated with IMC unlocks, we may expect to see unlocks clustered in a particular time interval after badge swipes. We checked for a correlation between door badge swipe times and IMC unlocks by examining the IMC power in a time window from 1 second before the badge to 10 seconds after, for each badge swipe in the time period 11 May 2015 to 5 June 2015 [3]. The time window starts a second before the door time because we use trend data, [1], which is sampled at one Hertz. The time window extends 10 seconds after the badge swipe to include the time for the door mechanism to re-arm. For each of the 15 doors, I computed the number of times that an IMC unlock occurred in the time window 1 second before to 10 seconds after a badge swipe, and compared this result to the total number of times that door was used while the IMC was locked. Badge swipes during times when the IMC was already unlocked were not included. Any discrepancy between door clock time and GPS time is estimated to be less than one second.

If door badging or door relay switch on and off were correlated with IMC unlocks, we would expect to see unlocks clustered in a particular time interval after the badge swipe. In order to compare any prospective correlation for many doors, I calculated the ratio of the number of times that an unlock occurred in the 10 second time-window near a badge swipe to the total number of badge swipes, for each door separately. A particularly large ratio for a specific door would suggest that that door may cause the unlocks. Our results do not suggest such a connection.

This ratio is less than 10% for all doors. The highest of these ratios occurred for the door "CUB022B Entrata Tunnel Ovest" (West Tunnel Entrance), at 9.1% (but note that this door was used only 11 times). There was no clustering of unlocks in the seconds immediately after badge swipes. Thus, analyses in this time range do not suggest badge swipes as a cause for the unlocks.

We note, however, that instances of unlocks increase around 30 seconds after a badge swipe at the "Esterno Accanto tunnel" door (see figures 2 and 3) . This door was badged 131 times and 63 instances of unlock occurred within 60 seconds of a badge swipe. We timed the entries of people into this room and note that a 30-second delay is approximately the time it takes a person to put on shoe covers and cross the room to a motion-activated light. We completed several tests for a connection between the lights and the unlocks and also

[1] The raw timeseries for each channel ("fast data"), typically sampled at 20kHz, is processed into mean, maximum, minimum, and rms "trend" data at 1 Hz.

tested for a connection to cell-phone noise with no significant results. By the time of those tests, an adjustment of the IMC control electronics by a different group had resolved the problem without identifying the cause.

2.2 Master Laser PZT saturation

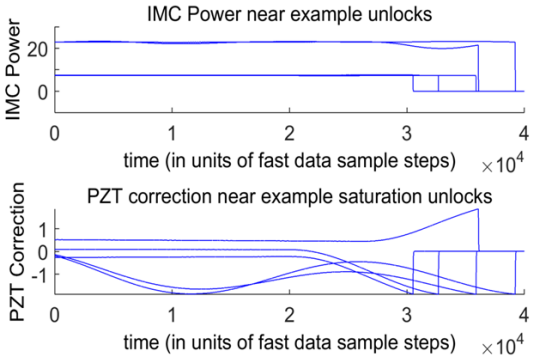

Figure 4: Plots of the Master Laser PZT Correction and the IMC Power in the 2 seconds surrounding an unlock, for five randomly selected example unlocks for which the PZT saturated. The five examples are overlaid. Note that the unlocks are fast, sharp drops.

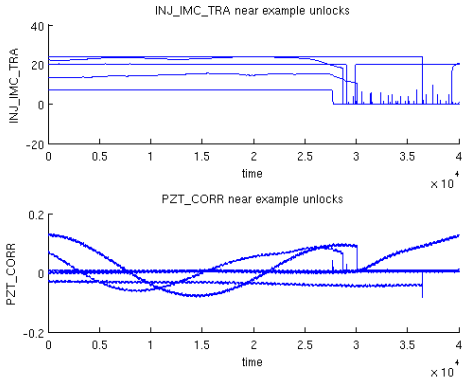

Figure 5: Plots of the Master Laser PZT Correction and the IMC Power in the 2 seconds surrounding an unlock, for five randomly selected example unlocks for which the PZT saturated. The five examples are overlaid. Note that the unlocks are fast, sharp drops.

Unlike the case of the doors, in which we searched for unlocks near a given set of discrete events, in the case of the ML PZT I created an automated search for all the unlocks in 9 weeks of trend data (sampled at 1 Hz) . Having found the event times in trend data, my script then searched fast data (sampled at 20kHz) around each event time to precisely locate the unlocks, and looked at the ML PZT behaviour in fast data just before the unlock. Since the ML PZT correction channel saturates at approximately plus or minus 2 volts, an unlock would occur if the ML PZT approaches that level.

Figures 4 shows examples (in fast data) of unlocks where the absolute value of the ML PZT correction exceeded 1.8 volts. Note that the IMC fast unlocks are sudden, sharp drops in power and the corresponding sharp drops in the ML PZT correction occur at the ML PZT saturation values (because it resets after saturating). For comparison figure 5 shows the PZT behaviour during (randomly selected) example events when the IMC unlocked but the ML PZT did not saturate. Note that the ML PZT drops do not occur at the maximum values of these ML PZT curves.

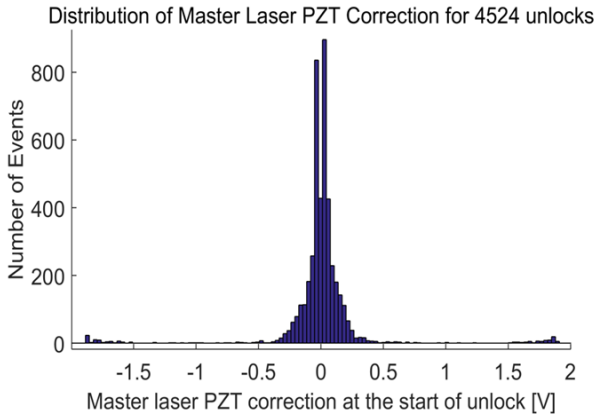

Figure 6: Distribution of ML PZT correction during fast IMC unlocks. Note the small peaks in the tails.

In this analysis we want to estimate what percentage of unlocks were due to ML PZT correction saturation. Figure 6 shows the distribution of ML PZT correction voltage in the instant before each unlock. Note that the distribution is strongly concentrated at low voltages, because the ML PZT correction is kept close to zero by the RFC loop (refer back to section 1.1 if needed) and thermal corrections. The tails of the distribution, however, are not monotonic but rather have small peaks near plus or minus 1.8 volts. 1.6 % of the unlocks (73 out of 4524 unlocks) are associated with ML PZT correction absolute value exceeding 1.8 Volts (i.e. saturation). We conclude that ML PZT correction saturation is a rare cause of unlocks.

3 Multipurpose Systematic Approach

The ML PZT and door-badging investigations of the sudden unlocks motivate the development of new multipurpose systematic statistically-rigorous means of search for connections between events of interest and prospective causes. The tool I developed, PROSPECTOR, accomplishes this analysis in the following basic components:
Inputs:
1.Define events in query channel.
2.Choose time to search.
3.Define conditions to exclude.
4.Choose channels to search.
Results:
PROSPECTOR returns a ranked list of the channels with the strongest prospective connection to the query events.
In the following sections, I use an example application to explain in detail how PROSPECTOR works, then I discuss its application to generalized problems, and then I discuss a few further investigations using PROSPECTOR.

3.1 Example application: IMC misalignments caused by chiller

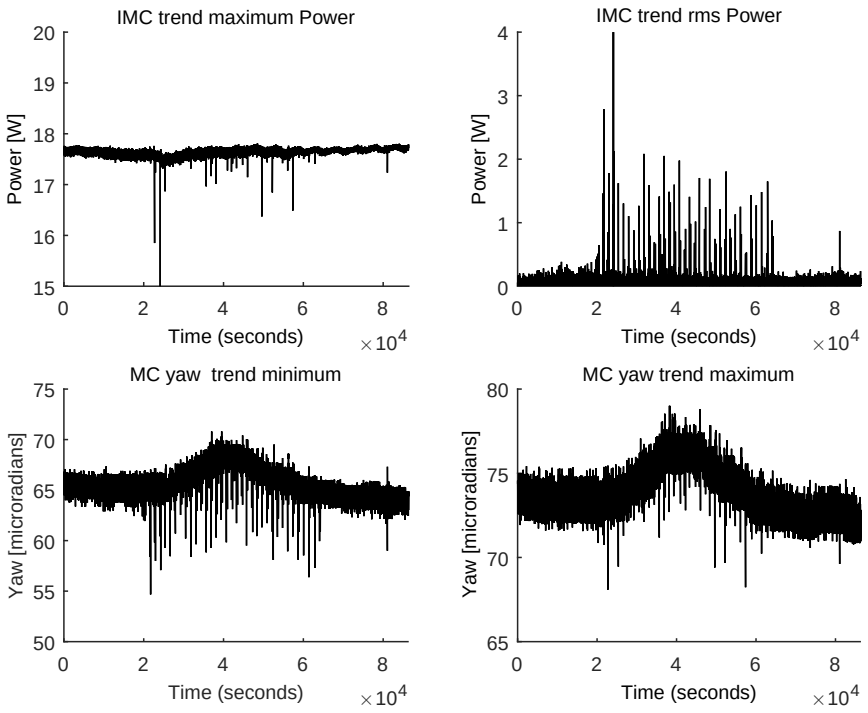

Figure 7: **MC mirror misalignments on 13 June.**Clockwise from upper left: Trend maximum power in the IMC. Trend rms power in the IMC. Trend maximum mirror yaw in the Mode Cleaner end mirror. Trend minimum mirror yaw in the mode cleaner end mirror. All plots are for the entire day of 13 June.

Brief mirror misalignments in the MC mirror, especially in the yaw (theta-y) direction, have been a known problem and provide an example problem for the PROSPECTOR noise-hunting tool. Drops in IMC power by one to a few Watts (out of a total 17 W) and spikes of a few microradians in mode cleaner end mirror yaw indicate the misalignment events. These events tend to occur during day-time hours and occur regularly with a period of 21 minutes. Before completing PROSPECTOR, by comparing channels by eye, I found that the IMC misalignments coincide with periodic peaks in the interruptible power supply mains current (IPS) monitors, caused every 21 minutes by the start of a water-chiller in the Mode Cleaner Building air-conditioning system.

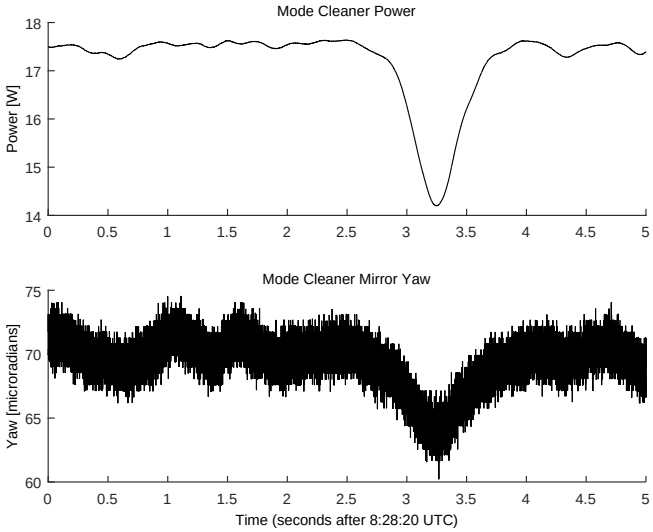

Figure 8: **Example event shown in fast data on 13 June 2015 at 8:28 UTC.** *Upper panel:* IMC power in fast data. *Lower panel:* Mode Cleaner mirror yaw (θ_y).

As an example of these misalignments, Figure 7 displays the IMC and mode cleaner theta-y in trend data for the entire day of 13 June. Figure 8 shows the shape of the drops in fast data. Figure 9 displays the IMC rms power, the IPS rms current, and the air-conditioning inflow water temperature over the same day. Note the periodicity. Figure 10 presents the same channels from 8:00 - 9:00 13 June (UTC). The red vertical lines mark the three IMC power drops. Note that these coincide exactly with the sudden current increases, and that typically smaller IMC drops occur when the current decreases. Note that the events only occur in the central part of the day, corresponding to a time of larger spikes in the IPS current. This is because two chillers operate during the hottest part of the day, and only the combined current spikes cause the problem.

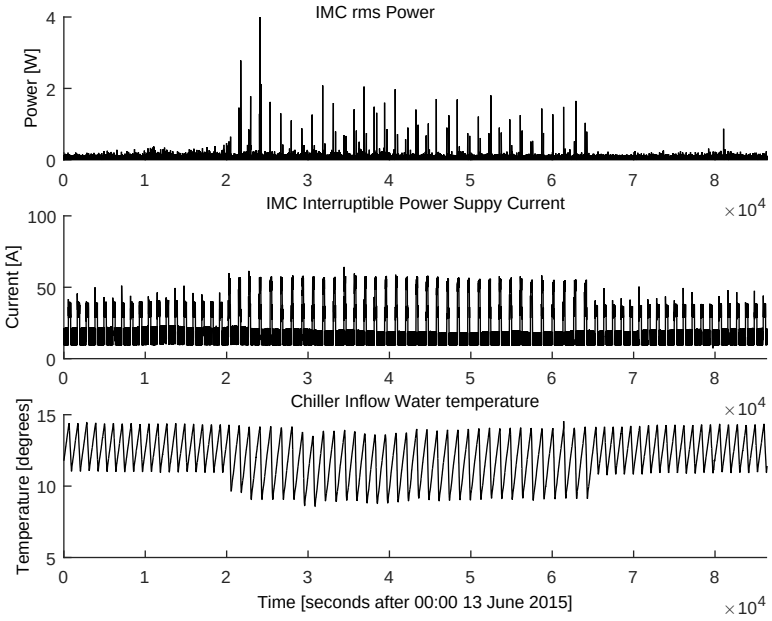

Figure 9: **Correlation between chiller and IMC misalignments, shown for 13 June 2015.** *top panel:* Mode Cleaner power in trend data. *middle panel:* mode cleaner building interruptible power supply current *lower panel:* mode cleaner building chiller water inflow temperature

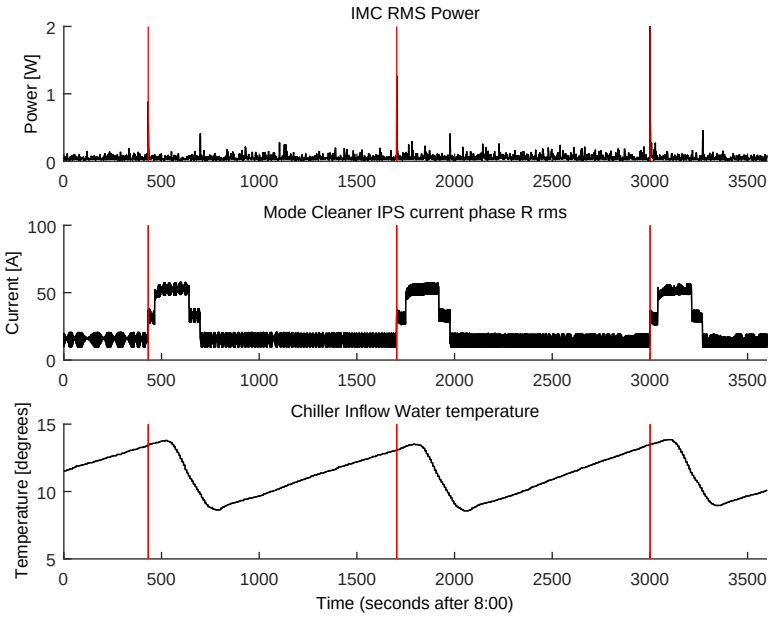

Figure 10: **Correlation between chiller and IMC misalignments, shown for 8:00 - 9:00 UTC, 13 June 2015.** Red lines mark the times of misalignment. *top panel:* Mode Cleaner power in trend data. *middle panel:* Mode cleaner building interruptible power supply current. *lower panel:* Mode cleaner building chiller water inflow temperature.

These drops are an ideal noise hunting problem for PROSPECTOR code. The goal of the PROSPECTOR code is to take events of interest, such as the drops in the IMC, systematically search a large number of channels, and identify the channels (in this case IPS current monitors) with the strongest prospective connection to the query events.
In the case of the chiller problem, I define the events as a drop in IMC power of more than 1 Watt and less than 10 Watts below the local mean power (the second criterion distinguishes drops from unlocks), which must correspond to a misalignment of at least 5 microradians in theta-y (mirror yaw). PROSPECTOR searches for all such events in the chosen time range (could be a day or many weeks) and then looks at the trend data in the chosen auxiliary channels (in this case all the mode cleaner building environmental channels) in the chosen time window. For each environmental channel, PROSPECTOR computes $state \equiv \sqrt{max^2 + min^2}$, where max and min refer to the trend maximum and trend minimum of the environmental channel. I use this quantity, rather than only trend minimum or maximum to efficiently make PROSPECTOR sensitive to either spikes or drops or a combination of the two (which could be lost in the trend mean).
Before further analysis, PROSPECTOR filters out times of unwanted conditions, filtering both the set of event times and the entire time range of interest. In this case, unwanted conditions are times when the IMC was already unlocked, and times when it was under automatic alignment (the events only occur during local control).
The fundamental statistical question to answer (for each channel) is whether the set of channel states at the event times is different from the state at a randomly selected subset of times during the overall time window. Thus, we must compare the distribution of channel state at event times to the overall distribution of channel state, for each channel. PROSPECTOR computes the empirical cumulative distributions for the channel state at event times and for the overall state, and compares the two distributions by computing the maximum difference between them. In other words, if x is the overall channel state, let $\{f(x)\}$ be the overall empirical cumulative distribution of the channel state. Let $\{g(x)\}$ be the empirical cumulative distribution of the channel state at events, interpolated to cover the same domain, x, as does f. Then the parameter of interest is the maximum of the $\{f(x) - g(x)\}$. This maximum difference is the parameter used to assess each environmental channel's prospective connection to the events of interest. PROSPECTOR returns a ranked list of environmental channels, and each channel's maximum difference.

Depending on the type (and rapidity) of the query events, it can be most appropriate to look at the environmental channels with a small offset from the event time, and such an offset can be built into the event definition in PROSPECTOR. In the case of the chiller, PROSPECTOR looks at the environmental channels 1 second before the misalignment itself. This offset is chosen because the misalignment is a gradual rather than sudden event (see figure 8).

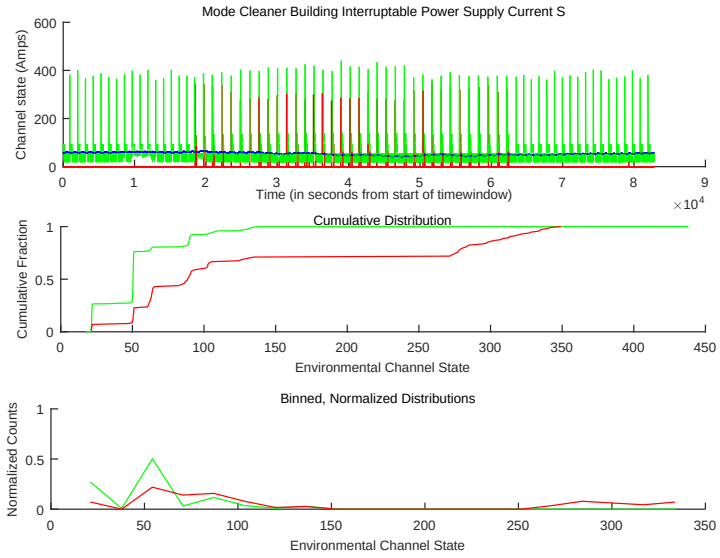

Figure 11: **PROSPECTOR results for mode cleaner IPS channel**. *top panel:* Green plots the mode cleaner IPS channel state during the entire day of June 13. Red plots the IPS channel state one second before event times. Blue marks the channel mean state during all times (including times that are filtered out for the analysis *middle panel:* Red and green plot cumulative distributions of the IPS channel state during event times and overall, respectively. *lower panel* Binned, normalized distributions for event times and overall, in red and green, respectively. Note that this plot is sensitive to choice of bins, and so the cumulative distributions in the middle panel are more informative.

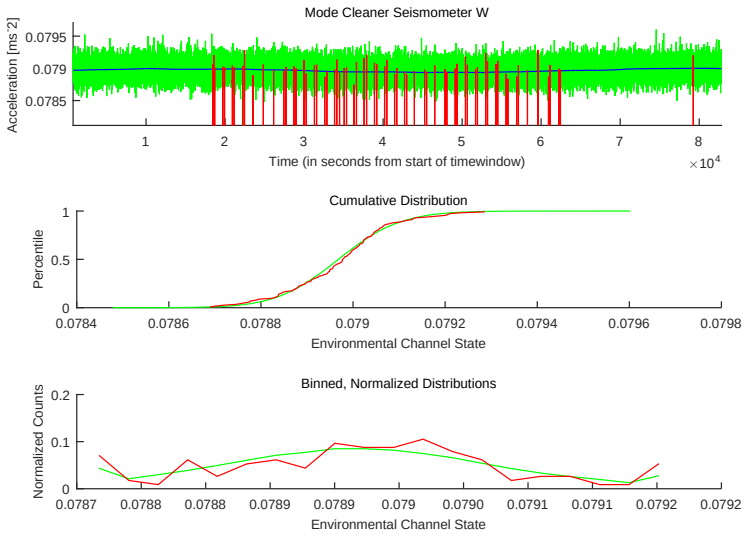

Figure 12: **PROSPECTOR results for a seismometer**. *top panel:* Green plots the mode cleaner west-oriented seismometer channel state during the entire day of June 13. Red plots the seismometer channel state one second before event times. Blue marks the channel mean state. *middle panel:* Red and green plot cumulative distributions of the west-oriented seismometer state during event times and overall, respectively. *lower panel* Binned, normalized distributions for event times and overall, in red and green, respectively. Note that this plot is sensitive to choice of bins, and so the cumulative distributions in the middle panel are more informative.

Figure 11 shows the results for an IPS channel which PROSPECTOR gave a high ranking as a prospective cause of the misalignments, as it should. The chiller draws from this interruptible power supply when it turns on, and so this channel identifies the correlation to the chiller. The middle panel of 11 shows the two cumulative distributions for this power supply. The maximum gap between the two cumulative distributions is 0.53. That is, the 75th percentile of the overall current data occurs at a lower amplitude than does the 22nd percentile of the current data at the event times. Hence, there is a difference of 53% in the cumulative distributions. This large difference suggests that the current data at event times are not the same as a randomly selected subset of the overall current data. Thus, the script returns this channel (rather than seismometers, microphone, etc.) as a prospective connection to the misalignments. The top panel of figure 11 clearly illustrates that misalignments tend to occur when the chiller turns on. For comparison, Figure 12 shows the results for a channel to which PROSPECTOR gave a very low ranking. For this channel, the maximum difference between the distributions is 0.06. Note that there are only small differences between the empirical cumulative distributions in figure 12.

3.2 Overview of the generalized version of the code

Defining Query Events PROSPECTOR must be given a specific definition of the events of interest. In the current version of the code, these events must be well-defined in trend data. Since PROSPECTOR is designed to search months of data, and to compare channels at event times to the overall channel behaviour during the months searched, the restriction to trend data is necessary to keep the computation to manageable runtimes. Clearly, PROSPECTOR will give the best results if the query events have a single cause. If the events are loosely defined, the search will likely return events with connections to many channels and so it is unlikely that any single channel will surface as a strong connection. This situation may be avoided by choosing stricter criteria to define the event of interest and by dividing a broad problem into well-defined components. For example, it is better to define a search for drops in IMC power that coincide in theta y mirror misalignments rather than to search for all IMC drops in the same query.

Choosing a timewindow to search and conditions to filter out As with any statistical tool, PROSPECTOR gives the most meaningful statistics for a large sample size of events. One of the chief advantages of a brute-force systematic tool is the ability to analyse hundreds or thousands of events. With an appropriately defined query, it is possible to collect ALL the occurrences of an event of interest and compare as many channels as you care to specify. The larger the time window, the more events will be found. However, the time to load data is the largest part of PROSPECTOR's run time, and so some restriction of the the time range to search is necessary. If the events of interest are known to have been a problem during a particular week, it is best to restrict the search to that week of data. Furthermore, if the search time window includes periods in which events occurred for multiple reasons, the results will be weaker. For example, searching for IMC misalignments over a timespan that included the chiller problems and also included a period of road work outside the mode cleaner returned weak results. Dividing the search into two separate time windows returned a strong result for the IPS when the chiller was the most significant problem and returns strong results for seismometers during the road work (see section 3.4).

Specifying additional filter criteria is another effective way to fine-tune the search. For events in the IMC, the code imposes a filter to exclude times when the IMC was already unlocked. Furthermore, the brief mirror misalignments only occur if the mode cleaner is under local control. Hence, we should not compare event times to times when the IMC was under automatic alignment. A simple filter in the script excludes automatic alignment. PROSPECTOR is designed to easily incorporate as many filters as desired, in order to tailor the query to a specific question.

Interpreting PROSPECTOR's results PROSPECTOR returns a list of channels, ranked by the maximum difference between the cumulative distribution of the channel state at event times compared to the distribution

overall. A significant difference in distributions does not necessarily imply a causal relationship. Some events produce effects that are detected in various environmental channels. For example, a UPS (uninterruptable power supply current monitor) channel in the mode cleaner actually ranks highest in the chiller problem, but examining fast (20kHz) data shows that the UPS increase occurs after the start of the misalignment, and the θ_y actuator causes these current draws. The purpose of PROSPECTOR is to find patterns among events in different channels, and in this purpose it has been successful. Establishment of cause versus effect is a separate question which typically requires checking the fast data. In the case of the brief IMC misalignments, the water temperature plots along with the IPS current indicate the chiller as the cause. Thus, prospector gives useful hints for noise hunting.

3.3 Magnetic Injection tests

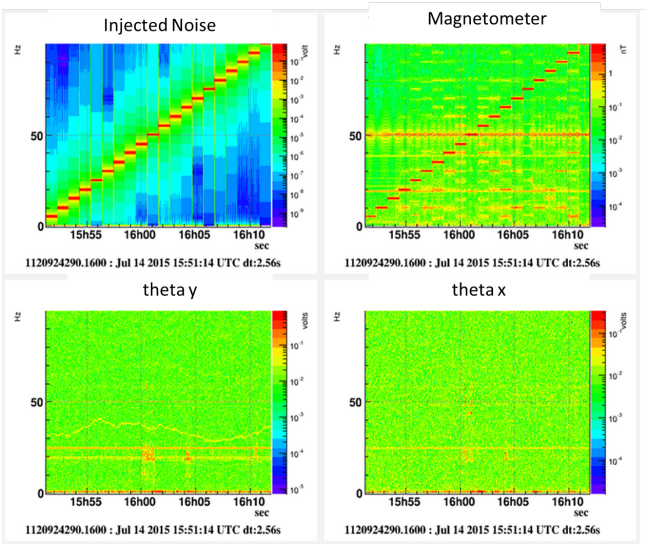

Figure 13: **Spectrograms of line injection frequency sweep.** *Top left:* Spectrogram of injected noise. *Top right:* Magnetometer reading. Lower left: mirror yaw. Lower right: mirror pitch.

Having established the chiller as the cause of the unlocks, the precise coupling mechanism between the events remains unclear. Magnetic noise from the power cables to the chiller is a possible coupling mechanism. We investigated this possibility by injecting magnetic noise into the mode cleaner [4]. We performed these tests by placing a large (1 m diameter) circular coil near the mode cleaner optical bench, and sending current through the coil in patterns controlled by python scripts. We ran several types of tests. First we injected lines of single-frequency sinusoidal noise, which swept from 10 Hz to 100 Hz in steps of 5 Hz, spending one minute at each frequency. I also injected constant low frequency noise at 350 MHz, and finally step function noise. Figure 13 presents spectrograms of the results for the line injection sweep. Note that the injected noise pattern (upper left spectrogram) is unambiguously detected by the mode cleaner building magnetometer (upper right spectrogram). However, we find no misalignments in pitch or yaw due to the injection. The other noise patterns produced similar results. Furthermore, PROSPECTOR gives a low ranking to the magnetometer for the time period of the magnetic injections. We conclude that these magnetic injections have not produced IMC misalignments, and that these tests have been a useful check of PROSPECTOR. We do not, however eliminate magnetic noise as the cause of the unlocks. It is possible that magnetic noise from the chiller causes problems at a different frequency or by means of a different path that was not covered by the magnetic injections.

3.4 Seismic Noise Hunting

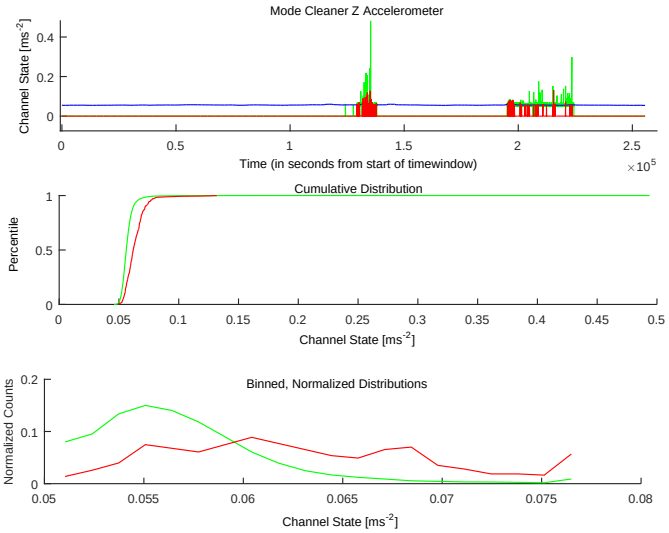

Figure 14: **PROSPECTOR results for an accelerometer**. *top panel:* Green plots the mode cleaner accelerometer channel state during a period that includes roadwork in July. Red plots the accelerometer state one second before event times. Blue marks the channel mean state during all times (including times that are filtered out for the analysis *middle panel:* Red and green plot cumulative distributions of the accelerometer channel state during event times and overall, respectively. *lower panel* Binned, normalized distributions for event times and overall, in red and green, respectively. Note that this plot is sensitive to choice of bins, and so the cumulative distributions in the middle panel are more informative.

As another demonstration that PROSPECTOR works, I ran the same misalignment search on a period of time in which road work -causing intense ground vibrations- was being done just outside the mode cleaner. The search specifications were the same as for the chiller problem. In this period, PROSPECTOR ranked seismometers and accelerometers higher than the IPS current. Most of the IMC power drops during that period were due to road work. Figure 14 illustrates the results using the mode cleaner accelerometer oriented along the z axis. In the upper panel, the times when road work was underway are the periods of increased seismic activity. The times with no green or red plotted are the times when the cavity was under automatic alignment and thus have been filtered out of the analysis.
Note that it would not be very meaningful to conclude that events occur at times of seismic activity when the search window is a period of generally high seismic activity. This is not what PROSPECTOR does. In this search window, PROSPECTOR is comparing the channel state at event times to the overall channel state during the days with roadwork. Thus, by giving top rankings to seismic channels, PROSPECTOR indicates that the misalignments occur at instants when seismic activity is larger than the already large overall state.

3.5 Use of an Environmental Channel as the Query Channel

The applications of PROSPECTOR that I've discussed so far all use events in the IMC as the query. This need not be the case. Events can be defined in any channel and PROSPECTOR will compare the behaviour of auxiliary channels at those event times. Rather than starting with a problem in the IMC and searching for prospective causes in environmental channels, it is just as possible to start with a set of environmental events and search for corresponding events in mode cleaner channels. In other words, rather than just searching for causes of problems, one could just as easily use PROSPECTOR to test whether particular noise events are problematic or benign.

Furthermore, one may use PROSPECTOR to search for corresponding events among different channels. This type of test can be useful when a noise event has been identified as a problem but the coupling mechanism remains to be found. As a test of this method, I defined the query events as the IPS spikes associated with the chiller. I then used PROSPECTOR to compare to all the mode cleaner environmental channels. PROSPECTOR gave a high ranking to the mode cleaner magnetometer (see figure 15). These results indicate that the chiller is indeed associated with magnetic noise. Magnetic noise could be part of the coupling mechanism in ways that the magnetic injections did not test. Further magnetic injections could test different coil orientations to more exhaustively probe any potential magnetic coupling.

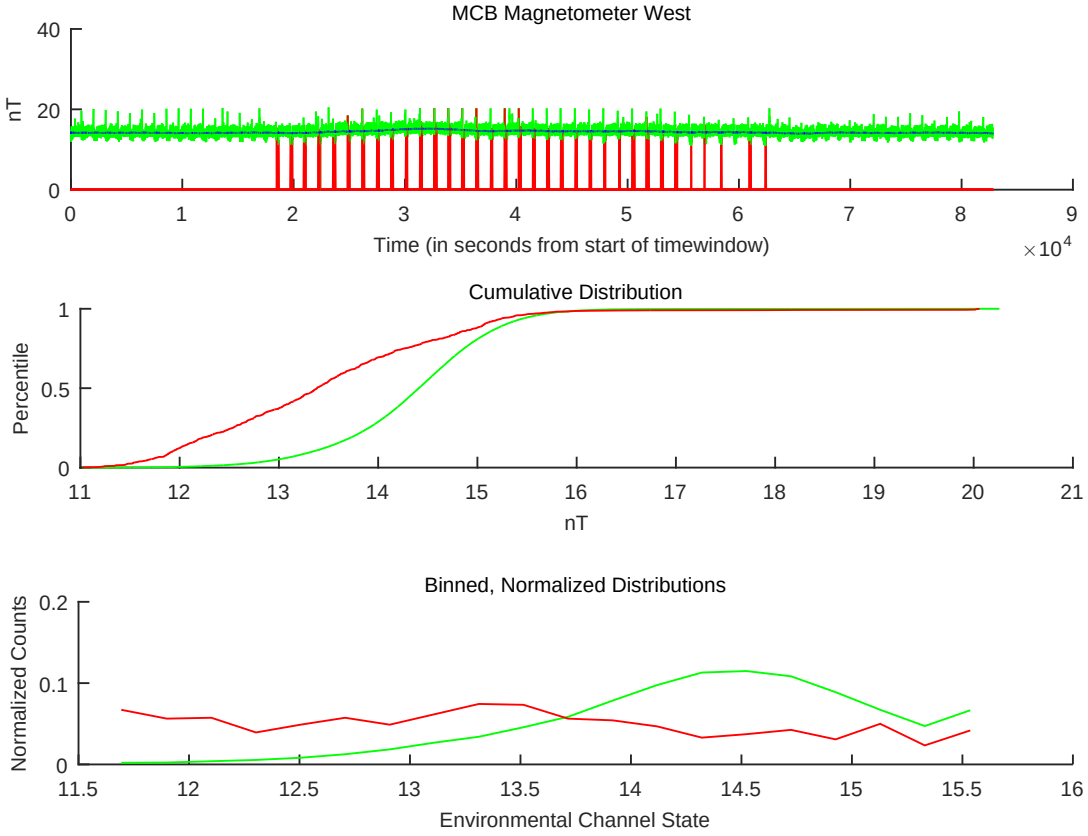

Figure 15: **PROSPECTOR results for the mode cleaner building magnetometer**. *top panel:* Green plots the magnetometer channel state during the entire day of June 13. Red plots the mode cleaner building magnetometer state one second before event times. Blue marks the channel mean state during all times (including times that are filtered out for the analysis *middle panel:* Red and green plot cumulative distributions of the mode cleaner building magnetometer channel state during event times and overall, respectively. *lower panel* Binned, normalized distributions for event times and overall, in red and green, respectively. Note that this plot is sensitive to choice of bins, and so the cumulative distributions in the middle panel are more informative.

4 Conclusions

In conclusion, PROSPECTOR has worked well for finding corresponding events in many channels. Query events may be defined in any channel, allowing PROSPECTOR to help answer a variety of questions. Cross checks between PROSPECTOR and EXCAVATOR [5] (another brute-force noise hunting tool) give confirm PROSPECTOR's results. It is important to emphasize that since PROSPECTOR uses trend data, it can search long periods of time. This ability for broad-ranging scans makes it an effective tool for noise hunting.

Future Directions In its current form, the PROSPECTOR code must call externally-defined functions to define the events of interest and to define the filters against unwanted conditions. This aspect customizes the search to a given problem while keeping the noise-hunting tool generalized to a breadth of problems. The customization step does, however, require user effort and is responsible for the bulk of PROSPECTOR's runtime. I would propose streamlining this process to allow the user to write a formal query, perhaps using a language such as SQL (Structured Query Language). Consider if each Virgo trend channel were a field in a data base table, and each second of time a row. For the types of noise hunting questions that PROSPECTOR answers, we are thus interested in collecting the all rows for which certain fields lie in a particular range of values. Structured queries would be a natural formalism for this approach.

The PROSPECTOR script will be available soon in the *VirgoDev* directory.

Acknowledgements

I am very grateful to the European Gravitational Observatory, the INFN, the NSF, and the University of Florida for the opportunity to do this work. I am tremendously grateful to Dr. Irene Fiori and Dr. Maddalena Mantovani for being wonderful advisors. I would also like to thank Federico Paoletti and Bas Swinkels for all their help and advice, and to thank everyone at Virgo for everything they taught me.

References

[1] Virgo logbook entry $https://tds.ego-gw.it/itf/osl_virgo/index.php?callRep=32486$ 2

[2] The Virgo Collaboration, 'Advanced Virgo technical design report' Virgo Note VIR-0128A-12 (2012), $https://tds.ego-gw.it/ql/?c=8940$. 3

[3] Virgo logbook entry $https://tds.ego-gw.it/itf/osl_virgo/index.php?callRep=32403$ 5

[4] Virgo logbook entry $https://tds.ego-gw.it/itf/osl_virgo/index.php?callRep=32505$ 14

[5] B.L. Swinkels, EXCAVATOR, a new tool for finding correlation between triggers and auxiliary channels, presentation at LIGO-Virgo meeting, Maryland (2013), available at https://tds.ego-gw.it/ql/?c=9772 17