ACCENT REALIZATION IN SPONTANEOUS SPEECH: GREEK H* AND L+H*

Stella Gryllia, Jiseung Kim, Riccardo Orrico, Na Hu, Amalia Arvaniti

Radboud Universiteit, Netherlands
stella.gryllia@ru.nl, jiseung.kim@ru.nl, orrico.riccardo@ru.nl, na.hu@ru.nl, amalia.arvaniti@ru.nl

ABSTRACT

Greek H* and L+H* contrast, such that the former pitch accent is realized as high pitch indicating new information, while the latter is realized as a "scooped" rise encoding contrastive information. This distinction is robust in controlled, scripted data but has not been tested in unscripted speech. To address this gap, we used Functional Principal Components Analysis (FPCA) to examine 1599 rising accents from unscripted speech elicited from eight native Greek speakers. The coefficients of the resulting PCs were analysed using LMEMs, which showed statistically significant differences between H* and L+H* for PCs 1-4. PC1 captured scaling differences, while PCs 2-4 reflected differences in contour shape, such that they support the analysis of L+H* as a "scooped" rise, and H* as high pitch. These results suggest that the distinction between H* and L+H* is relevant for Greek intonation phonology independently of speech type.

Keywords: intonation, variability, spontaneous speech, Greek, FPCA.

1. INTRODUCTION

English and other Germanic languages have a H*~L+H* accentual contrast [e.g., 1, 2, 3]. In English, H* indicates that the accented item is new in discourse and should become part of the common ground, while L+H* indicates it should be chosen from a closed set of alternatives and added to the common ground. Thus, H* and L+H* typically mark broad and narrow focus respectively. This contrast is robust in perception [4, 5], and in production with controlled, scripted data where contrastivity is critical [3, 6]. However, spontaneous speech has not provided equally strong evidence: [7] found that rising and high accents (L+H* and H* respectively) are present in spontaneous speech but this phonetic difference does not correspond to distinct pragmatic functions. In short, in spontaneous speech, English H* and L+H* are not as distinct as in scripted speech, shedding doubt on the validity of the contrast.

An analogous situation obtains in Greek, in which the descriptions of H* and L+H* are comparable to those for English in terms of phonetic realization and pragmatics [8, 9]. Based largely on qualitative data, these accents have been described as appearing in focal position followed by L-L% edge tones. In this context, H* creates a slightly declining plateau with the preceding accentual peak (if one is present) and is realized as a fall (Fig.1a); L+H* is realized as a rise from a low F0 point followed by a fall (Fig.1b).

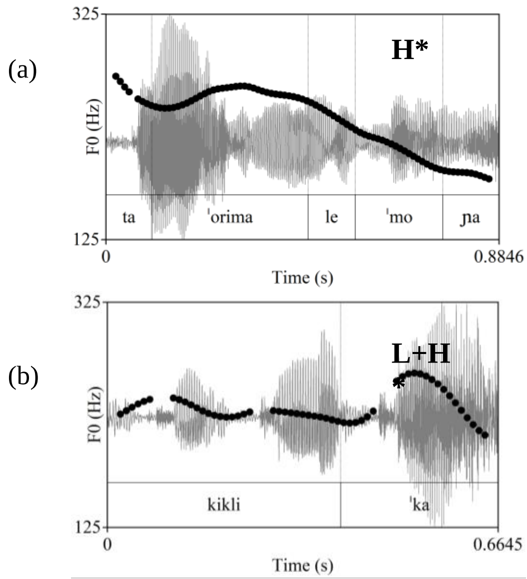

Figure 1: Waveforms and F0 tracks of [ta ˈorima leˈmoɲa] "the ripe lemons" with a H* accent on [leˈmoɲa] (a), and [kikliˈka] "in a circle" with a L+H* accent (b).

Recently, [10] applied Functional Principal Components Analysis (FPCA) on controlled scripted data to quantitatively examine this accentual contrast in Greek. FPCA is a data-driven, dimension-reduction method in which curves (time- and speaker-normalized F0 curves in [10]), are modelled as B-splines. Based on this modelling, FPCA returns the dominant modes of curve variation called Principal Components (PCs; [11, 12]; see 2.3 for details). In line with [8, 9], [10] showed that in Greek H* and L+H* differ in shape and scaling and are used in distinct pragmatic contexts. However, this quantitative evidence comes from scripted data only. Given the findings from English [7], it is reasonable to question whether the Greek accentual distinction is as robust in unscripted speech as it is in scripted speech [8, 9, 10] . We addressed this question by examining the H*~L+H* contrast in unscripted Greek data using FPCA.

2. METHODS

2.1. Participants

Eight native speakers of Standard Greek (4F), 22-27 years old ($\bar{X}$ = 24; SD = 1.7), who reported no speech or hearing disorders, took part. APF01 was bilingual in Greek (dominant) and Albanian (heritage); her data were included as they did not deviate from the average. All participants had learned English and additional languages via formal instruction (Italian, N = 3; French, N = 3; Spanish, N = 2; German, N = 1).

2.2. Materials and procedure

Unscripted speech was elicited by means of five tasks. In three monologues, participants narrated from memory a fable and a news item they had previously read, and told three stories using the app "Story Dice", in which players create a narrative based on icons depicted on six dice. Map tasks were used to elicit dialogues [13]: each speaker took part twice, once as instruction-giver (guiding the follower by using a map marked with a route) and once as instruction-follower (recreating the same route on their map which differed slightly from the instruction-giver's). In addition, participants discussed in pairs the functions of seven unusual objects appearing in a short video. Finally, the participants used the app "Guess who you are" (available in Greek) in groups of four to play an e-version of HedBanz (players place a card on a band on their head and guess the object on the card asking only yes/no questions).

Due to COVID-19 restrictions, the recordings were made using the experimenter's computer and the AVR application on the participants' phones. The recordings were saved as mono .wav files (256 kbps, 44.1 kHz sampling rate), a format that, according to [14], does not pose problems for F0 analysis.

2.3. Annotation and analysis

The data were annotated in Praat [15] after being orthographically transcribed. Trained annotators identified the nuclear accents of intermediate and intonational phrases that ended in L- and L-L% edge tones respectively. The annotators then used phonetic and pragmatic criteria to classify the accents (if high or rising) as H* or L+H*: accents were annotated as L+H* if they showed a rise from a low F0 point and their use was contrastive or corrective in context; high accents without a rise and accents that were not contrastive or corrective were annotated as H*. Finally, the annotators marked the *analysis window*, the interval over which the F0 analysis was to be performed; this included the accented word and its clitics, if any (e.g., του δεινόσαυρου [tu ðiˈnosavru] "the dinosaur.GEN" was one analysis window). Forced alignment was then performed using Praat's "Align interval" function with the language set to Greek, and erroneous segment boundaries were manually corrected. The resulting corpus included 1599 accents categorized as H* or L+H*; see Table 1.

	Speaker								
	01	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**Total**
H*	146	87	62	78	204	148	280	224	1229
L+H*	48	50	19	22	68	53	61	49	370
Total	194	137	81	100	272	201	341	273	1599

Table 1: Accent distribution by speaker.

The F0 contours of the analysis windows were extracted using Praat with an octave cost of 0.1 and the following pitch range and time steps (chosen to minimize tracking errors): APF01, APF05, APF07, APF08: 120–500 Hz, 5 ms time step; APM02, APM06: 70–280 Hz, 10 ms time step; APM03, APM04: 70–300 Hz, 10 ms time step. Remaining F0 halvings and doublings were visually identified [16] and manually corrected. Undefined F0 values were interpolated using stine interpolation [17]. The F0 contours were then scaled by speaker.

The next step was *landmark registration* to align the F0 curves around a specific time point and time-normalize them in the process. The onset of the accented vowel was used as landmark. The speaker-normalized and landmark-registered curves were then smoothed using k = 8 and λ = 10^4 (k refers to the number of spline knots and reflects the degree to which the original curves are faithfully modelled; λ reflects the degree to which smoothing is penalized).

The resulting F0 curves were subjected to FPCA following the procedures in [11, 12]. As mentioned in section 1, FPCA returns the dominant modes of curve variation (PCs). The shape of each F0 curve is treated as a function and receives a coefficient (or *score*) for each PC. Scores can then be used for statistical modelling to uncover differences between linguistic categories, here H* and L+H*.

The scores of the first five PCs were statistically analysed using linear mixed effects models (LMEMs; R: 4.2.1 [18]; lme4: 1.1.26 [19]). In all models, the PC scores were the dependent variable, and ACCENT (H*, L+H*) the independent variable. SPEAKER was included as random intercept and ITEM was nested within SPEAKER. This structure was selected because the data come from unscripted speech and thus each *item* was given a unique id (e.g., when APF01 says [ˈperasa ta ˈorima leˈmoɲa] "I went past the ripe lemons" *lemons* is item 01; when she says [ðeksiˈa ap ta ˈorima leˈmoɲa] "to the right of the ripe lemons" *lemons* is item 02).[1] The base model formula is given in (1); for PC1, R indicated that the model was "singular", so we simplified the random structure by dropping the random intercept for SPEAKER.

(1) lmer (PC# ~ ACCENT + (1|SPEAKER) + (1|SPEAKER: ITEM), data = data, REML = FALSE)

3. RESULTS

Data and analyses are available at osf.io/kca26. Fig.2 shows the first five PCs, which capture 98.2% of curve variability in this dataset. The black line in each panel represents the smoothed and normalized average F0 curve and is the same in all five panels. The red curves show the changes to this average F0 curve when the coefficient of the depicted PC is up to +1 standard deviation above the mean; the blue curves show the same changes when the coefficient is up to -1 standard deviation. Table 2 presents the output of LMEMs and, for each model, marginal R^2 (an estimate of the proportion of variance explained by fixed factors, here ACCENT) and conditional R^2 (an estimate of the proportion of variance explained by the combined effect of fixed and random factors, here ACCENT, SPEAKER, and ITEM).

Fig.2 shows that PC1 captured 76.4% of the curve variability and that this variability was related to F0 scaling, with higher scores (red curves) leading to higher scaling relative to lower scores (blue curves). As shown in Table 2, PC1 scores for L+H* were significantly higher relative to H* (see also Fig.3a). This suggests that L+H* was scaled higher than H*. Conditional R^2 indicates that 27% of the variance was due to the combined effects of ACCENT, SPEAKER, and ITEM, while marginal R^2 indicates that only 3% of the variance in PC1 scores was related to ACCENT.

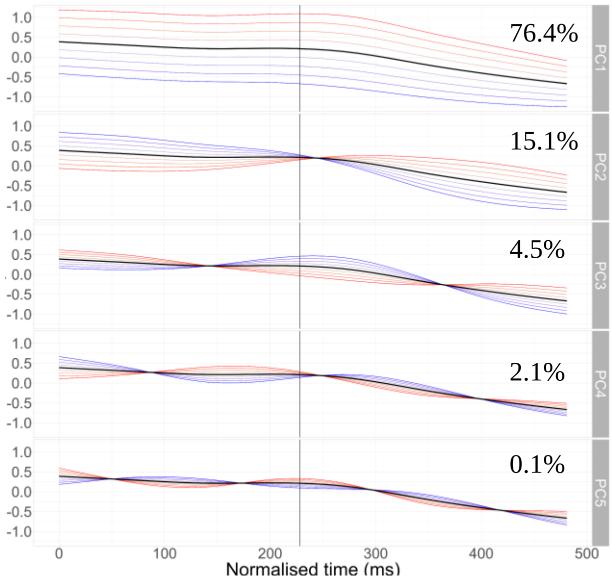

Figure 2: PC1-PC5 of the corpus; the vertical line indicates the onset of the accented syllable; percentages indicate variability captured per PC. For details, see text.

PC2 captured 15.1% of curve variability in the corpus. Fig. 2 suggests that this variability was related to F0 shape: higher PC2 scores resulted in a slightly "scooped" rise-fall, while lower scores resulted in a plateau-fall. The fact that PC2 scores were significantly higher for L+H* than H* (see Table 2 and Fig.3b), suggests that L+H* had a slightly rise-fall shape, while H* was realized as a plateau-fall. The conditional R^2 indicated that 68% of the variance was due to the combined effect of ACCENT, SPEAKER, and ITEM, while the marginal R^2 indicated that 10% of the PC2 variance was due to ACCENT.

PC3 and PC4 captured together 6.6% of curve variability. Fig.2 suggests that this variability was related to curve shape: higher PC3 and PC4 scores resulted in a fall, while lower scores resulted in a rise-fall. PC3 and PC4 scores for L+H* were significantly lower than those for H* (see Table 2 and Fig. 3c and 3d), indicating a difference between a rise-fall vs. a fall respectively. Conditional R^2 indicated that 54% and 64% of the variance (for PC3 and PC4 respectively) was due to the combined effect of ACCENT, SPEAKER, and ITEM, while marginal R^2 showed that just 2% of the variance in PC3 and PC4 was due to ACCENT.

Recapitulating, PC2, PC3, and PC4 captured differences between a rise-fall and a plateau-fall, with the differences among the three PCs relating mostly to the location of the peak in the rise-fall and the beginning of the fall in the plateau. These events can appear before or after the onset of the accented vowel (shown as the vertical black line in Fig.2).

PC1	**Est.**	**SE**	**df**	**t-value**	**Pr(>|t|)**
(Intercept)	-1.31	0.58	1024.9	-2.25	*
Accent: L+H*	7.16	1.05	1571.6	6.82	***
Marginal R^2: 0.03		*Conditional R^2: 0.27*			
PC2	**Est.**	**SE**	**Df**	**t-value**	**Pr(>|t|)**
(Intercept)	-2.04	0.83	8.35	-2.46	*
Accent: L+H*	5.92	0.42	1530.6	14.14	***
Marginal R^2: 0.09		*Conditional R^2: 0.68*			
PC3	**Est.**	**SE**	**Df**	**t-value**	**Pr(>|t|)**
(Intercept)	0.35	0.36	8.71	0.97	0.359
Accent: L+H*	-1.62	0.25	1584.6	-6.44	***
Marginal R^2: 0.02		*Conditional R^2: 0.54*			
PC4	**Est.**	**SE**	**Df**	**t-value**	**Pr(>|t|)**
(Intercept)	0.22	0.29	8.09	0.78	0.49
Accent: L+H*	-0.95	0.17	1533.7	-5.67	***
Marginal R^2: 0.02		*Conditional R^2: 0.64*			
PC5	**Est.**	**SE**	**Df**	**t-value**	**Pr(>|t|)**
(Intercept)	0.07	0.08	10.92	0.85	0.41
Accent: L+H*	-0.07	0.12	1444.62	-0.62	0.54
Marginal R^2: 0.00		*Conditional R^2: 0.70*			

Table 2: Results of LMEMs for PC1, PC2, PC3, PC4 and PC5; $p < .05$ = *, $p < .01$ = **, $p < .001$ = ***; the marginal and conditional R^2 of each model are presented below the relevant model in italics.

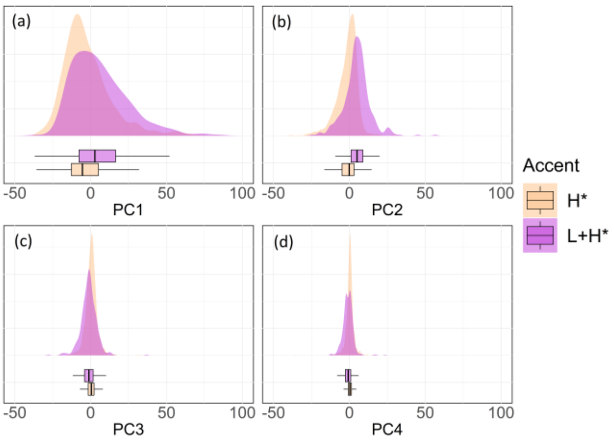

Figure 3: Density and boxplots of PC1, PC2, PC3, and PC4 scores, separately for H* and L+H* accents.

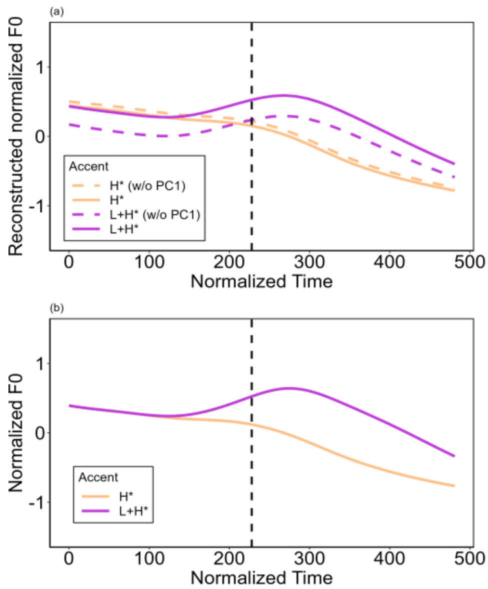

Figure 4: (a): Reconstructed F0 curves by ACCENT using (PC1), PC2, PC3, and PC4, and obtained by incorporating the average values of each PC score, as predicted by the linear model, into the FPCA equation. The equations for the two curves with all four PCs[2] are: for H*, $f(t) = \mu(t) - 1.31 \times PC1(t) - 2.04 \times PC2(t) + 0.35 \times PC3(t) + 0.22 \times PC4(t)$; for L+H*, $f(t) = \mu(t) + 5.85 \times PC1(t) + 3.87 \times PC2(t) - 1.27 \times PC3(t) - 0.73 \times PC4(t)$; (b) smoothed and averaged F0 curves pooled across speakers and tasks. In both plots, the vertical line indicates the onset of the accented syllable.

4. GENERAL DISCUSSION AND CONCLUSIONS

This study showed that H* and L+H* are distinct in Greek unscripted speech. In this dataset, H* was realized as a declining plateau followed by a fall, while L+H* had a slightly “scooped” rise-fall shape (Fig. 4). These findings align with the results reported in [9, 10] for scripted speech, providing supporting evidence that this accentual distinction between H* and L+H* is relevant for Greek intonational phonology.

This conclusion is reached by considering the shape and contribution of the PCs that had significantly different scores for ACCENT, the conditional and marginal R^2 for each LMEM, and the effect of each PC on curve reconstruction (Fig. 4).

As noted, PC1 captured scaling differences but its low conditional and very low marginal R^2 (compared to the other PCs) suggests that these scaling differences do not help distinguish H* from L+H* in Greek. They are most likely due to the changes in pitch level and span that are inevitably present in unscripted speech. This is supported by comparing the reconstructed curves with and without PC1. As shown in Fig. 4a, the curves are identical for H* and show only scaling differences for L+H*, suggesting that the scaling of L+H* is more variable (as indicated by the density plots in Fig. 3a as well). In contrast, while PC2, PC3, and PC4 explained relatively little of the variability in the corpus (21.7% all together), they captured contour shape differences of importance for the accentual contrast. This is reflected in the reconstructed curves which retain the same shape with and without PC1.

Conditional and marginal R^2 exhibited large differences in this analysis. The values of conditional R^2 suggest that, for all PCs, much of the variance was explained by the variables in the random structure. This is likely due to the nature of unscripted speech, in which syllable structure, accented syllable location, and the proximity of other tonal events are not controlled. Despite the extensive and uncontrolled variability in these unscripted data, however, the distinction between H* and L+H* did hold for Greek.

Finally, the use of FPCA followed by statistical modelling has important methodological repercussions for intonation research. Our results show that FPCA is suitable for analysing the intonation of unscripted speech. Most importantly, by using a data-driven, dimension-reduction technique followed by statistical modelling and by taking marginal and conditional R^2 values into account, and considering curve reconstruction, we have shown that it is possible to systematically distinguish variability due to linguistic factors of interest, such as the accentual contrast investigated here, from variability due to other factors, such as speaker-specific and item-specific differences. Further, this methodology allows us to deduce the features that are critical for distinguishing accentual contrasts. Here, for instance, it was shown that scaling differences were not as important as shape for distinguishing H* from L+H*. Taken together, these are helpful steps towards separating linguistic variation from general variability in the study of intonation and facilitate the study of spontaneous data.

5. ACKNOWLEDGEMENTS

The financial support of the European Research Council through grant ERC-ADG-835263 (SPRINT) to Amalia Arvaniti is hereby gratefully acknowledged. We thank our speakers who participated during the COVID-19 pandemic, Eleftheria Politi and Danae Tsinivits for data collection, Dafni Bagioka and Vicky Ioannidou for data annotation, and Katherine Marcoux for data processing in the early stages of this work.

6. REFERENCES

[1] Pierrehumbert, J. B. 1980. *The Phonetics and Phonology of English Intonation*. MIT PhD dissertation.

[2] Pierrehumbert, J. B., Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P. R., Morgan, J., Pollack, M. E. (eds) *Intentions in Communication*. Cambridge: MIT Press, 271–311.

[3] Greif, M., Skopeteas, S. 2021. Correction by Focus: Cleft Constructions and the Cross-Linguistic Variation in Phonological Form. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2021.648478.

[4] Ito, K., Speer, S. R. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58, 541–573. doi:10.1016/j.jml.2007.06.013

[5] Kurumada, C., Brown, M., Bibyk, S.A, Pontillo, D., Tanenhaus, M.K. 2014. Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition* 133(2), 335–342. doi:10.1016/j.cognition.2014.05.017.

[6] Kim, J., Arnhold, A. 2022. Prosodic focus-marking in Canadian English. [Conference poster]. 18th Conference on Laboratory Phonology.

[7] Jepson, K., Zhang, C., Lohfink, G., Marcoux, K., Arvaniti, A. 2021. H* and L+H* in English and Greek. [Conference presentation]. 4th Phonetics and Phonology in Europe.

[8] Arvaniti, A., Baltazani, M. 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In: Sun-Ah Jun (ed) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, 84-117.

[9] Georgakopoulos, T., Skopeteas, S. 2010. Projecting vs. interpretational properties of nuclear accents and the phonology of contrastive focus in Greek. *The Linguistic Review* 27, 315-341. doi: 10.1515/tlir/2020.012

[10] Lohfink, G., Katsika, A., Arvaniti, A. 2019. Variability and category overlap in the realization of intonation. *Proc. 19th ICPhS* Melbourne, 701–705. https://assta.org/proceedings/ICPhS2019.

[11] Asano, Y., Gubian, M. 2015. "Excuse meeee!!": (Mis)coordination of lexical and paralinguistic prosody in L2 hyperarticulation. *Speech Communication* 99, 183–200.

[12] Gubian, M., Torreira, F., Boves, L. 2015. Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49, 16–40. doi:10.1016/J.WOCN.2014.10.001.

[13] Anderson, A. H., Bader, M., Gurnan Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S. et al. 1991. The HCRC Map Task Corpus. *Language and Speech* 34, 351–366.

[14] Zhang, C., Jepson, K., Lohfink, G., Arvaniti, A. 2021. Comparing acoustic analyses of speech data collected remotely. *The Journal of the Acoustical Society of America* 149, 3910–3916. https://doi.org/10/1121/10/0005132.

[15] Boersma, P., Weenink, D. 2022. Praat doing phonetics by computer. [Online]. Available: http:// www.praat.org/.

[16] Baltazani, M., Przedlacka, J., Ünal-Logačev, Ö., Logačev, P., & Coleman, J. 2022. Intonation of Greek in contact with Turkish: a diachronic study. *Language Variation and Change*.

[17] Moritz, S. Bartz-Beielstein, T. 2017. ImputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 9(1), 207-218. https://doi.org/ 10.32614/RJ-2017-009.

[18] R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: http://www.r-project.org/.

[19] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 1.1, 1–48. [Online]. Available: https://www.jstatsoft.org/v067/i01.

---

[1] We also ran the analyses with a *crossed* random structure and the picture does not change with respect to the effect of ACCENT, except that R^2 conditional increases (which is expected given the model change).

[2] The equations for the two curves with PC2-PC4s[2] are: for H*, $f(t) = \mu(t) - 2.04 \times PC2(t) + 0.35 \times PC3(t) + 0.22 \times PC4(t)$; for L+H*, $f(t) = \mu(t) + 3.87 \times PC2(t) - 1.27 \times PC3(t) - 0.73 \times PC4(t)$